

Review

Not peer-reviewed version

Symbiosis in Health: The Powerful Alliance of AI and Propensity Score Matching in Real World Medical Data Analysis

[Peter Kokol](#)*, [Bojan Žlahtič](#), [Helena Blažun Vošner](#), [Jernej Završnik](#), Tadej Završnik

Posted Date: 8 December 2025

doi: 10.20944/preprints202512.0312.v1

Keywords: propensity score matching; artificial intelligence; real world medical data; synthetic thematic analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Symbiosis in Health: The Powerful Alliance of AI and Propensity Score Matching in Real World Medical Data Analysis

Peter Kokol ^{1,2,*}, Bojan Žlahtič ², Helena Blažun Vošner ^{1,3}, Jernej Završnik ^{1,3} and Tadej Završnik ⁴

¹ Community Healthcare Center Dr. Adolf Drolc Maribor, 2000 Maribor, Slovenia

² Faculty of Electrical Engineering and Computer Sciences, University of Maribor, 2000 Maribor, Slovenia

³ Alma Mater Europaea University, ECM Maribor, 2000 Maribor, Slovenia

⁴ University Clinical Center Maribor, 2000 Maribo, Slovenia

* Correspondence: peter.kokol@um.si

Abstract

Background: The rapid expansion of real-world data in medicine is driving the adoption of advanced methods like Artificial Intelligence (AI) and Propensity Score Matching (PSM). AI is widely applied across diagnostics, prediction, and treatment planning, while PSM is a crucial statistical technique used in quasi-experimental studies to mitigate confounding bias and approximate the reliability of randomized controlled trials. There is a growing research interest in combining these two methods to leverage their symbiotic strengths, but this association has not been holistically explored.

Methodology: This study employed Synthetic Thematic Analysis (STA), derived from synthetic knowledge synthesis, to systematically review the existing literature on AI and PSM in medicine. Publications were harvested from the Scopus database using a comprehensive search string limited to the Medical subject area. The resulting corpus (N=433 documents) was analyzed using bibliometric tools (Bibliometrix and VOSViewer) to map the research landscape, identify thematic clusters based on author keywords, analyze collaboration patterns, and synthesize findings from highly prolific publications. **Results:** The field is young and rapidly accelerating, showing an exponential increase from 2020 to 2024. China and the USA dominate research production and citation impact. The symbiotic relationship is published in high-impact medical journals and health informatics journals. STA identified four main thematic clusters: Prediction, Cancer Management, Diagnosing, and Deep Learning. AI and PSM are combined in two primary ways: AI used in PSM and PSM used in AI.

Conclusion: The symbiotic association between AI and PSM is a global and rapidly developing trend in medical research, driven by major international contributors. This convergence is enhancing methodological rigor in observational studies, primarily by improving prediction models and refining causal inference in complex areas like cardiovascular disease, cancer, and diagnostics.

Keywords: propensity score matching; artificial intelligence; real world medical data; synthetic thematic analysis

1. Introduction

The rapid expansion of real-world data and evidence [1–3] has driven the adoption of advanced technologies and methods in medicine, including artificial intelligence (AI) and propensity score matching (PSM). AI is applied in various medical fields, such as diagnostics, prediction, imaging and pattern recognition, risk assessment, robotics, education, treatment planning and many others [4–7,7–14]. In contrast, PSM is primarily a statistical methodology with a more specific purpose. It is widely used in quasi-experimental studies, such as retrospective analyses of medical claims, data from disease or product registries, digital health technologies, routine healthcare datasets,

observational studies, and electronic medical records. PSM leverages individual patient covariates to balance potential confounding factors when comparing different patient groups. Adjusting for baseline disparities enables analyses to approximate the reliability of randomized, prospective studies, which are often considered the gold standard in research [15–21].

The use of PSM and AI in data analysis has garnered significant research interest, evolving from traditional methods with baseline use of PSM to more advanced enhanced AI techniques and combined hybrid approaches and integrations [22,23]. Researchers typically begin with conventional propensity score matching to construct a methodological baseline, as this established technique provides fundamental insights into the data structure before applying alternative approaches [24,25]. However, there is growing interest in using AI and machine learning methods to enhance propensity score estimation (the use of neural networks, decision trees, using AI to automate processes of variable selection and model specifications relied on PSM), especially when dealing with high-dimensional data or complex relationships [26,27].

When thus combined, AI and PSM enhance medical research in a symbiotic way. First, PSM improves AI applications by increasing the reliability and efficiency of machine learning algorithms [28–31]. Second, integrating artificial intelligence (AI) with propensity score matching (PSM) presents a promising avenue to overcome limitations of traditional PSM and broaden its utility in medical and healthcare research. AI can augment multiple stages of the PSM workflow—including data preprocessing (e.g., natural language, image, and signal processing [32], propensity score estimation, covariate selection, and post-matching analysis. By mitigating biases, refining causal inference, and fostering more robust and interpretable models [33], AI-enhanced PSM holds substantial potential for advancing methodological rigor in observational studies [34].

The above symbiotic association between AI and PSM has not yet been thoroughly and holistically investigated. To close this gap, Synthetic thematic analysis (STA), derived from the synthetic knowledge synthesis [35,36], was used in this study to review for the first time the scope and content of the existing research literature and answer the main research question, namely How and in which context do AI and PSM complement each other? Furthermore, the main research question has been further decomposed into following more specific research question:

1. What are the dynamic and spatial features of the research literature production of the AI and PSM use in medicine?
2. How is the symbiosis association between AI and PSM reflected in most prolific source titles and most productive countries?
3. What research themes emerge in studies combining AI and PSM for medical data analysis?
4. What are the more prolific AI methods, medical applications, and diagnoses in combined AI and PSM analyses?
5. What are the dominant research trends in the combined use of PSM and AI?

2. Methodology

Synthetic thematic analysis (STA) in this study was performed with the algorithm shown in Figure 1:

1. Harvest the research publications corpus in the Scopus bibliographic database using the search string:

TITLE-ABS-KEY(("artificial intelligence" OR "machine learning" OR "deep learning" OR "intelligent system" OR "support vector machine" OR ("decision tree" AND (induction OR

heuristic)) OR "random forest" OR "Markov decision process" OR "hidden Markov model" OR "fuzzy logic" OR "k-nearest neighbours" OR "naive Bayes" OR "Bayesian learning" OR "artificial neural network" OR "convolutional neural network" OR "recurrent neural network" OR "generative adversarial network" OR "deep belief network" OR "perceptron" OR {natural language processing} OR {natural language understanding} OR {general language model}) AND "propensity score matching") AND (LIMIT-TO (SUBJAREA, "MEDI"))

2. Condense the content of the corpus using bibliometric mapping and induce the AI - PSM author's keywords research landscape employing VOSViewer software [37].

3. Analyse the connections between the codes in individual clusters and map them into prolific author's keywords association sub-networks

4. Use author keywords from sub-networks as parts of the search strings to identify prolific publications for each sub-network.

5. Analyse publications and categorize them into one of the following two categories

- a. AI use in PSM
 - b. PSM usage to improve AI algorithms
-

5. Synthesize the most prominent publications

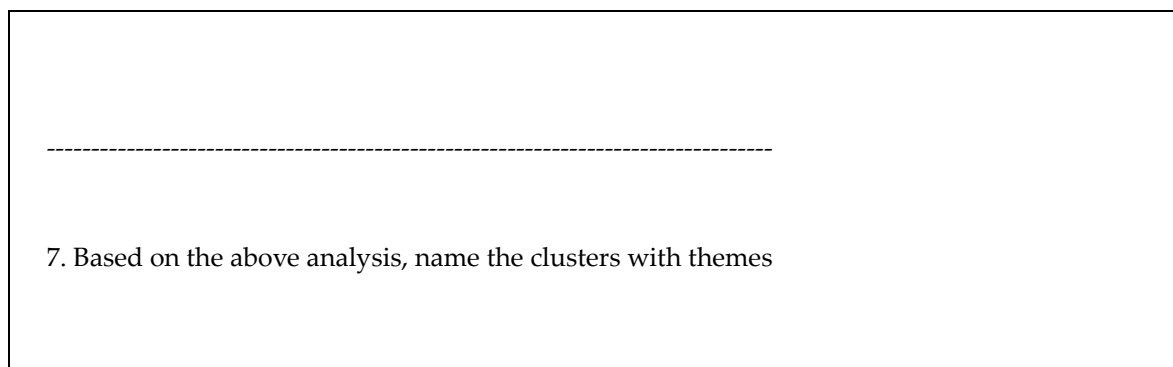


Figure 1. Algorithm 1 for determining prolific themes.

1. Use the corpus of publications harvested with Algorithm 1
2. Analyse the content of the corpus using R software to induce Topic map
3. Use author keywords from nodes as parts of the search strings to identify prolific publications for each
4. Synthetise the most prominent publications for each node
5. Name niche, emerging, declining, motor and basic nodes with themes

Figure 2. The algorithm for determining niche, emerging, declining, motor, and basic themes.

To obtain the necessary bibliometric data for extensive bibliometric study, Scopus was used due to its wider interdisciplinary coverage, inclusion of PubMed, advanced search options including AI, and the ability to export larger amounts of data for bibliometric analysis in one run. For the bibliometric analysis, two main bibliometric tools were used, namely Bibliometrix version 5.1.1 and VosViewer version 1.6.20. [38,39]. Bibliometrix was run in RStudio version 2025.09.1+401, which was utilising the R programming language version R-4.5.1 for Windows. VosViewer was run under the latest version of the programming language, Java 8 Update 461, using the Java Runtime Environment (JRE) 1.8.0_461. VosViewer was used to create the bibliometric and thematic mapping based on the author keywords found in the bibliometric data, while the Bibliometric library was used to perform the remaining part of the bibliometric analysis, including base bibliometric information gathering, publication overview, citation analysis, countries production, Lotka's Law, and Bradford's Law. Bibliometrix was run and utilized using the biblioshiny method, which allows for interactive web-based engagement.

3. Results

Dynamic and spatial features of the research literature production

The search based on the provided search string in Scopus was performed end of September 2025. The search yielded 433 documents from 283 sources that were authored by 3858 authors. Of the 433 documents, 415 were articles, one book, 11 conference papers, one conference review, and 4 reviews. The results encompass documents from 2011 until 2025 with no papers scheduled for 2026. The average citation was 10.73 per document, and the average document age was 2.02, indicating a relatively young and interesting field of study. The average number of co-authors per document is 21, and the international co-authorship is 18.94 indicating global interest and collaboration.

In Figure 3, there is a noticeable spike in the number of published papers over the last few years, indicating a heightened interest in the observed field from 2020 to 2024. There is a small dip in 2025. This is to be expected, as the results encompass data up to the end of September, which is still several months away from the end of the year. It's important to note that while the dip in the graph may suggest a larger disparity, the actual difference is just one paper. Given the remaining time for publication, the total number of publications is expected to far exceed those published in 2024. While the citation, as seen in Figure 4, declines over time, this is natural since citation is accumulated through time, and as the number of publications rises, the average number of citations decreases. However, there were 3 noticeable spikes in the average annual number of citations in the years 2011, 2014, and 2018. Those dates correspond to the top 3 most cited documents in the field [40–42].

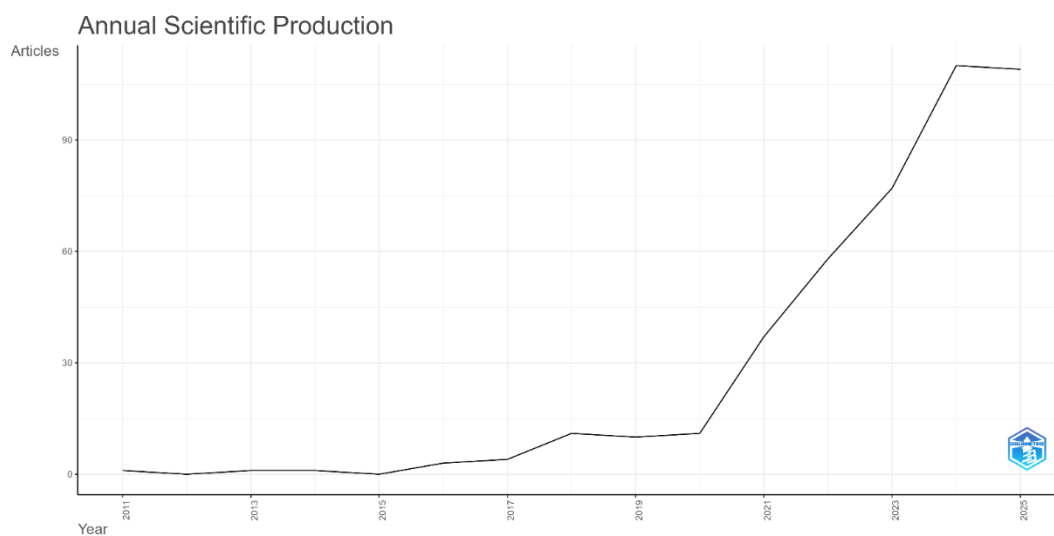


Figure 3. Number of publications over time.

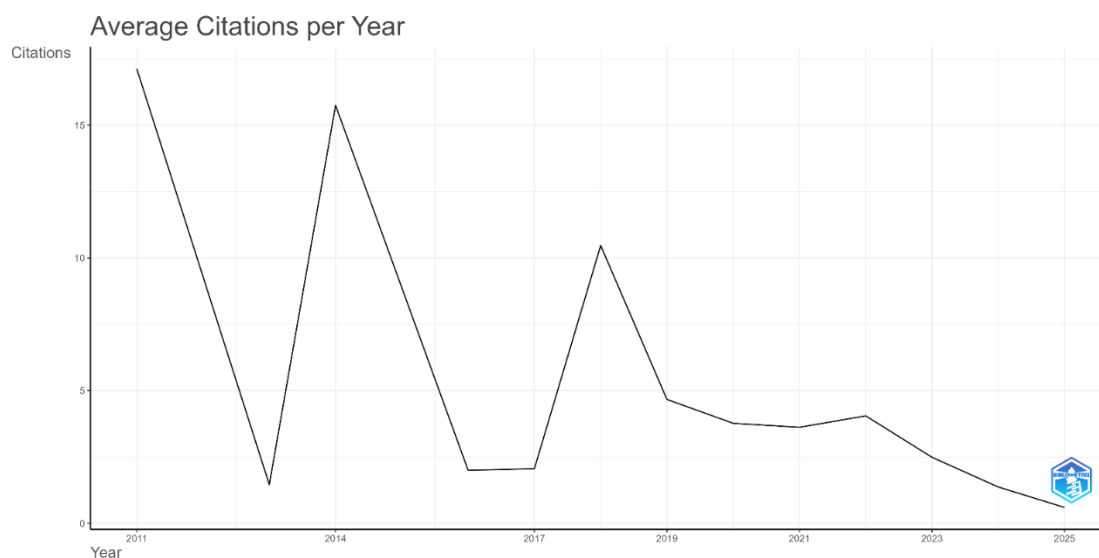


Figure 4. Average number of citations over time.

Observing the most relevant authors through Lotka's Law, there is a classical and clear power distribution with a very small number of authors present in the scientific space that contributed several papers. The top five contributors, Li C. (5 articles), Liu M. (4 articles), Wang Y. (4 articles), Qu J. (4 articles), and Zhang Y. (4 articles), all started publishing in this particular field of study in 2021 or later. All of them also contributed in 2025. When looking at the authors' local impact based on the

h-index and additionally taking into account their relevance, Li C. is the most prominent contributor, tightly followed by the previously mentioned remaining four contributors. Fujian Medical University (54 articles), Guangxi Medical University (54 articles), The Second Hospital of Xian Jiaotong University (46 articles), Capital Medical University (45 articles), Tongji Medical College of Huazhong University of Science and Technology (39 articles). The production over time for the outlined affiliation is starting to grow from 2021 and is noticeably growing each year.

Productive countries and source titles

Looking at Figure 5, it is clear that China and the United States of America are dominating in this field of study, with the former far outcompeting the other areas and the latter having a far more international presence in relation to cumulative production. Also noteworthy is that Germany has far more Multiple Country Publications (MCP) than Single Country Publications (SCP). Looking at the countries' production over time, there was very little publication from 2011 until 2020. After 2020 the production in China and the United States of America increased exponentially. Looking at the citations based on the country The United States of America and China again dominate the research field and are followed by the United Kingdom and Canada.

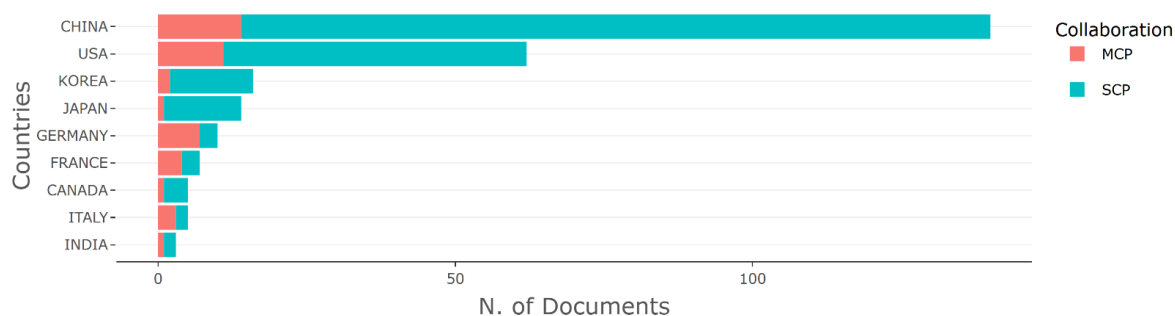


Figure 5. Most productive corresponding author's countries in the form of cumulative, international, and domestic publications.

Table 2 shows that the rank of countries' productivity in PSM-related research corresponds well with the rankings of countries in overall productivity across all disciplines, as well as in medicine, AI, medicine, and Statistics and probability. That might indicate that systemic national factors—like consistent R&D funding, shared infrastructure, high-quality education, research capacity and supportive policies—uniformly drive research productivity across academic disciplines/topics shown in Table 1. It might also signal the co-operation in symbiotic multidisciplinary research across medicine, AI and statistic focusing on PSM.

Table 1. Most productive countries research rankings according to Scimago [43].

Country	Rank all disciplines	Rank in medicine	Rank in artificial intelligence	Rank in Statistics and probability
China	2	2	1	2
United states	1	1	2	1
South Korea	13	14	12	16
Japan	5	5	4	10

health informatics journals, this distribution might suggest a structural interdependence between medical and informatics research, highlighting the possible synergistic nature of AI and PSM scholarship, wherein AI methodologies and clinical applications converge to advanced and more innovative health care.

Table 2. The bibliometric profile of journals publishing more than four or more papers.

Journal name	SNIP	Quarter	Research area
Frontiers in Oncology	0.831	2.	Cancer Research, Oncology
Frontiers in Cardiovascular Medicine	0.742	2.	Cardiology and Cardiovascular Medicine
BMC Infectious Diseases	1.106	1.	Infectious Diseases
Frontiers in Public Health	0.938	2.	Public Health, Environmental and Occupational Health
Journal of Clinical Medicine	1.022	1.	Medicine (all)
JMIR Medical Informatics	1.035	2.	Health Information Management Health Informatics
Cancers	1.030	2.	Cancer Research Oncology
Frontiers in Pharmacology	0.999	1.	Pharmacology Pharmacology (medical)
BMC Public Health	1.386	1.	Public Health, Environmental and Occupational Health
European Radiology	1.775	1.	Radiology, Nuclear Medicine and Imaging
Frontiers in Endocrinology	1.122	2.	Endocrinology, Diabetes and Metabolism
Frontiers in Medicine	0.879	1.	Medicine (all)

More prolific themes

The results of the STA are presented in Figure 6 and Tables 3–4. Following Zipf's Law [34], bibliometric mapping was applied to all keywords appearing in four or more publications, yielding a landscape of 45 author keywords grouped into four distinct clusters (Figure 1). Implementation of steps 2–6 in Algorithm 1 generated four thematic categories, 12 association sub-networks, and 28 high-impact publications (Table 2). A synthesis of these 28 publications is provided in Table 2.

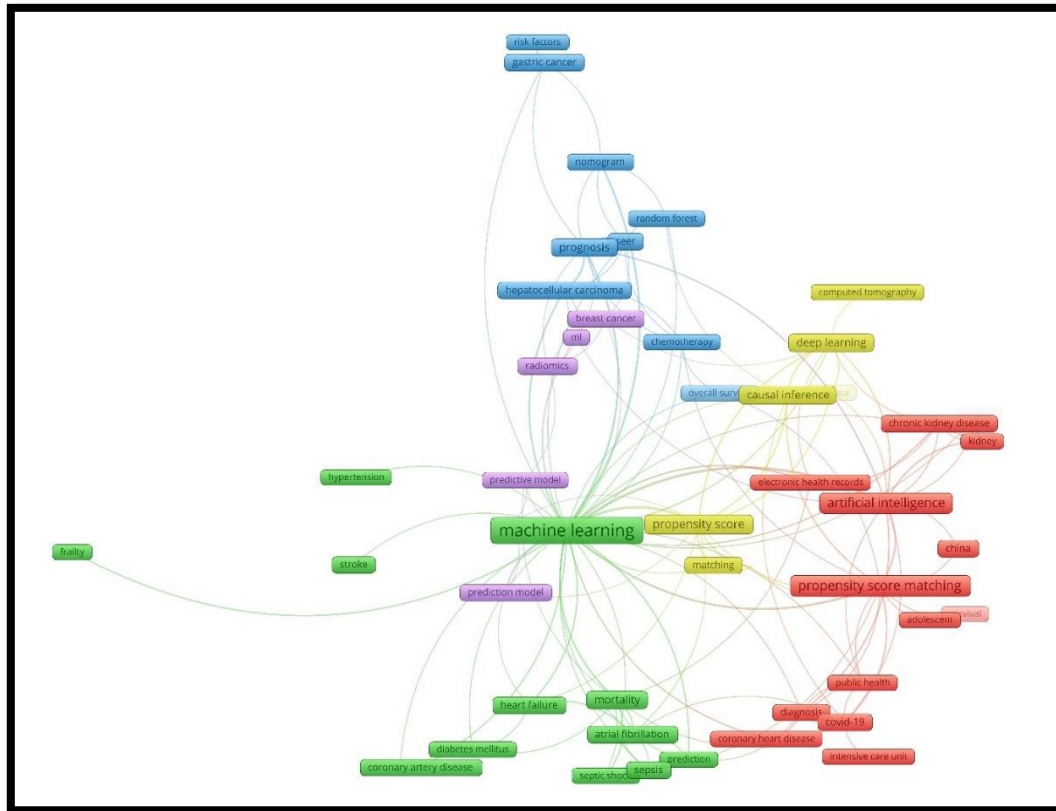


Figure 6. The authors keyword network.

Table 3. Theme, subthemes, and prolific publications.

Theme	Prolific author's keywords association sub-networks	Publications describing AI use in combination with PSM	Publications describing PSM use in AI
Prediction	Cardiovascular diseases – Diabetes mellitus	[45], [46–48]	[50,51]
Blue (14 author keywords)	Atrial fibrillation – sepsis-prediction	[49]	[52,53]
Cancer management	Breast cancer – SEER	[54]	[57,58]

Red (15 author keywords)	<p>Hepatocellular carcinoma – SEER – chemotherapy-survival</p> <p>Gastric cancer – random forest</p> <p>Natural language model, prediction modelling</p>	[55,56]	[59] [60,61] [62,63]
Diagnosing Green (14 author keywords)	<p>Coronary heart diseases – diagnosis</p> <p>Diagnosis – Intensive care unit – Public health</p> <p>Chronic kidney disease – Electronic health record</p>	[64] [65,66] [67]	[68,69] [70] [71]
Deep learning Yellow (7 author keywords)	<p>Casual inference – Big data – deep learning</p> <p>Monte Carlo simulation</p> <p>Computer tomography – deep learning</p>	[72–74] [75]	[76] [77]

Table 4.

Theme	Association sub-networks	Synthesis of publications
Prediction	Cardiovascular diseases – Diabetes mellitus-	<p>Binomial regression models and random forest regression was performed on a dataset of high risk COVID-19 subjects (inclusion criteria: age over 65 years old, presence of solid or haematological cancer, chronic kidney disease, chronic liver disease, chronic lung disease, uncontrolled diabetes, neurological disease, cardiovascular disease, obesity, cerebrovascular disease or being immunocompromised (AIDS, solid organ or blood stem cell transplantation, and all conditions requiring use of corticosteroids or other immunosuppressive medications)) after performing PSM based on being early treated or not [45].</p> <p>Lim et al. published a study in 2023 conducted a nationwide population-based cohort study comparing cardiovascular outcomes between new and existing users of dapagliflozin and empagliflozin in type 2 diabetes patients. Using a Korean cohort dataset, the authors employed a nearest-neighbours machine learning approach for propensity score matching prior to statistical analysis [47].</p> <p>The LIGHTNING study modelled, predicted, and compared hypoglycaemia rates of people with type 2 diabetes, comparing patients using first or second-generation insulin preparations. During analysis, authors first used conventional (PSM) and then advanced machine learning [50].</p> <p>A large-scale Indian patient database was analysed using the Spearman correlation coefficient method and Deep Learning to build a hazard model, which was used to predict CVD events and their time of occurrence that reportedly had a</p>

		<p>good performance rate. PSM was used first to match patients with and without CVD [51].</p> <p>Xie et al. investigated the utility of coronary computed tomography angiography (CCTA) in detecting cancer treatment-related coronary artery impairments and predicting major adverse cardiovascular events (MACE) in lung cancer patients undergoing chemotherapy or chemoradiotherapy. Their methodology involved: (1) AI-driven image recognition for initial assessment, (2) propensity score (PS) matching to compare patients with and without carcinoma, and (3) Cox regression modelling to evaluate differences in MACE-free survival rates. [46].</p> <p>Squicciarro et al [78] performed a retrospective cohort study (N=1,908) examining systemic inflammatory response syndrome (SIRS) impact on 30-day mortality post-cardiac surgery and developed predictive machine learning models. PSM was used to balance the training set. SIRS significantly increased mortality risk; models achieved AUC up to 0.82.</p>
	Atrial fibrillation – sepsis-prediction	<p>In a study by Wang et al. [52], authors developed a model to predict the risk of mortality in septic patients with atrial fibrillation using different ML algorithms. They used PSM to reduce the imbalance between the external validation and internal validation data sets.</p> <p>In another study, Ruan et al. [53] used five different ML algorithms to determine whether dyslipidaemia or obesity contributes more towards unfavourable clinical outcomes in patients suffering a first-ever ischemic stroke. PMS was employed to ascertain associations between indicators and prognosis.</p>

		<p>The study applied propensity score matching and a causal machine learning framework to predict heterogeneous treatment effects of LAAO versus DOAC in atrial fibrillation patients, enabling AI-driven individualized benefit estimation for improved patient selection and clinical decision-making [49].</p>
--	--	---

Cancer management	Breast cancer – SEER (Surveillance, Epidemiology, and End Results) database	<p>A study used the SEER database to identify the prognostic variables for patients with occult breast cancer, which is an uncommon malignant tumour for which the prognosis and treatment remain a controversial topic. Cox regression analysis was performed to construct prognostic models with the help of six machine-learning algorithms to predict overall survival. The authors further examined the impact of chemotherapy and surgery on survival outcomes in occult breast cancer patients stratified by molecular subtype, utilizing Kaplan-Meier survival analysis and propensity score matching. These findings were subsequently validated through subgroup Cox regression analysis [57].</p> <p>In a similar study, South Korean investigators used machine learning-based risk factor detection and breast cancer mortality prediction with the Shapley Additive Explanation (SHAP), which is an explainable artificial intelligence technique, to identify and interpret key features that have a significant impact on breast cancer mortality. To enhance the robustness and generalizability of their primary findings and balance the baseline covariates, they employed an exposure-driven 1:3 propensity score matching (PSM) analysis while minimizing a logistic regression model with the implications of potential confounders [58].</p> <p>Liang et al [54] analyzed 18,726 NHANES participants to examine breast cancer prevalence and neutrophil-percentage-to-</p>
-------------------	---	--

		albumin ratio (NPAR), revealing a significant positive association, potentially mediated by sex hormone levels, validated through advance multivariate, subgroup, and propensity score analyses.
	Hepatocellular carcinoma – SEER (Surveillance, Epidemiology, and End Results) database – chemotherapy-survival	<p>Patients diagnosed with hepatocellular carcinoma between January 2010 and December 2015 were identified through the SEER (Surveillance, Epidemiology, and End Results) database. The researchers first conducted univariate and multivariate logistic regression analyses to assess prognostic factors, then developed a 5-year survival risk prediction model using classical decision tree methodology. To address potential confounding variables related to chemotherapy administration, propensity score matching was implemented for both high-risk and low-risk patient cohorts [59].</p>
	Gastric cancer – random forests	<p>According to Huang C. et al., clinical data from 391 gastric cancer patients (including 86 peritoneal dissemination cases) were analyzed using a 1:3 propensity score matching approach. The researchers subsequently performed both univariate and multivariate conditional logistic regression analyses. Their methodology further incorporated classification tree analysis to establish decision rules, followed by random forest algorithm implementation to extract significant risk factors for peritoneal dissemination in gastric cancer [47].</p> <p>Another study aimed to explore the association of the tongue coating microbiota with the serum metabolic features and inflammatory cytokines in GC patients to seek a potential non-invasive biomarker for diagnosing GC. The tongue coating microbiota was profiled by 16S rRNA and 18S rRNA genes sequencing technology in the original population with 181 GC patients and 112 healthy controls (HCs). The propensity score matching method was used to eliminate potential confounders, including age, gender, and six lifestyle factors, and a matching</p>

		<p>population was created. For the diagnosis of GC a random forest model was constructed.</p>
	<p>Natural language model prediction modelling</p>	<p>As reported by Gao Z. et al, an innovative integrated strategy was developed to identify FDA-approved drugs for repurposing in cocaine use disorder (CUD) treatment. The study combined AI-driven drug prediction with clinical validation through the National Drug Abuse Treatment Clinical Trials Network (CTN), incorporating expert panel review and mechanistic action analysis. Based on combined AI prioritization and clinical expertise, ketamine emerged as the top candidate for further evaluation. The team conducted electronic health record (EHR) analysis comparing CUD outcomes in patients prescribed ketamine (for anesthesia/depression) against propensity-matched controls receiving alternative treatments. Complementary genetic and pathway enrichment analyses were performed to elucidate ketamine's potential mechanisms of action in CUD [55].</p> <p>In another study, PSM balanced the covariates across two groups of Alzheimer's disease and related dementia patients with oropharyngeal during hospitalization, whether at least 75% of their hospital diet consisted of a thick liquid diet or a thin liquid diet. Machine learning was used to predict hospital outcomes such as mortality, length of stay, and complications [62].</p> <p>Varun et al. analyzed data from 38,581 shoulder and hip replacement patients (2015-2021) to develop a random forest model predicting 30-day post-discharge outcomes (emergency department visits, unplanned readmissions, or discharge to skilled nursing facilities). The study incorporated 98 features spanning laboratory results, diagnoses, vital signs, medications, and utilization history. Notably, the researchers employed a Clinical BERT-finetuned NLP model to generate risk scores from clinical notes. To address potential biases, the methodology combined propensity score matching with comprehensive feature bias analysis, implementing Fairlearn toolkit's threshold</p>

		<p>optimization to mitigate gender and payer-related prediction disparities [79].</p> <p>Krishna et al. [56] performed natural language processing on clinical records of patients from the Veterans Health Administration database, which were antiarrhythmic drug-naive, to identify and compare baseline left ventricular ejection fraction between treatments with different drugs. They used 1:1 propensity score matching based on patient demographics, comorbidities, and medications, as well as Cox regression to compare strategies. A falsification analysis with non-plausible outcomes was performed to evaluate residual confounding.</p>
--	--	--

Diagnosing	Coronary heart diseases-diagnosis	<p>While coronary artery calcium (CAC) is an established predictor of cardiovascular disease (CVD), optimal screening strategies require identification of populations deriving maximal benefit from CAC detection. Kosuke et al. examined whether CAC's predictive value varies across demographic subgroups in the Multi-Ethnic Study of Atherosclerosis (MESA) cohort (ages ≥ 45, CVD-free at baseline). After 1:1 propensity score matching, the team employed machine learning causal forest modelling to: (1) quantify heterogeneity in CAC-CVD associations, and (2) predict individualized 10-year CVD risk increases when $CAC > 0$ versus $CAC = 0$. These machine learning estimates were subsequently benchmarked against absolute 10-year ASCVD risks calculated via 2013 ACC/AHA pooled cohort equations [68].</p> <p>A recent study explored whether gingival bleeding - a simple clinical indicator of periodontal disease - might serve as a marker for hypertension. Given the established link between cardiovascular diseases and systemic inflammation, with periodontitis potentially exacerbating this inflammatory burden, researchers analyzed NHANES III data from 5,396 adults aged ≥ 30 years who completed both blood pressure assessments and periodontal exams. Using survey-based propensity score matching that accounted for key confounders shared by hypertension and periodontal disease, they created matched cohorts with and without gingival bleeding. The analysis employed generalized additive models adjusted for inflammatory markers to evaluate associations between bleeding gums and both systolic blood pressure (mmHg) and uncontrolled hypertension. Further stratification by periodontal status (healthy, gingivitis, stable periodontitis, unstable periodontitis) provided additional insights, while machine</p>
------------	-----------------------------------	---

		<p>learning techniques helped determine variable importance in these relationships [69].</p> <p>Samuel et al. developed a computationally efficient algorithm that properly characterizes and samples from the conditional distribution of treatment following optimal propensity score matching, while accounting for Z-dependence. This innovation addresses a fundamental methodological challenge: unlike traditional matched-pair designs where pairs are fixed beforehand, propensity score matches are constructed post-treatment based on observed status. Consequently, standard permutation-based inference methods become invalid since treatment permutations could yield entirely different matched sets—a dependency (Z-dependence) that conventional approaches fail to consider [64].</p>
--	--	---

Diagnosing	Diagnosis – Intensive care unit- Public health	<p>Feller et al. [65] evaluated the added value of natural language processing (NLP) for enhancing HIV diagnosis prediction models. Their study included 181 HIV-positive patients receiving care at New York Presbyterian Hospital prior to confirmed diagnosis, along with 543 propensity-matched HIV-negative controls. Researchers extracted structured EHR data (demographics, laboratory results, diagnosis codes) and unstructured clinical notes from the pre-diagnosis period. They then developed three machine learning models: (1) a baseline model using only structured EHR data, (2) baseline plus NLP-derived topics, and (3) baseline plus NLP-extracted clinical keywords. Results demonstrated that incorporating NLP features significantly improved predictive accuracy for HIV risk assessment</p> <p>In a review paper, Zaccali and Tripepi [66] claim that trial emulation with PMS use in observational studies represents a significant advancement in epidemiology and can support improving public health outcomes. However, traditional PSM techniques face challenges like data quality, unmeasured confounding, and implementation complexity that could be overcome with machine learning techniques and developing methods to address unmeasured confounding.</p> <p>In a recent study, Fu et al. [70] collected information from selected participants before surgery and followed up until the day after surgery, then divided them into a normal and delayed ambulation group. Propensity score matching was applied to all participants by type of surgery and anaesthesia. All the characteristics in the two groups were compared using logistic regression, back propagation neural network (BPNN), and decision tree models. The accuracy, sensitivity, specificity, and area under the curve (AUC) values of the three models were compared to determine the optimal model.</p>
------------	--	---

	Chronic kidney disease – Electronic health record	<p>Krishnamurthy et al. [52] developed a machine-learning model to predict incidents of chronic kidney disease (CKD) 6-12 months before clinical onset using Taiwan's National Health Insurance claims data. The study employed propensity score matching to select 18,000 CKD cases and 72,000 matched controls, analysing two years of demographic, medication, and comorbidity history for each subject. Among various algorithms tested, convolutional neural networks (CNNs) demonstrated superior predictive performance. Tree-based feature importance analysis identified diabetes mellitus, advanced age, gout, and specific medication use (particularly sulphonamides and RAAS inhibitors) as the strongest predictors of CKD development.</p> <p>Using AI phenotyping and propensity score matching on 168,860 Veterans with heart failure, balanced on 77 covariates, high-dose RAS inhibitors showed lower kidney failure risk versus low-dose [67].</p>
--	---	--

Deep Learning	Casual inference – deep learning	<p>Cui et al. [76] conducted a large-scale analysis of ICU patients without primary gastrointestinal diseases using the MIMIC-IV database to evaluate the prognostic value of abdominal physical examinations (palpation and auscultation). Patients were stratified based on examination status, with 28-day mortality as the primary endpoint. The researchers employed multiple analytical approaches: Cox proportional hazards models, propensity score matching, and inverse probability treatment weighting. Following initial analysis, the examined cohort was randomly split into training (80%) and testing (20%) sets, while patients with primary GI conditions served as an external validation group. Six machine learning algorithms—Random Forest, Gradient Boosting Decision Trees, AdaBoost, Extra Trees, Bagging, and Multilayer Perceptron—were subsequently implemented to develop predictive models for in-hospital mortality.</p> <p>To optimally evaluate the relationship between participation in different types of social activities and depression in the elderly, Wang et al. [72] used propensity score matching (PSM) for analysis based on the counterfactual framework. The specific matching methods used were the k-nearest neighbours matching method, kernel matching method, and radius matching method.</p> <p>To reduce the underlying bias in observational studies, Ghosh et al. [75] developed a new deep learning architecture for propensity score matching and counterfactual prediction.</p> <p>Machine learning enhanced propensity score estimation by improving covariate balance, reducing bias in observational studies, and enabling robust causal inference, thereby advancing methodological rigor in treatment effect in patients with non-small lung cancer, by analysis across</p>
---------------	----------------------------------	---

		complex, high-dimensional healthcare and social science datasets [74]
	Monte Carlo simulation	Weyman et al. [75] addressed limitations of manual longitudinal propensity score matching by developing a machine learning-enhanced genetic matching approach that automatically optimizes covariate history balancing. Through Monte Carlo simulation studies, the authors demonstrated superior performance of their automated method compared to traditional manual matching techniques.
	Computer tomography – deep learning	Chen et al. [80] investigated radiomics and deep learning approaches for predicting sepsis risk following stone removal procedures (FURL/PCNL) in ureteral calculus patients. After propensity score matching, they developed: (1) a radiomics model for sepsis prediction, and (2) an enhanced deep learning model to boost predictive accuracy. LASSO regression identified 26 key predictive variables. The deep neural network (DNN) implementation showed improved AUC in internal validation, with subsequent external validation confirming model generalizability by addressing overfitting concerns.

The results of the deductive thematic analysis based on the authors' keywords are shown in Table 3. As we can see, the analysis did not reveal in detail which machine learning methods have actually been used because authors mainly used the generic keyword Machine learning. Consequently, we performed an additional analysis of terms found in publication abstracts. This analysis revealed that Boosting (n=30), Support vector machines (n=20), and Nearest neighbours (n=20). Random forests (n=18) and Decision trees (n=15) were the most popular methods. In addition, SHAP was used in 25 publications to make black box machine learning outputs explainable. The most popular applications were prognosis, prediction, decision support, and risk assessment. The most popular diagnoses were related to chronic diseases and cancer.

Table 3. Deductive thematic analysis - Ten most prolific AI algorithms and methods, medical applications, and diagnosis.

AI algorithms and services	n	Medical applications	n	Diagnosis	N
Machine learning	66	Prognosis	11	heart diseases	12
Artificial intelligence	23	Mortality	10	atrial fibrillation	6
Deep learning	11	Nomogram	5	covid-19	6
Decision tree	5	Prediction model	11	gastric cancer	6
Random forest	6	Diagnosis	4	hepatocellular carcinoma	6
Natural language processing	5	Decision support	6	Stroke	3
Big data	3	Risk assessment	14	Kidney diseases	11
SHAP/Explainable AI	2	epidemiology	5	Diabetes	9
Missing data imputation	1	Survival analysis	11	Cardiovascular diseases	18
Feature selection	2	Health services	6	Coronary diseases	7

Prolific term and topic trend analysis

Focusing on the terms, there is a clear trend developing where AI, machine learning, and deep learning are being introduced and considered for application since 2022, as seen in Figure 7. The symbiosis seems to evolve first from association of PSM and machine learning, through introduction of AI and deep learning and finally of concrete applications of both approaches in concrete medical application (cancer and sepsis).

Additional insight and confirmation of the symbiotic association of PDA and AI, can be obtained from the thematic map presented in Figure 8, where symbiosis of AI and PSM in chronic diseases presents as a motor theme in the upper right quadrant, while also being present in the lower left quadrant, as an emerging approach in various medical subfields..

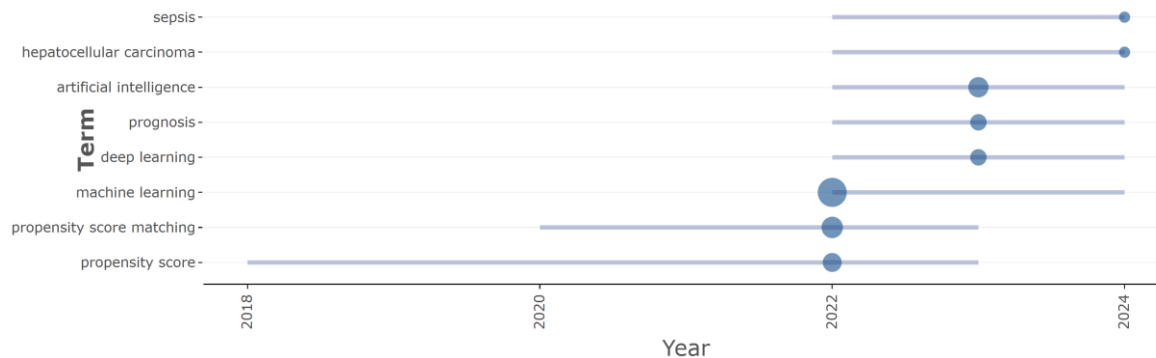
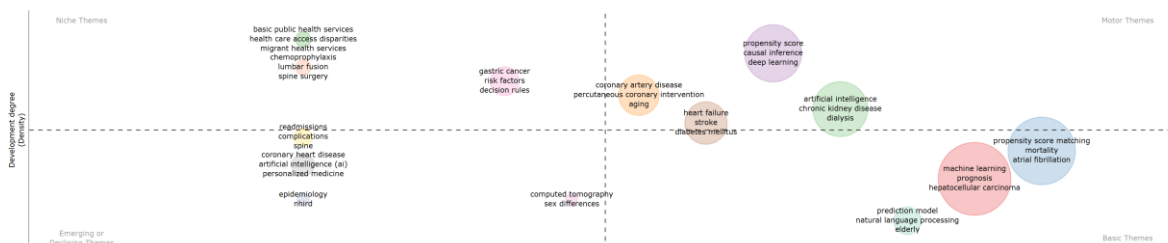


Figure 7. Mapping of trend topics based on author's keywords.



4. Discussion

This study systematically examined the evolving intersection between AI and PSM in medical research and identified a rapidly expanding methodological domain characterised by bidirectional methodological enhancement. Evidence from the thematic and bibliometric synthesis indicates that the combined use of AI and PSM has accelerated sharply since 2020, coinciding with the broader adoption of large-scale electronic health records, registry data, and digital health infrastructures. The findings suggest that the integration of AI into PSM workflows, and conversely the incorporation of PSM into AI-driven medical studies, is emerging as a critical methodological approach for improving causal inference, mitigating confounding bias, and advancing real-world digital medical evidence/big data analysis, aggregation and synthesis.

A key observation of this knowledge synthesis study is that AI techniques are increasingly being adopted to augment the PSM pipeline. Across included studies, machine learning and deep learning algorithms were employed to support covariate selection, estimate propensity scores in high-dimensional settings, identify non-linear and interaction effects, and improve pre-processing of structured and unstructured clinical data. Natural language processing models, particularly domain BERT, ChatGPT and Gemini adoptions, were used to extract clinically meaningful covariates from free-text clinical notes, thereby enhancing the completeness and quality of matching sets. The results demonstrate that AI-enhanced PSM is particularly valuable in complex observational datasets, where classical regression-based estimation techniques may be limited by covariate sparsity, non-linearity, or multidimensional feature relationships.

Importantly, the inverse symbiotic relationship—PSM used to strengthen AI applications—was also evident. In a substantial proportion of studies, PSM was applied prior to machine learning model training to reduce baseline imbalances, control confounding, and ensure that subsequent AI-based prediction models were trained on balanced and comparable cohorts. This sequencing reflects a methodological recognition that AI systems, when trained on imbalanced observational data, risk amplifying bias and generating unreliable or unfair predictions. PSM therefore functions as a foundational causal-inference layer that improves the validity, interpretability, and clinical acceptability of AI-based models. Notably, several studies coupled PSM with fairness-aware machine learning frameworks, suggesting an emerging alignment between causal inference methodology and ethical AI development.

Comparison with the broader literature confirms alignment with current methodological transitions in medical data science, including the adoption of target trial emulation, doubly robust learning methods, and explainable AI techniques. The increasing appearance of SHAP-based model interpretability approaches in the analysed corpus reflects a growing acknowledgement that explainability is essential when AI is deployed in clinical decision-making environments. Furthermore, the results correspond with regulatory trends emphasising transparency, fairness, and accountability in medical AI systems, as seen in evolving FDA and EMA guidance on machine learning-based medical technologies.

Several practical implications arise from these findings. First, the observed symbiosis between AI and PSM underscores the value of hybrid analytical pipelines for generating robust evidence from real-world clinical data. Such approaches may support precision medicine, enable reliable risk stratification, and enhance translational research by strengthening causal assumptions in observational studies. Second, the combined use of AI and PSM holds promise for hospitals and healthcare systems seeking to leverage routinely collected data for predictive modelling, outcome evaluation, and clinical decision support. Finally, the results highlight a methodological trajectory that can inform future clinical research training and the development of multi-disciplinary analytical frameworks integrating biostatistics, machine learning, and clinical epidemiology.

Despite substantial momentum, several gaps and challenges remain. The reviewed literature indicates limited methodological standardisation; AI-enhanced PSM implementations vary considerably across studies, and benchmarking frameworks for assessing performance and bias reduction are not yet mature. Although causal machine learning techniques were identified, their use remained limited relative to classical PSM frameworks, signalling a need for broader adoption of causal representation learning, counterfactual inference, and heterogeneous treatment-effect modelling. Furthermore, dependency on cross-sectional or static datasets limits applicability to time-varying clinical processes. Dynamic PSM and online learning approaches capable of updating matches as new data accumulate remain under-developed. Ethical considerations—including privacy protection, bias mitigation, and responsible model governance—were acknowledged only sporadically, reflecting another key research frontier.

This study also has methodological limitations. Although Scopus provides broad multidisciplinary coverage, relevant studies indexed exclusively in other databases may have been excluded. Keyword-based bibliometric retrieval may not fully capture studies using alternative terminologies for matching or AI methods. Author keyword heterogeneity limited granular quantification of algorithm-specific trends, necessitating complementary abstract-level analysis. Additionally, synthetic thematic analysis provides interpretive depth, but does not replicate the exhaustive appraisal of systematic reviews. Nonetheless, the triangulation of bibliometric mapping with thematic synthesis strengthens the credibility of the findings.

Future research should prioritise automated and scalable AI frameworks for propensity score estimation, integration of causal machine learning systems, and development of shared benchmark datasets for evaluating hybrid AI-PSM pipelines. Increased adoption of longitudinal and multi-modal clinical data, federated architectures, and transparent model governance protocols will be essential. As AI-integrated PSM methods advance, rigorous validation in prospective clinical settings and alignment with ethical and regulatory frameworks will remain critical to ensuring robust, fair, and clinically meaningful implementation.

5. Conclusion

In conclusion, this study reviews and synthetically analyses the current scientific literature in order to present the current state of integration and usage of AI in the domain of propensity score matching. This synthetic knowledge review provided several clear research themes, which demonstrate the bi-directional and complementary nature of AI in propensity score matching and integration in condition prognosis and diagnostics. The integration of AI in propensity score matching proved integral in the improvement of diagnosis accuracy and the understanding of

reasoning and helped overcome the limitations that are intrinsic to propensity score measurements. On the other hand, critical shortcomings were outlined, for instance, the need for advanced techniques for flawless integration of AI into propensity score matching, the need for clarity and interpretability of AI models, robust solutions for dealing with confounding unmeasured factors, and the need for standardization of evaluation. The other major problem, which is still one of the greatest obstacles when it comes to AI usage in medicine, is the ethical concern. Future research will have to focus on those critical points, the integration of AI, adding robustness, integrating, and formulating ethical concepts that will enable AI to be safe and acceptable. The implementation and verification of those new approaches will have to be performed on different longitudinal medical datasets since the synergistic usage of AI in propensity score matching provides new avenues and progress in the field of precise medicine. The main goal must be to improve decision-making in healthcare and the optimization of patient outcomes.

Author Contributions: Writing—review and editing, Writing—original draft, Supervision, Conceptualization: P.K., B.Ž., T.Z., Data analysis, Methodology development, Visualization: P.K., H.B.V., and B.Ž.; Writing—review and editing, supervision: All, All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Public Involvement Statement: There is no public involvement in any aspect of this article.

Use of Artificial Intelligence: AI or AI-assisted tools were not used in drafting any aspect of this manuscript. A.I was used only for language and grammar checking

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dang A. Real-World Evidence: A Primer. *Pharmaceut Med* 2023;37:25–36. <https://doi.org/10.1007/s40290-022-00456-6>.
2. Li Q, Lin J, Chi A, Davies S. Practical considerations of utilizing propensity score methods in clinical development using real-world and historical data. *Contemporary Clinical Trials* 2020;97:106123. <https://doi.org/10.1016/j.cct.2020.106123>.
3. Rivas JG, Kraft P, Evans-Axelsson S, Hijazy A, Beyers K, De Meulder B, et al. Real-world Evidence on Baseline Characteristics and Treatment in Metastatic Hormone-sensitive Prostate Cancer: Findings from the PIONEER 2.0 Big Data Investigation Group. *Eu Uro Op Sci* 2025;81:82–91. <https://doi.org/10.1016/j.euros.2025.09.010>.
4. Al-Antari MA. Artificial Intelligence for Medical Diagnostics—Existing and Future AI Technology! *Diagnostics* 2023;13:688. <https://doi.org/10.3390/diagnostics13040688>.
5. Artificial intelligence meets medical robotics | Science n.d. https://www.science.org/doi/full/10.1126/science.adj3312?casa_token=HoLADs-riL4AAAAA%3AIU3aQJbwQEY0iPYzPU33NHeof8CLJxIq8kJonOrHDAyKUZ1yYmEgCiA1wbPSyJFsiEKks2hnpeys2U (accessed December 13, 2024).
6. Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. *Brain* 2022;145:457–75. <https://doi.org/10.1093/brain/awab439>.
7. Briganti G, Le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. *Front Med* 2020;7. <https://doi.org/10.3389/fmed.2020.00027>.

8. Liao J, Li X, Gan Y, Han S, Rong P, Wang W, et al. Artificial intelligence assists precision medicine in cancer treatment. *Front Oncol* 2023;12. <https://doi.org/10.3389/fonc.2022.998222>.
9. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. *The Lancet Digital Health* 2021;3:e195–203. [https://doi.org/10.1016/S2589-7500\(20\)30292-2](https://doi.org/10.1016/S2589-7500(20)30292-2).
10. Shick AA, Webber CM, Kiarashi N, Weinberg JP, Deoras A, Petrick N, et al. Transparency of artificial intelligence/machine learning-enabled medical devices. *Npj Digit Med* 2024;7:1–4. <https://doi.org/10.1038/s41746-023-00992-8>.
11. Tian M, Shen Z, Wu X, Wei K, Liu Y. The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier. *Academic Journal of Science and Technology* 2023;8:57–61. <https://doi.org/10.54097/ajst.v8i2.14945>.
12. van de Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RER, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. *BMJ Health Care Inform* 2022;29:e100495. <https://doi.org/10.1136/bmjhci-2021-100495>.
13. Lu Y, Jin J, Zhang H, Lu Q, Zhang Y, Liu C, et al. Traumatic brain injury: Bridging pathophysiological insights and precision treatment strategies. *Neural Regen Res* 2026;21:887–907. <https://doi.org/10.4103/NRR.NRR-D-24-01398>.
14. Xiong X, Zheng L-W, Ding Y, Chen Y-F, Cai Y-W, Wang L-P, et al. Breast cancer: pathogenesis and treatments. *Signal Transduct Target Ther* 2025;10. <https://doi.org/10.1038/s41392-024-02108-4>.
15. Katip W, Rayanakorn A, Oberdorfer P, Taruangsri P, Nampuan T. Short versus long course of colistin treatment for carbapenem-resistant *A. baumannii* in critically ill patients: A propensity score matching study. *Journal of Infection and Public Health* 2023;16:1249–55. <https://doi.org/10.1016/j.jiph.2023.05.024>.
16. Krenzien F, Schmelzle M, Pratschke J, Feldbrügge L, Liu R, Liu Q, et al. Propensity Score-Matching Analysis Comparing Robotic Versus Laparoscopic Limited Liver Resections of the Posterosuperior Segments: An International Multicenter Study. *Annals of Surgery* 2024;279:297. <https://doi.org/10.1097/SLA.0000000000006027>.
17. Langworthy B, Wu Y, Wang M. An overview of propensity score matching methods for clustered data. *Stat Methods Med Res* 2023;32:641–55. <https://doi.org/10.1177/09622802221133556>.
18. Meneguzzo P, Antoniadou A, Garolla A, Tozzi F, Todisco P. Predictors of psychopathology response in atypical anorexia nervosa following inpatient treatment: A propensity score matching study of weight suppression and weight loss speed. *International Journal of Eating Disorders* 2024;57:1002–7. <https://doi.org/10.1002/eat.24135>.
19. Wang SV, Schneeweiss S, Franklin JM, Desai RJ, Feldman W, Garry EM, et al. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses. *JAMA* 2023;329:1376–85. <https://doi.org/10.1001/jama.2023.4221>.
20. Zhu P, Liao W, Zhang W-G, Chen L, Shu C, Zhang Z-W, et al. A Prospective Study Using Propensity Score Matching to Compare Long-term Survival Outcomes After Robotic-assisted, Laparoscopic, or Open Liver Resection for Patients With BCLC Stage 0-A Hepatocellular Carcinoma. *Annals of Surgery* 2023;277:e103. <https://doi.org/10.1097/SLA.0000000000005380>.
21. Jochum F, Dumas É, Gougis P, Hamy A-S, Querleu D, Lecointre L, et al. Survival outcomes of primary vs interval cytoreductive surgery for International Federation of Gynecology and Obstetrics stage IV ovarian cancer: a nationwide population-based target trial emulation. *Am J Obstet Gynecol* 2025;232:194.e1-194.e11. <https://doi.org/10.1016/j.ajog.2024.07.044>.
22. Yang S, Hussain M, Ammar Zahid RM, Maqsood US. The role of artificial intelligence in corporate digital strategies: evidence from China. *Kybernetes* 2025;54:3062–82. <https://doi.org/10.1108/K-08-2023-1583>.
23. Park J-B, Bae JH. Effectiveness of a novel artificial intelligence-assisted colonoscopy system for adenoma detection: a prospective, propensity score-matched, non-randomized controlled study in Korea. *Clin Endosc* 2025;58:112–20. <https://doi.org/10.5946/ce.2024.168>.

24. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives†. *European Journal of Cardio-Thoracic Surgery* 2018;53:1112–7. <https://doi.org/10.1093/ejcts/ezy167>.
25. Kim DW. Statistical Methods for Baseline Adjustment and Cohort Analysis in Korean National Health Insurance Claims Data: A Review of PSM, IPTW, and Survival Analysis With Future Directions. *J Korean Med Sci* 2025;40. <https://doi.org/10.3346/jkms.2025.40.e110>.
26. Ghimire L, Waller E. The Future of Health Physics: Trends, Challenges, and Innovation. *Health Physics* 2025;128:167–89. <https://doi.org/10.1097/HP.0000000000001882>.
27. Xiao X, Alharbi K, Zhang P, Qin H, Yue X. Bayesian federated causal inference and its application in manufacturing. *J Intell Manuf* 2025. <https://doi.org/10.1007/s10845-025-02665-7>.
28. Hennecken J. Predicting Subclinical Atrial Fibrillation using Artificial Intelligence and validate using propensity-score matching and Explainable AI. Master Thesis. 2024.
29. Ishiyama M, Kudo S, Misawa M, Mori Y, Maeda Y, Ichimasa K, et al. Impact of the clinical use of artificial intelligence–assisted neoplasia detection for colonoscopy: a large-scale prospective, propensity score–matched study (with video). *Gastrointestinal Endoscopy* 2022;95:155–63. <https://doi.org/10.1016/j.gie.2021.07.022>.
30. Kim H, Choi JS, Kim K, Ko ES, Ko EY, Han B-K. Effect of artificial intelligence–based computer-aided diagnosis on the screening outcomes of digital mammography: a matched cohort study. *Eur Radiol* 2023;33:7186–98. <https://doi.org/10.1007/s00330-023-09692-z>.
31. Prospero M, Ghosh S, Chen Z, Salemi M, Lyu T, Zhao J, et al. Causal AI with Real World Data: Do Statins Protect from Alzheimer’s Disease Onset? Proceedings of the 5th International Conference on Medical and Health Informatics, New York, NY, USA: Association for Computing Machinery; 2021, p. 296–303. <https://doi.org/10.1145/3472813.3473206>.
32. Karim ME. Can supervised deep learning architecture outperform autoencoders in building propensity score models for matching? *BMC Med Res Methodol* 2024;24:167. <https://doi.org/10.1186/s12874-024-02284-5>.
33. Lourenço L, Weber L, Garcia L, Ramos V, Souza J. Machine Learning Algorithms to Estimate Propensity Scores in Health Policy Evaluation: A Scoping Review. *IJERPH* 2024;21:1484. <https://doi.org/10.3390/ijerph21111484>.
34. Whata A, Chimedza C. Evaluating Uses of Deep Learning Methods for Causal Inference. *IEEE Access* 2022;10:2813–27. <https://doi.org/10.1109/ACCESS.2021.3140189>.
35. Kokol P, Kokol M, Zagoranski S. Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress* 2022;105:00368504211029777. <https://doi.org/10.1177/00368504211029777>.
36. Kokol P. Synthetic Knowledge Synthesis in Hospital Libraries. *Journal of Hospital Librarianship* 2024;24:10–7. <https://doi.org/10.1080/15323269.2023.2291282>.
37. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84:523–38. <https://doi.org/10.1007/s11192-009-0146-3>.
38. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 2017;11:959–75. <https://doi.org/10.1016/j.joi.2017.08.007>.
39. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84:523–38. <https://doi.org/10.1007/s11192-009-0146-3>.
40. Austin PC. Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. *Statistics in Medicine* 2011;30:1292–301. <https://doi.org/10.1002/sim.4200>.
41. Austin PC, Small DS. The use of bootstrapping when using propensity-score matching without replacement: a simulation study. *Statist Med* 2014;33:4306–19. <https://doi.org/10.1002/sim.6276>.
42. Benedetto U, Head SJ, Angelini GD, Blackstone EH. Statistical primer: propensity score matching and its alternatives†. *European Journal of Cardio-Thoracic Surgery* 2018;53:1112–7. <https://doi.org/10.1093/ejcts/ezy167>.
43. Scimago Journal & Country Rank n.d. <https://www.scimagojr.com/> (accessed November 16, 2025).

44. Islam N, Islam S, Roy PB. A bibliometric technique for analyzing trends in public health research. *Data Science and Informetrics* 2025. <https://doi.org/10.1016/j.dsim.2025.01.010>.
45. Colaneri M, Fama F, Fassio F, Holmes D, Scaglione G, Mariani C, et al. Impact of early antiviral therapy on SARS-CoV-2 clearance time in high-risk COVID-19 subjects: A propensity score matching study. *International Journal of Infectious Diseases* 2024;149. <https://doi.org/10.1016/j.ijid.2024.107265>.
46. Xie Y, Shen H, Xu Q, Tu C, Yang R, Liu T, et al. Evaluating coronary arteries and predicting MACEs using CCTA in lung cancer patients receiving chemotherapy or chemoradiotherapy. *Radiotherapy and Oncology* 2024;200:110498. <https://doi.org/10.1016/j.radonc.2024.110498>.
47. Lim J, Choi Y-J, Kim BS, Rhee T-M, Lee H-J, Han K-D, et al. Comparative cardiovascular outcomes in type 2 diabetes patients taking dapagliflozin versus empagliflozin: a nationwide population-based cohort study. *Cardiovascular Diabetology* 2023;22. <https://doi.org/10.1186/s12933-023-01911-7>.
48. Squicciarro E, Lorusso R, Consiglio A, Labriola C, Haumann RG, Piancone F, et al. Impact of Inflammation After Cardiac Surgery on 30-Day Mortality and Machine Learning Risk Prediction. *J Cardiothorac Vasc Anesth* 2025;39:683–91. <https://doi.org/10.1053/j.jvca.2024.12.013>.
49. Ngufor C, Zhang N, Van Houten HK, Holmes DR, Graff-Radford J, Alkhouli M, et al. Causal Machine Learning for Left Atrial Appendage Occlusion in Patients With Atrial Fibrillation. *JACC Clin Electrophysiol* 2025;11:977–86. <https://doi.org/10.1016/j.jacep.2024.12.013>.
50. Pettus J, Roussel R, Liz Zhou F, Bosnyak Z, Westerbacka J, Berria R, et al. Rates of Hypoglycemia Predicted in Patients with Type 2 Diabetes on Insulin Glargine 300 U/ml Versus First- and Second-Generation Basal Insulin Analogs: The Real-World LIGHTNING Study. *Diabetes Ther* 2019;10:617–33. <https://doi.org/10.1007/s13300-019-0568-8>.
51. Kumar S, Gupta P, Dekker ALAJ, Bermejo I, Kar S. Development and Validation of Multicenter Study on Novel Artificial Intelligence Based Cardiovascular Risk Score (AICVD) 2021. <https://doi.org/10.21203/rs.3.rs-1080955/v1>.
52. Scopus - Document details - DEVELOPMENT OF A MACHINE LEARNING MODEL FOR PREDICTING 28-DAY MORTALITY OF SEPTIC PATIENTS WITH ATRIAL FIBRILLATION n.d. <https://doi.org/10.1097/SHK.0000000000002078>.
53. Ruan H, Ran X, Li S, Zhang Q. Dyslipidemia versus obesity as predictors of ischemic stroke prognosis: a multi-center study in China. *Lipids in Health and Disease* 2024;23:72. <https://doi.org/10.1186/s12944-024-02061-9>.
54. Liang H, Pan K, Wang J, Lin J. Association between neutrophil percentage-to-albumin ratio and breast cancer in adult women in the US: findings from the NHANES. *Front Nutr* 2025;12. <https://doi.org/10.3389/fnut.2025.1533636>.
55. Gao Z, Winhusen TJ, Gorenflo M, Ghitza UE, Davis PB, Kaelber DC, et al. Repurposing ketamine to treat cocaine use disorder: integration of artificial intelligence-based prediction, expert evaluation, clinical corroboration and mechanism of action analyses. *Addiction* 2023;118:1307–19. <https://doi.org/10.1111/add.16168>.
56. Pundi K, Fan J, Kabadi S, Din N, Blomström-Lundqvist C, Camm AJ, et al. Dronedarone Versus Sotalol in Antiarrhythmic Drug-Naive Veterans With Atrial Fibrillation. *Circulation: Arrhythmia and Electrophysiology* 2023;16:456–67. <https://doi.org/10.1161/CIRCEP.123.011893>.
57. Qu J, Li C, Liu M, Wang Y, Feng Z, Li J, et al. Prognostic Models Using Machine Learning Algorithms and Treatment Outcomes of Occult Breast Cancer Patients. *Journal of Clinical Medicine* 2023;12. <https://doi.org/10.3390/jcm12093097>.
58. Park SW, Park Y-L, Lee E-G, Chae H, Park P, Choi D-W, et al. Mortality Prediction Modeling for Patients with Breast Cancer Based on Explainable Machine Learning. *Cancers* 2024;16. <https://doi.org/10.3390/cancers16223799>.
59. Hu J, Gong N, Li D, Deng Y, Chen J, Luo D, et al. Identifying hepatocellular carcinoma patients with survival benefits from surgery combined with chemotherapy: based on machine learning model. *World Journal of Surgical Oncology* 2022;20. <https://doi.org/10.1186/s12957-022-02837-2>.

60. Huang C, Liu Z, Xiao L, Xia Y, Huang J, Luo H, et al. Clinical Significance of Serum CA125, CA19-9, CA72-4, and Fibrinogen-to-Lymphocyte Ratio in Gastric Cancer With Peritoneal Dissemination. *Frontiers in Oncology* 2019;9. <https://doi.org/10.3389/fonc.2019.01159>.
61. Xu S, Xiang C, Wu J, Teng Y, Wu Z, Wang R, et al. Tongue Coating Bacteria as a Potential Stable Biomarker for Gastric Cancer Independent of Lifestyle. *Digestive Diseases and Sciences* 2021;66:2964–80. <https://doi.org/10.1007/s10620-020-06637-0>.
62. Makhnevich A, Perrin A, Talukder D, Liu Y, Izard S, Chiuzaan C, et al. Thick Liquids and Clinical Outcomes in Hospitalized Patients with Alzheimer Disease and Related Dementias and Dysphagia. *JAMA Internal Medicine* 2024;184:778–85. <https://doi.org/10.1001/jamainternmed.2024.0736>.
63. Digumarthi V, Amin T, Kanu S, Mathew J, Edwards B, Peterson LA, et al. Preoperative prediction model for risk of readmission after total joint replacement surgery: a random forest approach leveraging NLP and unfairness mitigation for improved patient care and cost-effectiveness. *Journal of Orthopaedic Surgery and Research* 2024;19. <https://doi.org/10.1186/s13018-024-04774-0>.
64. Pimentel SD, Yu R. Re-evaluating the impact of hormone replacement therapy on heart disease using match-adaptive randomization inference 2024. <https://doi.org/10.48550/arXiv.2403.01330>.
65. Feller DJ, Zucker J, Yin MT, Gordon P, Elhadad N. Using Clinical Notes and Natural Language Processing for Automated HIV Risk Assessment. *Journal of Acquired Immune Deficiency Syndromes* 2018;77:160–6. <https://doi.org/10.1097/QAI.0000000000001580>.
66. Zoccali C, Tripepi G. Clinical trial emulation in nephrology. *Journal of Nephrology* 2024. <https://doi.org/10.1007/s40620-024-02158-5>.
67. Patel SS, Raman VK, Zhang S, Sheriff HM, Fonarow GC, Heidenreich PA, et al. Renin Angiotensin Inhibition and Lower Risk of Kidney Failure in Patients with Heart Failure. *Am J Med* 2025;138:1384–1393.e5. <https://doi.org/10.1016/j.amjmed.2025.04.038>.
68. Inoue K, Seeman TE, Horwich T, Budoff MJ, Watson KE. Heterogeneity in the Association between the Presence of Coronary Artery Calcium and Cardiovascular Events: A Machine-Learning Approach in the MESA Study. *Circulation* 2023;147:132–41. <https://doi.org/10.1161/CIRCULATIONAHA.122.062626>.
69. Pietropaoli D, Monaco A, D’Aiuto F, Aguilera EM, Ortu E, Giannoni M, et al. Active gingival inflammation is linked to hypertension. *Journal of Hypertension* 2020;38:2018–27. <https://doi.org/10.1097/HJH.0000000000002514>.
70. Fu S, Chen L, Lin H, Jiang X, Zhang S, Zhong F, et al. Prediction Model for Delayed Behavior of Early Ambulation After Surgery for Varicose Veins of the Lower Extremity: A Prospective Case-Control Study. *Archives of Physical Medicine and Rehabilitation* 2024;105:1908–20. <https://doi.org/10.1016/j.apmr.2024.06.004>.
71. Krishnamurthy S, Kapeleshh KS, Dovgan E, Luštrek M, Gradišek Piletič B, Srinivasan K, et al. Machine learning prediction models for chronic kidney disease using national health insurance claim data in Taiwan. *Healthcare (Switzerland)* 2021;9. <https://doi.org/10.3390/healthcare9050546>.
72. Wang X, Guo J, Liu H, Zhao T, Li H, Wang T. Impact of Social Participation Types on Depression in the Elderly in China: An Analysis Based on Counterfactual Causal Inference. *Frontiers in Public Health* 2022;10. <https://doi.org/10.3389/fpubh.2022.792765>.
73. Ghosh S, Bian J, Guo Y, Prospero M. Deep propensity network using a sparse autoencoder for estimation of treatment effects. *Journal of the American Medical Informatics Association* 2021;28:1197–206. <https://doi.org/10.1093/jamia/ocaa346>.
74. Luo Q, Zheng Z, Luo W, Zhu J. Development and external validation of interpretable machine learning models for personalized multiple treatment recommendations in non-small cell lung cancer. *Int J Med Informatics* 2026;206. <https://doi.org/10.1016/j.ijmedinf.2025.106160>.
75. Weymann D, Chan B, Regier DA. Genetic matching for time-dependent treatments: a longitudinal extension and simulation study. *BMC Medical Research Methodology* 2023;23. <https://doi.org/10.1186/s12874-023-01995-5>.
76. Cui X, Shi Y, He X, Zhang M, Zhang H, Yang J, et al. Abdominal physical examinations in early stages benefit critically ill patients without primary gastrointestinal diseases: a retrospective cohort study. *Frontiers in Medicine* 2024;11. <https://doi.org/10.3389/fmed.2024.1338061>.

77. Chen M, Yang J, Lu J, Zhou Z, Huang K, Zhang S, et al. Ureteral calculi lithotripsy for single ureteral calculi: can DNN-assisted model help preoperatively predict risk factors for sepsis? *European Radiology* 2022;32:8540–9. <https://doi.org/10.1007/s00330-022-08882-5>.
78. Squicciarro E, Lorusso R, Consiglio A, Labriola C, Haumann RG, Piancone F, et al. Impact of Inflammation After Cardiac Surgery on 30-Day Mortality and Machine Learning Risk Prediction. *J Cardiothorac Vasc Anesth* 2025;39:683–91. <https://doi.org/10.1053/j.jvca.2024.12.013>.
79. Digumarthi V, Amin T, Kanu S, Mathew J, Edwards B, Peterson LA, et al. Preoperative prediction model for risk of readmission after total joint replacement surgery: a random forest approach leveraging NLP and unfairness mitigation for improved patient care and cost-effectiveness. *J Orthop Surg Res* 2024;19:287. <https://doi.org/10.1186/s13018-024-04774-0>.
80. Chen M, Yang J, Lu J, Zhou Z, Huang K, Zhang S. Ureteral calculi lithotripsy for single ureteral calculi: can DNN-assisted model help preoperatively predict risk factors for sepsis? *European Radiology* 2022;32:8540–9. <https://doi.org/10.1007/s00330-022-08882-5>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.