Mini Review

NaturaPredicta™: AI-Enhanced Prediction Model of Botanical Ingredients

Sun Ho Kim*

SEAH bio solution, Seoul, Korea

*Corresponding author: Sun Ho Kim, PhD, SEAH bio solution, B618, 167, Songpa-daero, Songpa-gu, Seoul, 05855, Korea **Email: frank@seahbio.com**

Abstract

Traditional Korean medicine has a rich history of utilizing botanical ingredients for disease prevention and treatment. In this commentary, we provide an in-depth analysis of NaturaPredicta™, a groundbreaking model featured in our earlier article titled 'NaturaPredicta™: NLP-based Functional Scoring Method for Predicting the Bioactivity and Similarity of Botanical Ingredients [1]. Yet, the process of uncovering the bioactive compounds within these ingredients has traditionally been labor-intensive and time-consuming. Recent advancements in artificial intelligence (AI) are changing the game, making it possible to predict bioactivity and identify bioactive compounds in botanical ingredients using innovative in silico methods. NaturaPredicta™ leverages natural language processing (NLP) techniques based on deep learning algorithm applied to the vast PubMed® database, offering a unique approach that differs from conventional methods relying on molecular structure-based codes. This novel approach has the potential to revolutionize the field by reducing the time and cost required for traditional experimental methods, thus expediting the development of new Health Functional Food (HFF) ingredients.

Keywords: NaturaPredicta™; In Silico Prediction; Botanical Ingredient; Bioactivity Prediction; Artificial Intelligence (AI); Natural Language Processing (NLP); Deep Learning Algorithm; Health Functional Food (HFF)

Introduction

The global demand for Health Functional Food (HFF) has been on a steady rise, with South Korea emerging as a notable and prominent player in this market [2]. As individuals increasingly prioritize their health and well-being, the allure of natural products, especially those derived from botanical ingredients, continues to grow. These botanical ingredients, deeply embedded in Korean culture and traditional medicine, are not just part of the country's heritage but also essential components of the modern wellness industry. They are used in their natural forms and as derivatives, harnessing their potent and bioactive components. In recent years, scientific research has expanded our understanding of the medicinal properties of these botanical ingredients. Traditional knowledge is being validated and integrated into modern medicinal practices, particularly in the realm of preventive healthcare. Using natural products, especially botanical ingredients, for new HFF development is gain ing prominence and relevance. However, the road to discovering and characterizing the complex bioactive compounds within botanical ingredients is fraught with challenges. The intricate chemical structures of these compounds demand meticulous examination, resulting in labor-intensive and time-consuming research. Moreover, the vast number of different botanical ingredients used in traditional medicine adds a practical hurdle to comprehensive study.

The Role of AI in Botanical Ingredient Analysis

Enter the era of Artificial Intelligence (AI) and in silico methods, which are transforming drug discovery and biomedical research. AI-driven techniques, in particular, have paved the way for predicting bioactivity and identifying bioactive compounds within botanical ingredients with remarkable efficiency. Unlike traditional experimental methods, these computational approaches harness the

power of AI to predict molecular properties without the need for laborious wet laboratory work. This convergence of AI and in silico methods has proven instrumental in recognizing intricate patterns in vast datasets of chemical compounds, enabling precise predictions of bioactivity based on their chemical structures. AI-driven in silico methods offer several distinct advantages, including the ability to handle large datasets of chemical compounds, rapid screening, and efficient prediction capabilities [3]. One notable application of AI in this context is deep learning, specifically Convolutional Neural Networks (CNN), which has been used to classify and predict the biological activities of essential oil-producing plants based on their chemical compositions [4]. This method has shown promise in predicting the bioactivities of botanical ingredients, which can lead to groundbreaking developments in the field of HFF. Moreover, Natural Language Processing (NLP), a subset of AI, has found its way into healthcare and drug development, where it can effectively analyze and generate insights from unstructured text and speech data. NLP has been used in various stages of drug development, including target identification, clinical design, regulatory decision-making, and pharmacovigilance. These AI-powered language models have demonstrated their potential in accelerating drug discovery and development, including the search for new treatments for diseases like COVID-19.

Introducing NaturaPredicta™

In this landscape of AI-driven innovation, NaturaPredicta™ emerges as a groundbreaking model that harnesses NLP techniques to predict the bioactivity and similarity of botanical ingredients. What sets NaturaPredicta™ apart is its unique approach to data collection and analysis. Instead of relying on traditional molecular structure-based codes, it uses functional score probabilities derived from existing approved ingredients. This innovative method reduces the reliance on labor-intensive laboratory work and offers a more efficient and cost-effective way to predict bioactivity.

Methodology

The NaturaPredicta™ methodology comprises three major stages: data collection, NLP-based functional scoring, and functional score comparison.

Stage 1: Data Collection

The first stage involves an exhaustive search and compilation of botanical ingredients approved as HFF since the regulation was implemented. This search encompasses various types of ingredients, focusing on botanicals with individualized entries and substantial representation in PubMed®. Additionally, ingredients oriented toward health claims for chronic diseases are selected based on the Functional Food Material Information System (FFMIS) database [5].

Stage 2: NLP-based Functional Scoring

In the second stage, NaturaPredicta™ collects abstracts from scientific literature by searching for functional category keywords. Functional scores are then calculated using Bio BERT [6] a specialized version of BERT for bio text analysis. The model assigns scores to each functional attribute in a single abstract, determining wheth-

er a specific ingredient possesses these functionalities. The scores of individual functionals across all the abstracts are aggregated and converted into a unit vector for fair comparison across ingredients.

Stage 3: Functional Score Comparison

The third stage employs similarity scoring to compare functional scores, utilizing cosine similarity to score the similarities between HFF ingredients and target ingredients. This process identifies HFF ingredients that exhibit the highest cosine similarity to the target ingredient, thereby pinpointing candidate ingredients with similar functional patterns.

Results and Implications

NaturaPredicta™ marks a transformative epoch in health functional food (HFF) development, driven by its remarkable innovation—the curated HFFI dataset. This dataset serves as a comprehensive repository of bioactivity information, shedding light on the intricate world of bioactive compounds concealed within botanical ingredients. Within this expansive dataset, patterns and similarities emerge, offering insights into the complex interplay of bioactive compounds within these ingredients. Specifically, NaturaPredicta™ draws intriguing connections between specific target ingredients (TIs) and established HFF ingredients, such as the affinity between *Arachis hypogaea* (peanut) and *Zingiber officinale* (ginger) or *Curcuma longa* (turmeric). This revelation paves the way for further research and the development of innovative HFF ingredients. A deeper exploration of these TIs reveals an intricate network of interrelatedness with numerous HFF ingredients, spanning functions from "cholesterol lowering" to "cognitive function improvement" and "antioxidative" attributes. These multifaceted insights provide invaluable guidance for HFF industry researchers and developers, pointing them toward promising leads and bioactive compounds teeming with therapeutic potential. What sets NaturaPredicta™ apart is its seamless blend of tradition and technology, offering a streamlined, efficient, and cost-effective approach to unraveling the bioactivity hidden within botanical ingredients. In an era where the demand for health and wellness products continues to rise, NaturaPredicta™ stands as a symbol of innovation, poised to redefine the landscape of healthcare and well-being. By leveraging AI and NLP technologies, it empowers researchers and developers in the HFF industry to expedite the discovery of novel ingredients with specific therapeutic functionalities. Traditionally, unearthing bioactive compounds entailed labor intensive and time-consuming laboratory experiments, a process that was resource-intensive and costly. In stark contrast, NaturaPredicta™ automates data collection and analysis, significantly reducing the time and costs associated with traditional experimental methods. This acceleration aligns seamlessly with the surging demand for natural products that promote health and well-being, addressing the evolving preferences of consumers worldwide.

Conclusion

In the ever-evolving landscape of health and wellness, NaturaPredicta™ stands as a beacon of innovation, guiding us toward a healthier future. It represents the culmination of centuries-old

traditional wisdom and the cutting-edge technology of today, unlocking the vast potential of botanical ingredients in a manner previously inconceivable. As we delve deeper into this groundbreaking transformation of botanical medicine, it becomes clear that NaturaPredicta™ is not merely a tool; it is a herald of a new era in healthcare and well-being. At its core, NaturaPredicta™ embodies the fusion of two worlds: the rich heritage of traditional Korean medicine and the limitless possibilities of artificial intelligence (AI) and natural language processing (NLP). It is a testament to the adaptability of science and the resilience of tradition. In an era where technology often takes precedence over tradition, NaturaPredicta™ showcases how these two seemingly disparate realms can harmoniously coexist, opening new horizons for human health. One of the most striking aspects of NaturaPredicta™ is its ability to bridge the gap between ancient botanical wisdom and modern scientific rigor. For centuries, healers and herbalists have passed down knowledge of botanical ingredients, extolling their virtues for preventing and treating various ailments. While this traditional wisdom has undoubtedly been effective, it often lacked the empirical data and scientific validation demanded by contemporary medicine. NaturaPredicta™ serves as a bridge between these worlds, translating centuries of accumulated knowledge into quantifiable insights. The significance of this bridge cannot be overstated. It allows us to honor and preserve the cultural heritage embedded in botanical ingredients while subjecting them to the scrutiny of data-driven analysis. It harmonizes intuition and evidence, giving us a more comprehensive understanding of the potential therapeutic benefits of these natural wonders. By doing so, it ensures that traditional wisdom remains relevant and adaptable in an increasingly data-centric world. Furthermore, NaturaPredicta™ exemplifies the democratization of healthcare. It empowers researchers, developers, and healthcare professionals to explore the healing potential of botanical ingredients with unprecedented efficiency and accessibility. No longer bound by the constraints of time-consuming "wet laboratory" experiments, they can harness the power of AI and NLP to expedite discoveries that benefit us all. This democratization extends beyond professionals; it reaches the masses. It encourages individuals to take control of their wellbeing by making informed choices about the health functional foods they consume. As we reflect on NaturaPredicta™ and its transformative impact, we must also consider its global implications. The quest for health and well-being transcends borders and cultures. What begins as a pioneering effort in South Korea holds the promise of reaching every corner of the globe. The botanical ingredients that have been cherished in Korean culture for generations can now contribute to the health and happiness of people worldwide. This international exchange of knowledge and innovation reinforces the interconnectedness of our global community.

In closing, NaturaPredicta™ is not just a scientific breakthrough; it is a testament to the potential of human ingenuity when coupled with the wisdom of tradition. It reminds us that progress need not come at the expense of heritage, and that innovation can serve as a bridge between the past and the future. It beckons us to explore the uncharted territories of botanical medicine, where each ingredient is a treasure trove of healing potential waiting to be unlocked. The journey ahead is one of collaboration and exploration, where AI and NLP work hand in hand with traditional knowledge to uncover new insights and possibilities. It is a journey that promises a brighter, healthier future for all of humanity. NaturaPredicta™ is our guide on this path, illuminating the way forward with the promise of botanical ingredients and the transformative power of technology. As we embrace this new era of botanical medicine, we do so with hope and anticipation, knowing that the synergy of tradition and technology will lead to remarkable advancements in the field of natural medicine and nutrition, ultimately benefiting individuals and communities worldwide.

References

1. Kim SH, Choi D (2023) NaturaPredicta™ NLP-based Functional Scoring Method for Predicting the Bioactivity and Similarity of Botanical Ingredients. Food Suppl Biomater Health 3(3): e17.

2. Han K, Kim D, Lee K, Ji I (2023) Effect of Intake of Vitamin D on Health Status in Human. Food Suppl Biomater Health 3(2): e12.

3. Rifaioglu AS, Atas H, Martin MJ, Cetin-Atalay R, Atalay V, et al. (2019) Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bio informs 20(5): 1878–1912.

4. El-Attar NE, Hassan MK, Alghamdi OA, Awad WA (2020) Deep learning model for classification and bioactivity prediction of essential oil-producing plants from Egypt 10(1): 21349.

5. Functional Food Material Information System.

6. Lee J, Yoon W, Kim S, Kim D, Kim S, et al. (2020) Bio BERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36(4): 1234-1240.