

A Data-Driven Framework for Competitive Intelligence in Battery Safety Technologies: Integrating Text Mining and Patent Index Analysis

Donghae Kim ¹, Junghwan Lee ², and Dongwook Kim ^{3,*}

¹ ChoChunShik Graduate School of Mobility, KAIST, Daejeon, Republic of Korea.; eastsea517@kaist.ac.kr

² Department of Management Information Systems, Chungbuk National University, Cheongju, Republic of Korea; junghwan@chungbuk.ac.kr

³ European Telecommunications Standards Institute(ETSI), Sophia Antipolis, France; dongwook.kim@etsi.org

* Correspondence

<https://doi.org/10.5392/IJoC.2026.22.3.001>

Manuscript Received 9 June 2026; Received 22 June 2026; Accepted 10 August 2026

Abstract: Current studies on secondary battery thermal runaway control often overlook the competitive dynamics and infringement risks that arise from portfolio overlaps. This study proposes an applicant-centered framework that integrates patent indices and text-mining-based similarity to map the competitive landscape. We analyzed 1,295 valid patents and utility models (2005–2025) from Korea, China, Japan, and the U.S. using the WIPSON database. Key patent indices based on citations, patent family, and Technology Strength were utilized to quantify applicant influence. By vectorizing claims and abstracts using Term Frequency–Inverse Document Frequency (TF-IDF) and cosine-similarity networks, we found that while Tesla demonstrated a high citation impact with fewer filings, CATL, LG Energy Solution, Samsung SDI, and BYD formed a dense network core. Universities and Tesla occupied peripheral positions, indicating a niche research focus. Time-series analysis revealed a shift from structural protection (2005–2009) to pack-level management (2010–2019) and data-driven control (2020–2025). This framework supports strategic benchmarking, collaboration targeting, and white-space-based R&D planning.

Keywords: Secondary-battery Thermal Runaway Control; Patent Index Analysis; TF-IDF Text Mining; Cosine Similarity Network; Applicant Portfolio Similarity

1. Introduction

Lithium-ion batteries are widely used in electric vehicles (EVs) and energy storage systems (ESS) because of their high energy densities. However, battery fires and thermal runaway, increasing the temperature very rapidly within seconds, continue to occur, leading to casualties and property damage [1] and posing a critical challenge for battery safety. Recent studies have confirmed that thermal runaway is accompanied by intense heat generation and gas release through a series of coupled reactions, including electrolyte decomposition, collapse of the solid electrolyte interphase, and lithium plating [2]. As these mechanisms manifest differently based on the cell material and design configuration, systematic investigations have become necessary.

With the adoption of battery for diverse mobility applications thermal runaway incidents continues, and consequently international test standards and regulations have been established and implemented to prevent such events [3]. In particular, a series of fire accidents in aviation mobility (eVTOL, UAM) and micromobility (for example, electric scooters and electric bicycles) have demonstrated the limitations of conventional mobility-centered safety criteria [4]. Governments and standardization bodies have responded by strengthening the test procedures and design guidelines to prevent thermal runaway and ensure battery safety [5]. Such regulatory measures have extended beyond the technical requirements and include operational guidance. For example, the European Union introduced new regulations governing the full lifecycle of batteries and waste batteries, emphasizing that batteries should be designed to pose no risk to human safety and property [6]. Similarly, the United States Federal Aviation Administration coded the requirements for suppressing thermal

runaway propagation in aviation lithium-ion batteries [7]. China and Japan have also mandated module-level safety testing [8, 9].

Consequently, these developments directly influence the development of corporate technological strategies and the development of patent portfolios [10]. Global battery and mobility companies focus on diverse areas, such as thermal runaway detection and interruption, flame-retardant and fire-spread-mitigation materials, and structures that block intercell heat transfer [11], while strengthening research and development (R&D) strategies and safety-certification-based business models to comply with regulatory requirements [10]. Therefore, thermal runaway control technologies have emerged not only as a means of ensuring safety but also as key determinants of corporate competitiveness and market-entry strategies [11].

However, thermal-runaway-control technologies require substantial time and cost for conception and implementation and inherently involve safety-related risks. Accordingly, efforts are required to reduce inefficient trial and error by systematically analyzing R&D trends and data-driven identification of the technological position and market influence of each player [12]. Traditionally, bibliographic databases have been used to assess research intensity and scholarly networks within a given technological domain [13]. Although academic publications primarily reflect scientific achievements, patents are regarded as a critical element in technology management research because they reveal the technological and commercial strategies of firms and institutions more directly [14-16]. Databases provided by the World Intellectual Property Organization and the European Patent Office present information on applicants, technology classifications, citation relationships, and patent families, thereby enabling quantitative analyses of competitive landscapes and technological clusters [17]. Such patent-based analyses provide a framework for identifying research–industry linkages within the technology ecosystem and serve as scientific evidence not only for establishing R&D strategies, but also for securing competitive advantage within an industry. Accordingly, their importance has become increasingly prominent in technological management research and patent informatics [18].

Patent data have been continuously analyzed from a bibliographic perspective. In particular, a study conducted in 2019 classified wind-power patent texts into five technology domains and analyzed the technology life-cycle stage of each domain using text mining [19]. In addition, another study mapped research trends in lithium-ion battery state-of-health (SoH) estimation through a bibliometric analysis, identifying major themes and their evolution over time based on large-scale literature metadata [20]. Although these studies are useful for proposing market directions and identifying white spaces based on patent texts, they are limited because they do not sufficiently capture competitive relationships, collaboration potential, and patent infringement risks arising from portfolio overlap across applicants. Moreover, they are constrained by the lack of thorough preprocessing of high-quality patents or key players, given that all patents within a domain are analyzed without sufficient filtering or refinement.

This study proposes an analytical framework that combines applicant-level patent index analysis with text-mining-based portfolio analysis of secondary-battery thermal runaway control technologies. For patent index analysis, patent value indicators such as cites per patent (CPP) and patent family size (PFS) were used to evaluate the technological influence and market diffusion of major applicants and identify key players leading the thermal runaway control technology ecosystem within the database. Subsequently, for the same dataset, the abstracts and representative claim texts of the patents held by each applicant were integrated, and text mining was conducted based on term frequency–inverse document frequency (TF–IDF) vectors and cosine similarity. This approach quantitatively evaluates the similarities and technological proximity of patent portfolios among the applicants.

More specifically, this study sought to answer the following three research questions.

RQ.1 Who are the major applicants in the field of secondary-battery thermal runaway control technologies, and what values and characteristics do their patent portfolios exhibit from a patent index perspective?

RQ.2 Based on the patent texts of major applicants, what network structure is formed by interapplicant portfolio similarities and technological proximity?

RQ.3 When integrating the results of the patent index analysis and text-mining-based portfolio similarity analysis, what types of leaders, followers, and niche players can be identified within the thermal runaway-control technology ecosystem, and what do these findings imply?

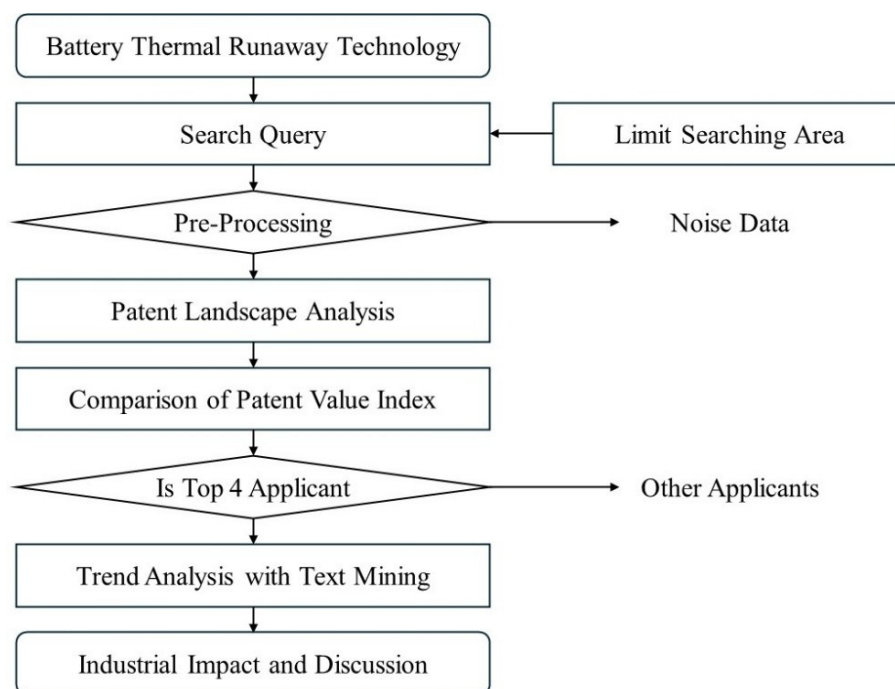


Figure 1. Overall flowchart

In summary, as shown in Figure 1, this paper presents an applicant-centric analytical framework that integrates patent indices and text mining in the field of secondary-battery thermal runaway control technologies. This framework identifies major applicants in the market and quantitatively evaluates the proximity of their technological portfolios. By moving beyond a simple comparison of filing volumes, it enables simultaneous assessment of the patent quality, strategic focus, and technological similarities and differentiations among applicants. Accordingly, this study aimed to provide quantitative evidence that can reduce unnecessary trial and error in establishing future R&D portfolios and determining technology investment directions for thermal runaway control technologies.

The remainder of this paper is organized as follows. Section 2 describes the data collection process and the methodology, including the patent indices and text-mining techniques employed. Section 3 presents the empirical results and analyzes the technological proximity and strategic positioning of major applicants. Section 4 discusses the implications of the findings for R&D strategy and technology investment, and Section 5 concludes the paper by summarizing the main contributions and outlining future research directions.

2. Methodology

2.1 Data Description

The dataset used in this study covered the entire scope of the broad category of “secondary-battery thermal runaway control technologies” and included patents related to both thermal runaway control materials and control systems. The search encompassed both patents and utility models, and the jurisdictions were limited to four IP5 members: Korea (KR), China (CN), Japan (JP), and the United States (US). Only published and granted documents were included, whereas rejected, withdrawn, invalidated, and expired documents were excluded. The time window for the search was set from January 1, 2005, to the data collection cut-off date (February 2025). To prevent the duplicate counting of the same invention, consolidation was performed based on the granted records by merging duplicate entries into the earliest granted record, which was designated as the representative document [21].

2.2 Data Collection

This study extracted data from the WIPS ON database [22], and the following query was used to retrieve patents in the domain of “secondary-battery thermal runaway control technologies.”

Search query (English keywords only):

((secondary battery cell battery lithium battery lithium-ion battery LIB Li-battery lithium-battery ESS energy storage system electric vehicle electric automobile EV (Electrical NEAR1 vehicle)).TI.) AND ((thermal runaway thermal explosion (thermal ADJ1 runaway) thermal-runaway).KEY.) AND ((H01M-010* F01P* G06*).IPC.) AND ((KR JP CN US).CC.) AND ((20*) NOT (2000* 2001* 2002* 2003* 2004*).AD.) AND ((application registration).CSTK.)

In addition to the English terms listed above, the corresponding Korean keywords for each English keyword set were included in the actual query to ensure comprehensive coverage of the Korean-language records. In the query, the whitespace between terms was treated as OR, and the asterisk (*) is a truncation symbol that retrieved all terms beginning with the specified prefix. The query combined: (i) keywords related to secondary batteries and application products within the title field “TI,” (ii) thermal-runaway-related keywords within the “KEY” field (searched across the title, representative claim, and abstract), (iii) relevant International Patent Classification (IPC) codes, (iv) filing jurisdictions (CC), (v) application year (AD), and (vi) document status (CSTK). Keywords explicitly denoting “control technology” were excluded because they are overly generic and could introduce substantial noise into the retrieval results.

The search was further constrained by IPC codes. H01M-010* corresponds to secondary cells and their manufacturing methods (including the subclasses). F01P* corresponds to the cooling of machines and engines, including the cooling of internal combustion engines (including the subclasses). G06* corresponds to computing, calculating, and counting (including the subclasses). Additional restrictions were applied using CC (jurisdiction), AD (application year), and CSTK (document status) field operators.

2.3 Preprocessing

In this study, IPC codes were used to identify data with low technological relevance, and a frequency-based preprocessing was performed. Specifically, the IPC codes in the patent dataset were defined as input variables, and a quantitative frequency analysis was performed using the Counter class in Python. This procedure corresponds to a preprocessing step to ensure the technological relevance and consistency of the dataset while eliminating noise. The overall preprocessing workflow is presented in Figure 2.

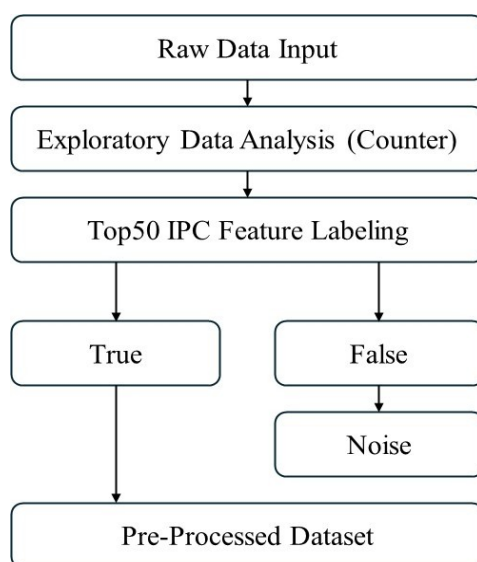


Figure 2. Preprocessing flowchart

First, all IPC codes were examined to calculate the frequency of occurrence of each code, and the top 30 IPCs were identified. This step was conducted to derive the major technology groups represented in the patent dataset and assess the degree of technological concentration. Next, a new binary column was added to the original data frame to indicate whether the IPC code of each patent belonged to the top 30 technology classes. Only patents containing at least one of the top 30 IPCs were defined as valid data, resulting in a final dataset of 1,295 patents for analysis. The results are summarized in Table 1.

Table 1. Preprocessed dataset (overall number of patents = 1,295).

CN	JP	KR	US
1033	32	74	156

2.4 Patent Index-Based Analysis

Various quantitative indicators can be employed to evaluate the technological and market positions of each technology field or applicant based on patent data. The CPP, PFS, and citation index (CI) are widely used to quantitatively assess the qualitative level and industrial influence of the patent portfolio of an assignee (Ernst, 2003). As expressed by Eq. (1), CPP was calculated by dividing the total number of forward citations received by the assignee's granted patents by the number of granted patents held by the assignee (N_a). As expressed by Eq. (3), CI was calculated by dividing the total number of citations made by the assignee's granted patents by N_a . These indices critically measure the patent quality, supporting the assessment of technological advancement and patent value. As expressed by Eq. (2), PFS was calculated by dividing the total number of family jurisdictions associated with the assignee's granted patents by N_a . Accordingly, the PFS is a quantitative measure of the diffusion range and global ripple effect of the assignee's technologies.

$$CPP_a = (\text{Number of citations of registered patents}_a) / N_a \quad (1)$$

$$PFS_a = (\text{Total number of family}_a) / N_a \quad (2)$$

$$CI_a = (\text{Number of patents cited}_a) / N_a \quad (3)$$

The patent impact index (PII) was calculated by dividing a specific assignee's CPP by the overall CPP computed across all assignees, as expressed by Eq. (4). The technology strength (TS) index was calculated by multiplying the assignee's PII with the number of granted patents held by the assignee, as expressed by Eq. (5) [23]. PII captures an assignee's performance level relative to the average technological level, whereas TS reflects the technological capability of a specific firm or country. These indicators jointly account for both the qualitative and quantitative dimensions of a technological output.

$$PII_a = CPP_a / CPP_{total} \quad (4)$$

$$TS_a = PII_a \times N_a \quad (5)$$

2.5 Text-Mining-Based Analysis

TF-IDF vectorization is a method for extracting important terms from each patent document and evaluating how discriminative those terms are relative to other patents. Term frequency (TF) represents the frequency of a specific term within a document and is calculated using Eq. (6). IDF is a weighting scheme in which a term that appears frequently across many documents is assigned a lower importance because such a term has limited discriminative power. The IDF was calculated using Eq. (7). The final TF-IDF value was obtained by multiplying the TF and IDF, as expressed in Eq. (8). Accordingly, this approach enables the vectorization and comparison of patent texts by reflecting not only term frequency, but also term importance.

$$TF(t, d) = (\text{Number of appearances of word } t \text{ in patent } d) / (\text{Total number of words in patent } d) \quad (6)$$

$$IDF(t) = \log((\text{Total number of patents}) / (\text{Number of patents containing word } t)) \quad (7)$$

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (8)$$

The transformed vectors were used to compute the technological similarity between firms using cosine similarity. Cosine similarity measures the angular distance between two vectors to quantify the similarity between two patents, that is, the extent of technological overlap. The similarity value is computed on a scale of 0 to 1, in which values closer to 1 indicate that the technological contents of the two patents are highly similar. Cosine similarity is defined by Eq. (9).

$$\text{cosine similarity} = (A \cdot B) / (||A|| \times ||B||) \quad (9)$$

3. Results

3.1 Quantitative Comparison Based on Patent Index

The results for the top 12 applicants (representing a total of 1,295 patents) are presented in Figure 3. Tesla Motors had a high CPP value (approximately 37.44). The mean reference line was presented in two forms to prevent distribution distortion. When Tesla was included, the average CPP was 5.41, whereas it decreased to 2.50 when Tesla was excluded. LG Chem, CATL, and Samsung SDI were the only applicants whose CPP and PFS values exceeded their respective averages. By contrast, Beihang University and Tsinghua University yielded above-average CPP values but below-average PFS values, reflecting the tendency of universities and research institutions to emphasize technological impacts rather than multi-jurisdictional rights expansion. By contrast, LG Energy Solution yielded a below-average CPP but an above-average PFS.

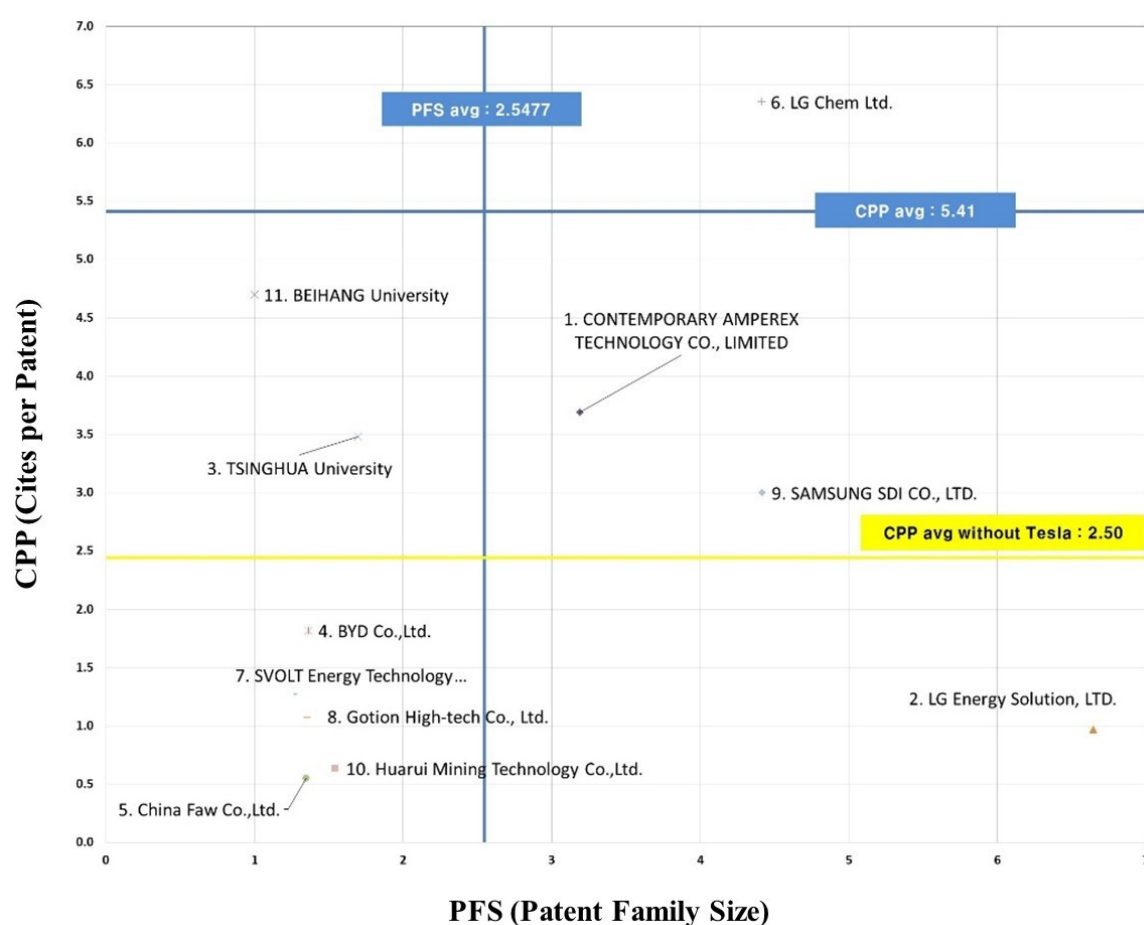


Figure 3. CPP/PFS for the top 12 applicants by number of patents

Overall, the findings indicate that the qualitative and quantitative dimensions of patent portfolios do not necessarily move in tandem. For instance, research-oriented applicants tend to show high CPP but low PFS, suggesting that knowledge diffusion exhibits a significant influence. By contrast, players with strengths in PFS, such as LG Energy Solution, exhibit a relatively pronounced orientation toward securing portfolio coverage rather than maximizing citation-based technological impact.

After calculating the citation-related indices, the results were divided into four groups (Figure 4). This classification was based on two axes. First, for the CPP axis, the average CPP, excluding Tesla (an extreme outlier), was used as the threshold; applicants with CPP values higher than the mean were categorized as the leading group, whereas those lower than the mean were categorized as the follower group. Second, for the CI axis, the overall mean CI was used as the threshold: applicants higher than the mean were defined as belonging to a mature market, whereas those lower than the mean were defined as belonging to an early-stage market.

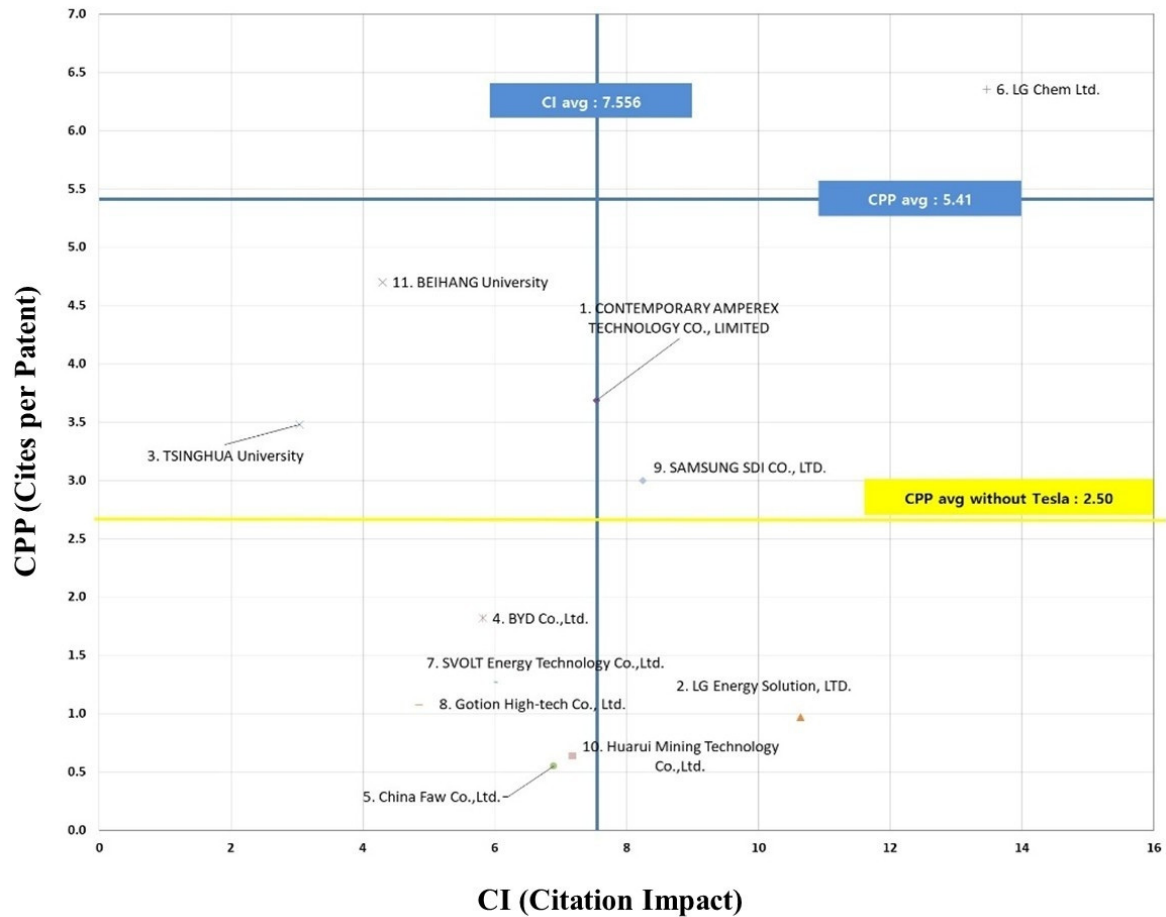


Figure 4. CPP/CI for the top 12 applicants by number of patents

The mature-market-leading group included LG Chem and Samsung SDI. The early-stage-market-leading group comprised Beihang University and Tsinghua University. The mature-market follower group included LG Energy Solution, and the early-stage-market follower group included BYD. The CPP, PFS, and CI indices for the top 12 applicants by filing volume are summarized in Table 2.

Table 2. Values of CPP, PFS, and CI for the top 12 applicants by number of patents.

Major Applicant	CPP	PFS	CI
Contemporary Amperex Technology	3.69	3.19	7.54
LG Energy Solution	0.97	6.65	10.65
Tsinghua University	3.48	1.70	3.04
BYD	1.82	1.36	5.82
China Faw	0.55	1.35	6.90
LG Chem	6.35	4.41	13.47
SVOLT Energy Technology	1.27	1.27	6.00
Gotion High-tech	1.07	1.36	4.86
SAMSUNG SDI	3.00	4.42	8.25
Huarui Mining Technology	0.64	1.55	7.18
Beihang University	4.70	1.00	4.30
Tesla Motors	37.44	2.33	12.67

PII normalizes an assignee's total forward citations by the overall forward citations and can be interpreted with 1 as the baseline. A PII value lower than 1 indicates a below-average market impact, whereas a value higher than 1 indicates an above-average impact. The analysis showed that Tesla recorded the highest PII (16.63). The PII values for CATL (1.64), LG Chem (2.82), and Samsung SDI (1.33) also exceeded 1. In addition, university applicants, including those from Tsinghua University (1.55) and Beihang University (2.09), exhibited PII values higher than 1. The fact that research-oriented applicants maintained PII values higher than 1 implies that even after normalization, as also indicated by the CPP results, they retained comparatively high technological influence relative to other players. This suggests a strong potential for knowledge diffusion and foundational technologies in terms of citation impact.

Nevertheless, the limitations of the PII are clear. Although the PII indicates the magnitude of relative influence, it does not directly provide absolute and quantitative information on the scale of patenting activity. Similarly, in the TS analysis, Tesla recorded a dominant value (149.70), mainly driven by its high CPP. In particular, the value for CATL, which ranked below LG Chem in the PII analysis, increased to 78.69 in TS, approximately half of Tesla's level and nearly 1.5 times that of LG Chem (47.96). This result indicates that CATL filed more patents than LG Chem. Specifically, even if the per-patent influence is comparable to or lower than that of LG Chem, the combination of portfolio scale and impact leads to enhanced overall technology strength. In this regard, LG Chem is quality-focused, whereas CATL adopts a scale-accumulation strategy. Overall, the technological indices suggest that Tesla and CATL significantly influence the ecosystem. The results are presented in Figure 5 and Table 3.

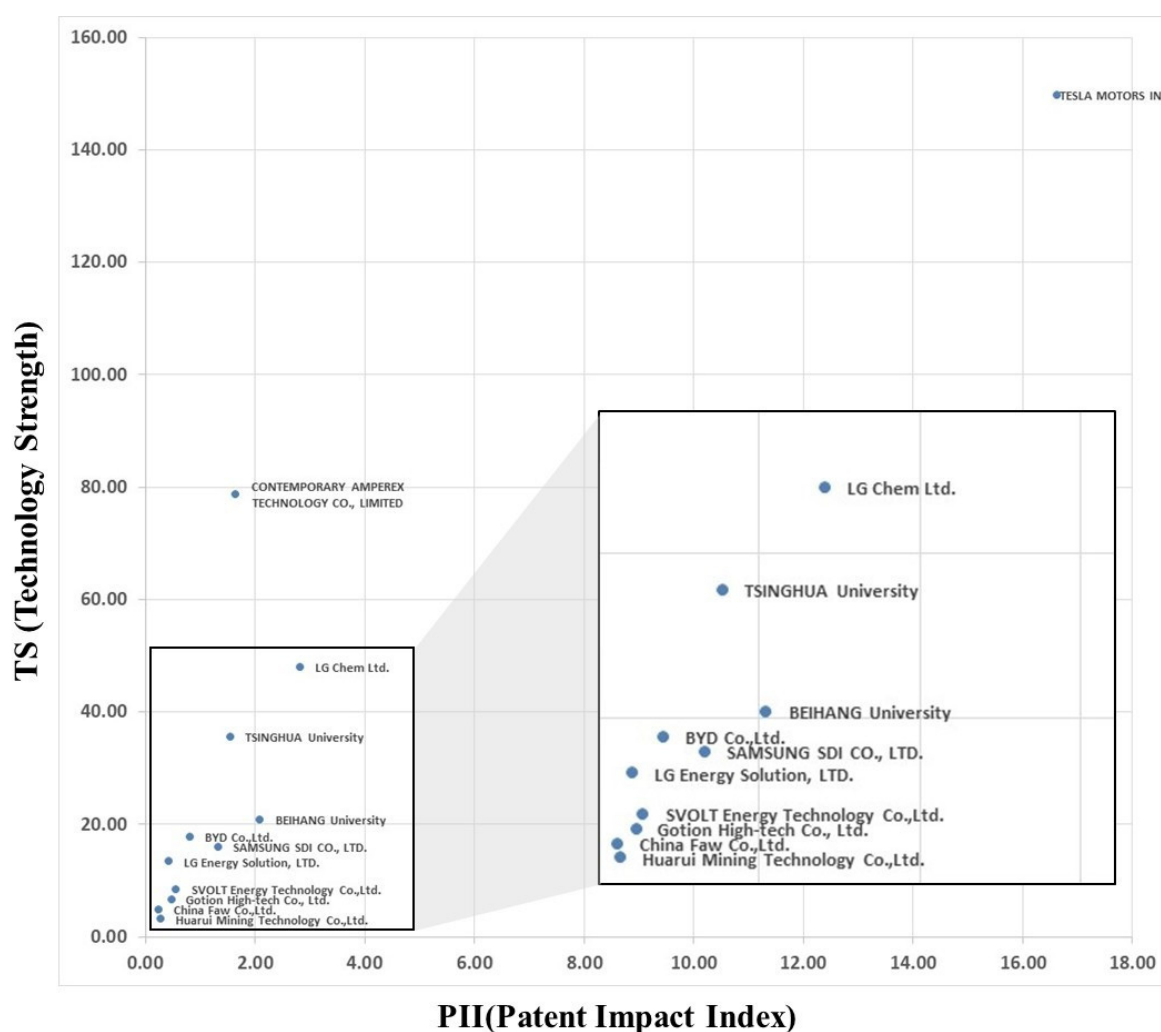


Figure 5. PII/TS for the top 12 applicants by number of patents

Table 3. Values of PII, TS for the top 12 applicants by number of patents

Major Applicant	PII	TS
Contemporary Amperex Technology	1.64	78.69
LG Energy Solution	0.43	13.36
Tsinghua University	1.55	35.56
BYD	0.81	17.79
China Faw	0.24	4.89
LG Chem	2.82	47.96
SVOLT Energy Technology	0.56	8.46
Gotion High-tech	0.48	6.65
SAMSUNG SDI	1.33	15.99
Huarui Mining Technology	0.28	3.13
Beihang University	2.09	20.88
Tesla Motors	16.63	149.70

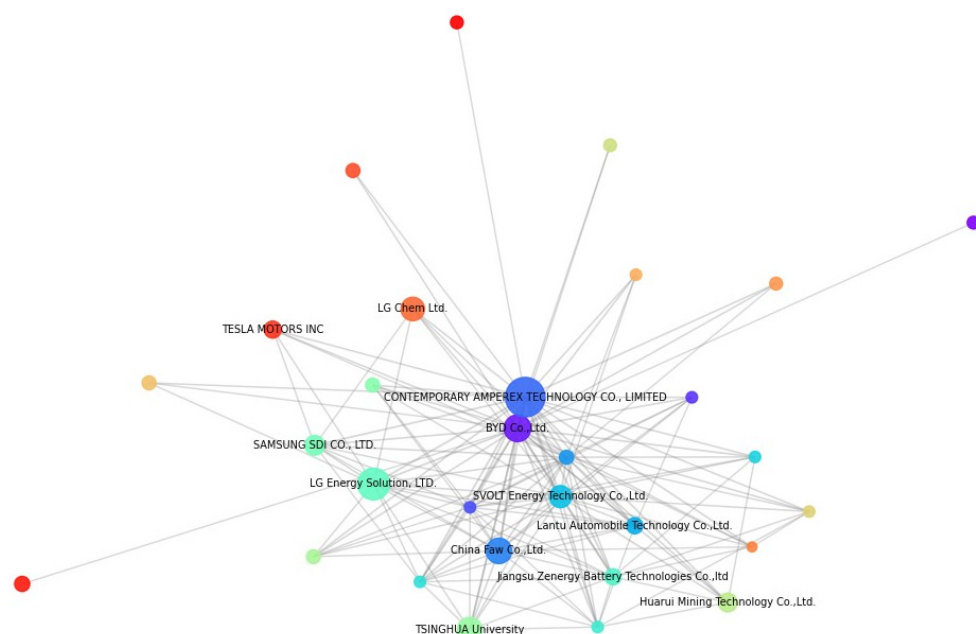
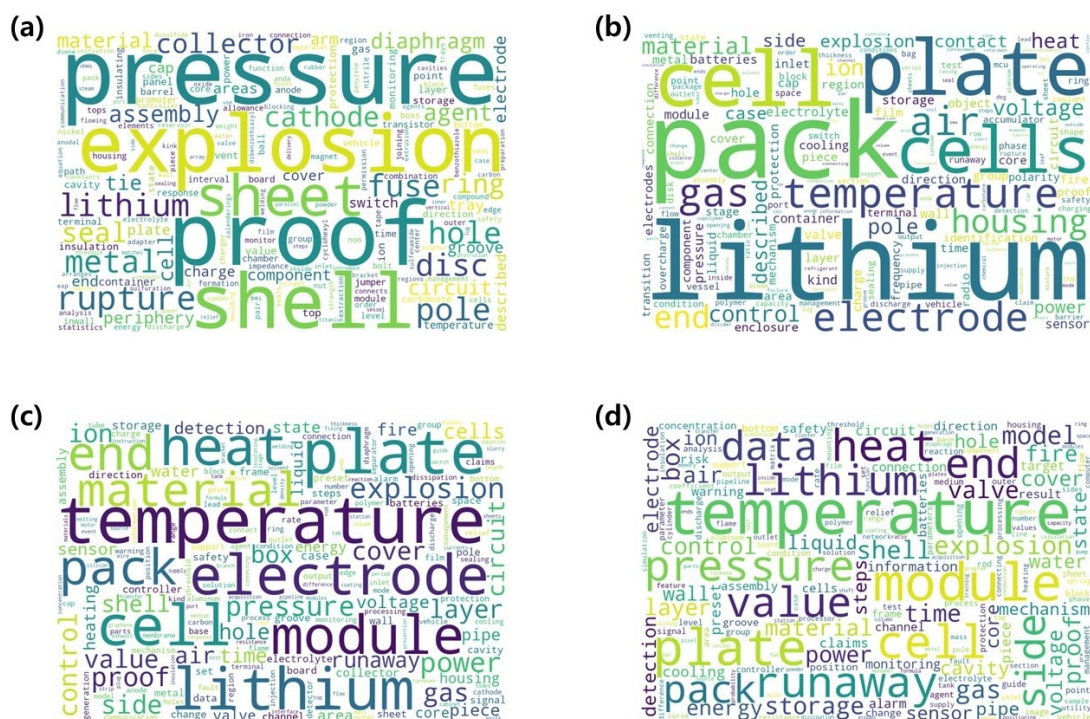
A quantitative indicator analysis based on these patent indices distinguishes each applicant's influence (qualitative dimension) from its filing scale (quantitative dimension). However, value-based indicators alone are insufficient for explaining thematic similarities across portfolios or the proximity that underlies competitive and collaborative relationships. In this study, a cosine-similarity network for the same sample was constructed by vectorizing representative claims and abstract texts using TF-IDF, and the results were cross-validated across two dimensions (value and text).

3.2 Bibliometric Comparison Based on Text Mining

The text mining of independent claims, shown in Figure 6, revealed clear year-to-year technological shifts in thermal runaway control technologies. In the word clouds for 2005–2009, prominent keywords included *pressure* (64 occurrences), *explosion* (53), and *shell* (48). For 2010–2014, *lithium* (136), *pack* (135), and *cell/cells* (128 and 121, respectively) were among the most frequently occurring terms, and keywords associated with failure-related factors at the pack level, such as *temperature* (117), *gas* (107), and *air* (106), also emerged. For 2015–2019, the overall frequency of the main keywords increased significantly, with *temperature* (921), *plate* (819), and *lithium* (735) appearing the most frequently. Finally, for 2020–2024, new terms such as *data* (760) and *model* (661) appeared, and the relative prominence of *control* (556) increased.

To examine the interapplicant patent portfolio similarity, the representative claims and abstract texts of each firm were used as the primary analytical sources. After concatenating the patent texts for each player into a single document, term importance was quantified using TF-IDF vectorization.

After computing the pairwise similarities, edges were drawn only when the similarity value was at least 0.7. This threshold was applied to include only firms with high technological similarities in the network, thereby identifying relationships with an increased likelihood of competition or collaboration. A network was constructed by connecting the edges representing interfirm similarity to the nodes representing the firms. The node size was scaled proportionally to the number of patents held by each firm, whereas the edge thickness represented the degree of technological similarity between the two firms. Specifically, the edge thickness was set proportional to the cosine similarity value; thicker edges indicated higher similarity and closer technological proximity between the two firms than thinner edges. The network density reflected the extent to which technologically similar firms formed clusters. The resulting network is shown in Figure 7.



4. Discussion

The methodological contribution of this study lies in the combined examination, within the same sample, of value-based indicators that reflect patent impact, extensibility, scale, and text-based portfolio similarity (thematic proximity), thereby enabling the simultaneous diagnosis of both the level of influence and similarity

to competitors. Value indicators alone are insufficient to explain the overlap in technological content and competitive landscapes, whereas text similarity alone cannot capture the market diffusion and industrial impact of technologies. Accordingly, by cross-validating these two dimensions, this study mitigates overinterpretation that could arise from single-metric analyses—for example, assuming that a highly cited player is necessarily aligned with mainstream technologies, or that a portfolio-isolated player is inherently less important. Moreover, the study provides interpretations that are applicable to competition–collaboration assessment and the identification of white-space opportunities.

4.1 Typology of Applicants based on Patent Value Indices

Players in the mature-market-leading group (LG Chem and Samsung SDI) can be regarded as applicants exhibiting strong technological leadership because they have a substantial number of patents that are both highly cited and associated with technologies that have reached an advanced stage. Players in the early-stage-market-leading group (Beihang University and Tsinghua University) yielded high CPP and PII values but low PFS, indicating that their international diffusion is relatively limited, whereas their technological innovativeness is high; thus, they can be characterized as research-institution-type applicants. Players in the mature-market follower group (for example, LG Energy Solution) participate in advanced technology areas but exhibit relatively low influence indicators, suggesting strategies focused on complementing incumbent technologies or targeting white-space domains. Players in the early-stage-market follower group (for example, BYD) have many patents with low impact and low maturity. These applicants appear to be in the exploratory phase regarding technology applicability and may possess growth potential, depending on future market dynamics. Tesla Motors recorded overwhelmingly high values in both PII and TS, indicating that, despite a limited number of filings, its individual patents exert significant influence. Thus, the number of patents is not necessarily proportional to market-level impact.

4.2 Interpreting Portfolio Similarity Network Considering Patent Indices

A cross-analysis of the text-based portfolio similarity network and patent index results reveals several notable patterns. First, CATL, LG Energy Solution, BYD, SVOLT Energy Technology, and China FAW are positioned near the center of the network, while also exhibiting CPP, PFS, and TS values above a specific level. These applicants form mutually similar portfolios around core themes in the EV battery ecosystem, such as thermal management, battery module safety, and thermal runaway protection systems, suggesting competitive–collaborative relationships with considerable potential for technological overlap. In particular, CATL and LG Energy Solution obtained high TS and above-average CI values, indicating that both their index-based influence and network centrality were high. Thus, they can be regarded as hub applicants within the technology domain.

By contrast, Tsinghua University and Beihang University were located on the periphery of the network, despite their high CPP and PII values. This pattern suggests that although these institutions hold influential patents, they focus on topic areas or research directions that differ from those of other major applicants. This mismatch may be attributable to the fundamental research orientation of academic institutions or to heterogeneous terminology and expressions in patent specifications.

Tesla Motors recorded top-tier PII and TS values but appeared comparatively isolated in the portfolio similarity analysis. This suggests that Tesla's core patents may be grounded in proprietary design concepts or system architectures that are not widely shared with other applicants, or that its relatively closed IP strategy may have driven it in directions that diverge from generalized technology clusters. Such discrepancies demonstrate that significant influence in patent indices does not necessarily correspond to high text-based similarity, implying that competitive advantage can be shaped not only by mainstream technologies, but also by exclusive and differentiated innovations.

4.3 Strategic Implications for Competition, Collaboration, and White Space

The combined analysis of text mining and patent indices improves the assessment of the competitive and collaborative potential within the thermal runaway control technology ecosystem. Applicants that are strongly interconnected at the center of the network address identical or highly similar technological challenges. Therefore, they are likely to engage in intense competition in industrial practice while also retaining opportunities for collaboration in other areas, such as standardization, joint responses to regulatory requirements, and certification framework establishment.

Applicants located on the periphery of the network, although exhibiting high index values, can be interpreted as players with niche or disruptive innovation potential. The patent portfolios of Tesla Motors and some university applicants appear to form distinctive technology clusters that diverge from the dominant clusters. Accordingly, these portfolios may play crucial roles in future market restructuring and emerging regulatory environments. Such applicants may occupy white-space domains characterized by relatively few direct competitors and clear technological differentiation, which can carry high strategic value in the long term.

In addition, the findings support firms and research institutions seeking to establish or refine R&D portfolios for thermal runaway control technologies by providing evidence for decisions regarding who to benchmark, which applicants to avoid competing against, and which applicants to consider for collaboration, joint filing, or technology transfer. For example, hub-type applicants positioned at the network center exert a strong influence on the current market, whereas niche-type applicants positioned at the periphery may become increasingly important as new application areas emerge or regulatory modifications create new opportunities.

5. Conclusions

This study quantitatively analyzed patent data in the domain of secondary-battery thermal runaway control technologies and evaluated the proximity of technological portfolios among major applicants by combining patent-index-based metrics with TF-IDF-based text mining. A total of 1,295 thermal-runaway-control-related patents filed in KR, CN, JP, and US from January 2005 to February 2025 were collected from the WIPS ON database, and a valid patent set was selected through IPC code-based preprocessing. Subsequently, applicant-level patent portfolios were vectorized, and an integrated analytical framework combining patent indices and text mining was applied to address the three research questions (RQ.1–RQ.3) proposed in the Introduction, as summarized below.

For the first question, RQ.1, “Who are the major applicants in secondary-battery thermal runaway control technologies, and what values and diffusion characteristics do their patent portfolios exhibit from a patent index perspective,” an analysis using CPP, PFS, CI, PII, and TS identified CATL, LG Chem, and Samsung SDI as core applicants with high competitiveness at both the technological and market extensibility levels. These organizations exceeded the averages in both CPP and PFS. Although Tesla filed a limited number of patents, it exhibited a high-impact portfolio with high CPP, PII, and TS values. Tsinghua University and Beihang University showed high CPP and PII but low PFS. Hence, they were categorized as research-institution-type applicants characterized by strong technological innovativeness but limited international diffusion. By contrast, LG Energy Solution recorded below-average CPP but above-average PFS, indicating a diffusion-oriented portfolio with broad coverage but relatively limited citation impact at the individual patent level.

For the second question (RQ.2), “Based on patent texts of major applicants, what network structure is formed by portfolio similarities and technological proximity across applicants,” a cosine similarity network was constructed using TF-IDF vectors derived from combined representative claims and abstract texts. CATL, LG Energy Solution, Samsung SDI, and BYD were located in the network core and exhibited high connectivity and technological proximity. This indicates that these firms address common challenges along similar technological axes, leading to major technology clusters. By contrast, universities and research institutions formed distinctive topic areas at the network periphery, suggesting research directions that diverged from established commercial clusters. Accordingly, industry–academia collaboration between the industrial cluster at the core and research institutions at the periphery can be interpreted as a key mechanism that connects separated clusters and accelerates the translation of research outcomes into commercialization, thereby driving technological advancement in this domain.

For the third question (RQ.3), “When the results of patent index analysis and text-mining-based portfolio similarity analysis are interpreted jointly, what types of leaders, followers, and niche players can be identified within the thermal runaway control technology ecosystem, and what implications do they provide for potential competition–collaboration and portfolio strategies,” a cross-examination of the two analytical dimensions identified CATL, LG Chem, Samsung SDI, LG Energy Solution, and BYD as hub-type leaders, which showed high index values and high network centrality. Although these applicants are in direct competition within the same technology cluster, they also have the potential to function as collaborative partners from a regulatory perspective. By contrast, applicants such as Tsinghua University, Beihang University, and Tesla, which show high influence in index-based measures but relative isolation in the network, can be interpreted as peripheral niche players differentiated from mainstream clusters. Tesla simultaneously demonstrated a dominant influence on index-based measures and an independent position in the text-similarity network, suggesting that it may

build a competitive advantage through differentiated designs and system architectures, rather than through generalized, widely shared technologies. Such typologies can inform portfolio strategies, including the selection of benchmarking targets, competition avoidance versus focus strategies, and decisions on joint research and licensing, depending on the applicant's positioning.

With respect to the time-series text-mining results, the 2005–2009 period was dominated by physically protective keywords, such as pressure, explosion, and shell. From 2010 to 2019, the relative prominence of keywords related to thermal and gas behavior control, such as temperature, gas, and pack, increased. From 2020 to 2024, keywords associated with intelligent management and control, including data, model, and control, emerged as major terms. This evolution indicates that thermal runaway control technologies have progressed from a phase centered on structural protection to pack- and module-level thermal management considering heat and gas behavior, and further toward data-driven smart management systems and model-based control.

In summary, this study proposes a quantitative framework that integrates applicant-level patent index analysis and text-mining-based portfolio similarity analysis to elucidate the competitive landscape, paradigm shifts, and strategic positioning of major players in the thermal runaway-control technology ecosystem. These findings have academic and practical implications for establishing R&D portfolios, analyzing collaboration and competition structures, and setting investment directions in thermal runaway control technologies.

This study has several limitations. First, because the scope of data collection was limited to KR, CN, JP, and US among the IP5, patent trends in Europe and other emerging countries were ignored. Consequently, global patent activity may have been underestimated, and future studies should employ expanded datasets, including Europe and additional jurisdictions. Second, quantitative indicators such as CPP, PFS, CI, PII, and TS can be affected by citation lag, cross-country differences in citation practices, and cross-citations among affiliates. Therefore, these metrics should be interpreted as measures of relative influence within a given period rather than as an absolute patent value. Addressing these issues, for example, by introducing correction metrics that account for citation time-series effects, remains an important future research direction. Third, portfolio proximity analysis based on text mining and cosine similarity focuses on word-level similarity and may not sufficiently reflect differences in implementation maturity or experimental performance. This limitation is particularly relevant for recent artificial-intelligence-related technologies, where heterogeneous patent phrasing introduces uncertainty into semantic interpretation. Therefore, future studies should consider combining paragraph-level embeddings and semantic clustering based on large language models. Such extensions may be applicable not only to thermal runaway control technologies, but also to related technology domains characterized by high risk and high cost.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] A.W. Golubkov, D. Fuchs, J. Wagner, H. Wiltse, C. Stangl, G. Fauler, G. Voitic, A. Thaler, V. Hacker, "Thermal-runaway experiments on consumer Li-ion batteries with metal-oxide and olivin-type cathodes," *RSC Adv.*, vol. 4, no. 7, pp. 3633–3642, 2014, doi: <https://doi.org/10.1039/c3ra45748f>.
- [2] P. J. Bugryniec, E. G. Resendiz, S. M. Nwophoke, S. Khanna, C. James, S. F. Brown, "Review of gas emissions from lithium-ion battery thermal runaway failure — Considering toxic and flammable compounds," *J. Energy Storage*, vol. 87, 111288, 2024, doi: <https://doi.org/10.1016/j.est.2024.111288>.
- [3] V. Ruiz, A. Pfrang, A. Kriston, N. Omar, P. Van den Bossche, and L. Boon-Brett, "A review of international abuse testing standards and regulations for lithium ion batteries in electric and hybrid electric vehicles," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1427–1452, 2018, doi: <https://doi.org/10.1016/j.rser.2017.05.195>.
- [4] M. Segalovich, A. Berl, U. Aviv, E. Jaffe, R. Shelef, J. Haik, M. Cleary, R. Kornhaber, and M. Harats, "The hidden danger of lithium battery-powered electric bicycles and scooters: A case series of the Israeli National Burn Center experience," *J. Burn Care Res.*, vol. 43, no. 2, pp. 504–507, 2022, doi: <https://doi.org/10.1093/jbcr/irab240>.
- [5] NSW Government, New safety requirements for lithium-ion e-micromobility products like e-bikes and e-scooters, NSW Fair Trading, 2025.
- [6] European Commission, Proposal for a Regulation of the European Parliament and of the Council concerning batteries and waste batteries, repealing Directive, Brussels, COM(2020) 798 final. Accessed: Dec. 10, 2020.

- [7] US Department of Transportation, Pipeline and Hazardous Materials Safety Administration (PHMSA), Hazardous materials: Enhanced safety provisions for lithium batteries transported by aircraft (FAA Reauthorization Act of 2018), Fed. Regist., vol. 87, pp. 77995–78011, 2022. [Online] Available: <https://www.federalregister.gov/d/2022-27563>
- [8] Xinhua News Agency, China's new EV battery safety standard to take effect in July 2026, Ministry of Industry and Information Technology, Beijing, 2025.
- [9] I. Hoey, Japan updates lithium-ion battery safety standards for enhanced monitoring and fire prevention, Int. Fire Saf. J., Apr. 10, 2024.
- [10] J. Zhang, T. Fan, S. Yuan, C. Chang, K. Wang, Z. Song, and X. Qian, "Patent-based technological developments and surfactants application of lithium-ion batteries fire-extinguishing agent," J. Energy Chem., vol. 88, pp. 39–63, 2024, doi: <https://doi.org/10.1016/j.jechem.2023.08.037>.
- [11] X. Feng, D. Ren, X. He, and M. Ouyang, "Mitigating thermal runaway of lithium-ion batteries," Joule, vol. 4, no. 4, pp. 743–770, 2020, doi: <https://doi.org/10.1016/j.joule.2020.02.010>.
- [12] A. L. Porter and S. W. Cunningham, Tech Mining: Exploiting New Technologies for Competitive Advantage, John Wiley & Sons, Hoboken, New Jersey, 2005. [Online] Available: <https://onlinelibrary.wiley.com/doi/book/10.1002/0471698466>
- [13] M. Meyer, "What is special about patent citations? Differences between scientific and patent citations," Scientometrics, vol. 49, pp. 93–123, 2000, doi: <https://doi.org/10.1023/A:1005613325648>.
- [14] F. Narin, K. S. Hamilton, and D. Olivastro, "The increasing linkage between U.S. technology and public science," Resour. Policy, vol. 26, no. 3, pp. 317–330, 1997, doi: [https://doi.org/10.1016/S0048-7333\(97\)00013-9](https://doi.org/10.1016/S0048-7333(97)00013-9).
- [15] H. Ernst, "Patent information for strategic technology management," World Pat. Inf., vol. 25, no. 3, pp. 233–242, 2003, doi: [https://doi.org/10.1016/S0172-2190\(03\)00077-2](https://doi.org/10.1016/S0172-2190(03)00077-2).
- [16] D. Dekker, D. Christopoulos, and H. McGregor, "Dynamics of knowledge production: A relational-event analysis of patent citation hazards in hydrogen technologies," Sustainable Futures, vol. 10, 101460, 2025, doi: <https://doi.org/10.1016/j.sftr.2025.101460>.
- [17] H. Ernst and N. Omland, "The Patent Asset Index—A new approach to benchmark patent portfolios," World Pat. Inf., vol. 33, no. 1, pp. 34–41, 2011, doi: <https://doi.org/10.1016/j.wpi.2010.08.008>.
- [18] T. U. Daim, G. Rueda, H. Martin, and P. Gerdri, "Forecasting emerging technologies: Use of bibliometrics and patent analysis," Soc. Change., vol. 73, no. 8, pp. 981–1012, 2006, doi: <https://doi.org/10.1016/j.techfore.2006.04.004>.
- [19] M. D. Madvar, F. Ahmadi, R. Shirmohammadi, and A. Aslani, "Forecasting of wind energy technology domains based on the technology life cycle approach," Energy Rep., vol. 5, pp. 1236–1248, 2019, doi: <https://doi.org/10.1016/j.egyr.2019.08.069>.
- [20] A. Jose and S. Shrivastava, "The advancements and research trends in state of health estimation for lithium-ion batteries over the last two decades: A bibliometric examination of prospective trends," Sustainable Futures, vol. 10, 101332, 2025, doi: <https://doi.org/10.1016/j.sftr.2025.101332>.
- [21] S. Lee, B. Yoon, and Y. Park, "An approach to discovering new technology opportunities: Keyword-based patent map approach," Technovation, vol. 29, no. 6–7, pp. 481–497, 2009, doi: <https://doi.org/10.1016/j.technovation.2008.10.006>.
- [22] J. S. Eom, D. C. Kim, J. H. Han, and W. G. Bae, "Enhancing mechanical performance of thick steel plates for offshore wind structures: A classification and patent landscape study," World Pat. Inf., vol. 84, 102419, 2026, doi: <https://doi.org/10.1016/j.wpi.2025.102419>.
- [23] L. Aristodemou and F. Tietze, "Citations as a measure of technological impact: A review of forward citation-based measures," World Pat. Inf., vol. 53, pp. 39–44, 2018, doi: <https://doi.org/10.1016/j.wpi.2018.05.001>.

