Standigm
AI for new drug R&D

SPICA™

AI Powered Novelty Assessment

Executive Summary

The high cost and uncertainty of late-stage intellectual-property (IP) conflicts remain a major bottleneck in drug discovery. **SPICA™** eliminates this risk early by providing an automated, real-time estimate of a compound's patentability. Powered by a graph-transformer architecture that excels at capturing both local chemistry and long-range scaffold context, the system is **pretrained on 100 million synthetically accessible molecules** and **fine-tuned on 10 million patented structures**, aligning its embedding space with legal—not merely structural—similarity.

A single SMILES query is vectorized, compared against the full patent corpus via ultra-fast similarity search, and returned with a **0–1 novelty score** and a ranked list of closest prior-art compounds in just a few seconds. Internal validation on 63 proprietary molecules shows **96.8 % concordance with expert patent attorneys**; the system also retrieved the exact patents flagged by experts in **93 % of cases**. Longitudinal **analyses of FDA-approved drugs** confirm that SPICA™'s score tracks the expected decline in novelty after patent filing, underscoring its real-world relevance.

Accessible through a **REST API or batch upload**, SPICA™ integrates seamlessly with electronic lab notebooks and generative-design pipelines, and its database is refreshed continuously as new patents appear. By moving patentability checks to the very first design cycle, SPICA™ shortens lead-optimization timelines, lowers external legal spend, and safeguards future market exclusivity.

Introduction

The process of drug development is complex and time-consuming, often spanning several years and costing millions of dollars. One of the critical phases in this process is lead optimization, where the focus is on enhancing the properties of a potential drug candidate to ensure its efficacy, safety, and patentability. In this phase, patentability is a paramount factor as it determines the commercial viability and market exclusivity of the drug. SPICA™ offers a

groundbreaking approach to predicting the patentability of designed drugs, leveraging advanced machine learning techniques and a vast patent database.

Overview of SPICA™

Standigm's SPICA™ is a state-of-the-art predictive tool designed to assess the patentability of drug candidates swiftly and accurately. It utilizes a machine learning model trained on an extensive dataset of patent information, allowing it to provide insights into the likelihood of patent conflicts and assess the novelty of the compound.

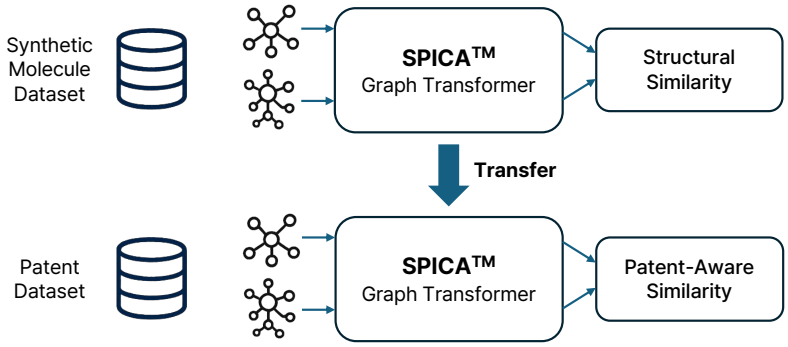

Figure 1. Overview of the SPICA™ training process.

Graph Transformer

SPICA™ employs a powerful graph transformer-based architecture as its backbone, which has recently shown superior performance compared to traditional graph neural networks. By utilizing attention mechanisms, graph transformers capture both local details and long-range interactions between atoms within a chemical compound, enabling a more thorough understanding of complex molecular structures.

Metric Learning

Metric learning is a process where the model learns to form an embedding space reflecting similarities and differences between data points. Leveraging state-of-the-art metric learning techniques, SPICA™ constructs a rich embedding space, accurately discerning subtle structural similarities and differences to identify novel compounds and predict patent conflicts.

Extensive Training Dataset

SPICA™ is first pretrained on a large-scale dataset comprising over 100 million paired synthetic molecules to learn intricate structural similarities between molecules. The pretraining phase is then complemented by fine-tuning on a comprehensive patent database consisting of over 10 million patented molecules, ensuring that the model's representations are both robust and relevant to assessing the novelty of compounds.

Vector Similarity Search

SPICA™'s efficiency is further amplified through vector search techniques, enabling rapid retrieval of relevant compounds from the database. This accelerates the similarity search process, enhancing the model's speed and scalability in analyzing large patent datasets. With SPICA™, searching for similar compounds and assessing the novelty of a molecule can be done in just a few seconds.

Detailed Workflow of SPICA™

Search for Similar Compounds in Standigm's Patent Database

Users can input the chemical structure of the query compound using a list of SMILES (Simplified Molecular Input Line Entry System) strings. SPICA™ then searches its extensive patent database, which contains more than 10 million entries of small molecules in patents. This database includes detailed information on patented compounds, their structures, and related documentation. The model identifies compounds in the database that are structurally similar to the query compound. This similarity search is crucial as it highlights potential prior art that could impact the patentability of the new compound.

Quantitative Novelty Scoring

Based on the similarity search, SPICA™ calculates a quantitative novelty score for the query compound. This score reflects the uniqueness of the compound in comparison to molecules in related patents and indicates the likelihood of obtaining a new patent without conflicts.

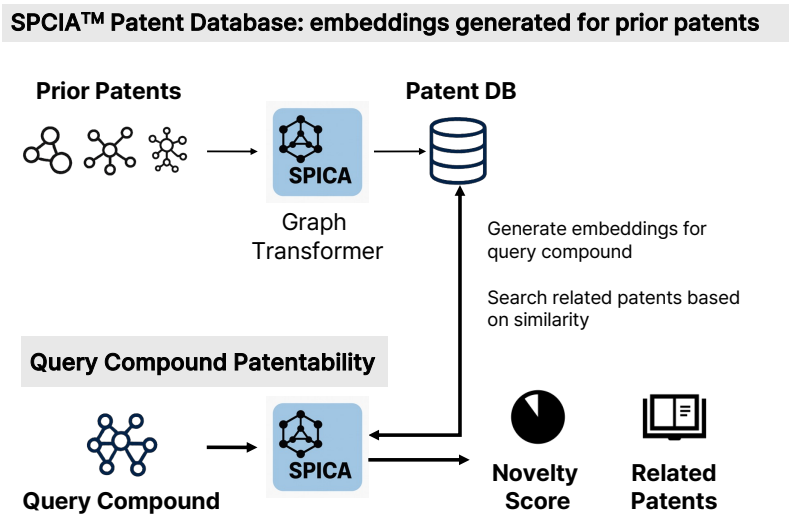

Figure 2. SPICA™ workflow for novelty evaluation.
A query molecule is processed by the SPICA™ GNN for vector search and novelty scoring.

Novelty Assessment on Our In-house Dataset

We evaluated SPICA™ on our in-house proprietary dataset consisting of 63 compounds with human expert-annotated patentability. SPICA™'s patentability classifier agrees with human experts on 61 of 63 molecules (96.8%) and demonstrated an AUROC of 0.99, showcasing exceptional accuracy. Moreover, SPICA™ was able to retrieve expert-identified prior patent via similarity search in 93% of cases, underscoring its robust patent search capabilities.

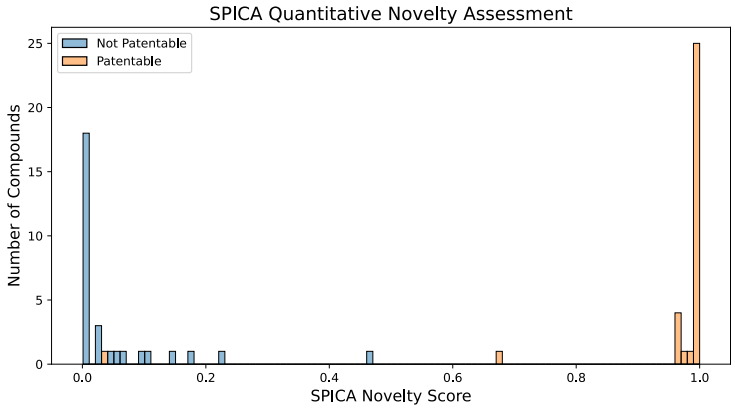

Figure 3. SPICA™ quantitative novelty assessment on
a proprietary in-house dataset of 63 compounds, annotated by human experts.

Case Study:
SPICA™ Novelty Score Analysis of FDA-Approved Drugs

The graphs below show the change in SPICA™ novelty score over the years for four FDA-approved drugs. The graphs reveal that these compounds maintain a high novelty score prior to their original patents. However, there is a noticeable decline in the novelty score following the patent applications. This trend demonstrates the reliability and accuracy of the novelty score generated by SPICA™.

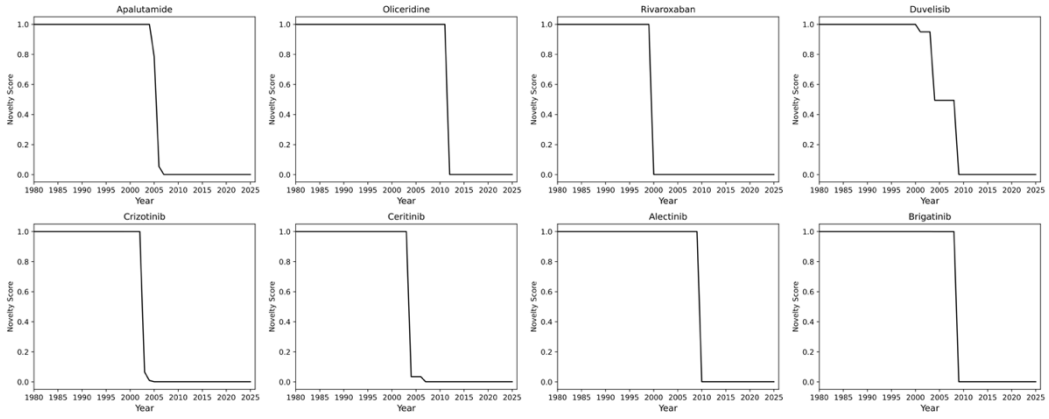

Figure 4. Example of SPICA™ Novelty scores for years of eight FDA approved drugs.

Conclusion

SPICA™ represents a significant advancement in the field of drug development, offering unparalleled speed and accuracy in patentability prediction. By integrating advanced machine learning techniques and a comprehensive patent database, SPICA™ empowers drug discovery scientists to make informed decisions, streamline the lead optimization process, and accelerate the path to market. As a result, SPICA™ not only enhances the efficiency of drug development but also supports the creation of innovative and commercially viable therapeutic solutions.

For more information and to explore how SPICA™ can enhance your drug discovery efforts, visit Standigm's SPICA™ page and Standigm's website.

Contact Us

For more information about Standigm and SPICA™ and how it can benefit you, please reach out to us:

- Email: business@standigm.com
- Website: www.standigm.com
- Latest SPICA™ White Paper: https://www.standigm.com/ai-saas/spica

Legal Information

The information contained in this white paper is for informational purposes only and does not constitute a commitment or guarantee of any kind. Standigm assumes no responsibility for errors or omissions.

About Standigm

Standigm is a Korean AI drug discovery company with over 10 years of experience, backed by SK, Pavilion Capital, and Kakao Ventures, with $70M+ in funding and partnerships with global pharma. Standigm's proprietary AI platforms—Standigm ASK™ for target identification and Standigm BEST™ for compound design—power continuous generation of commercially valuable pipelines.