

Entry

# Predictive Modeling in Medicine

Milan Toma <sup>1,\*</sup>  and Ong Chi Wei <sup>2</sup> <sup>1</sup> New York Institute of Technology, College of Osteopathic Medicine, Old Westbury, New York, NY 11568, USA<sup>2</sup> A\*Star Institute of High Performance Computing, 1 Fusionopolis Way #16-16 Connexis, Singapore 138632, Singapore; ong\_chi\_wei@ihpc.a-star.edu.sg

\* Correspondence: tomamil@tomamil.com

**Definition:** Predictive modeling is a complex methodology that involves leveraging advanced mathematical and computational techniques to forecast future occurrences or outcomes. This tool has numerous applications in medicine, yet its full potential remains untapped within this field. Therefore, it is imperative to delve deeper into the benefits and drawbacks associated with utilizing predictive modeling in medicine for a more comprehensive understanding of how this approach may be effectively leveraged for improved patient care. When implemented successfully, predictive modeling has yielded impressive results across various medical specialities. From predicting disease progression to identifying high-risk patients who require early intervention, there are countless examples of successful implementations of this approach within healthcare settings worldwide. However, despite these successes, significant challenges remain for practitioners when applying predictive models to real-world scenarios. These issues include concerns about data quality and availability as well as navigating regulatory requirements surrounding the use of sensitive patient information—all factors that can impede progress toward realizing the true potential impact of predictive modeling on improving health outcomes.

**Keywords:** predictive; modeling; medicine; computational; models; forecast; future



**Citation:** Toma, M.; Wei, O.C. Predictive Modeling in Medicine. *Encyclopedia* **2023**, *3*, 590–601. <https://doi.org/10.3390/encyclopedia3020042>

Academic Editors: Samir Nammour and Raffaele Barretta

Received: 20 March 2023

Revised: 6 May 2023

Accepted: 9 May 2023

Published: 11 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Predictive modeling involves the use of mathematical or computational methods to create models that can forecast future outcomes. While equation-based models are used for the former approach, simulation techniques are required for the latter. Predictive modeling has numerous applications in medicine, such as clinical decision-making and clinical trials; however, its potential remains largely untapped in this field due to various challenges. The application of these techniques to the medical domain is particularly challenging because it deals with a dynamic nature of this discipline and complexity of patient populations treated in modern healthcare settings. Furthermore, developing and implementing effective predictive models requires a deep understanding of data being used along with adequate resources to support model development and implementation. A useful glossary table consisting of commonly used terms can be found in Table 1.

The development of various software tools in the medical domain has significantly improved the process of creating predictive models. With both open-source and commercial products available, researchers now have access to more options than ever before for their academic studies. These advancements are expected to continue benefiting medicine in years to come. However, it is essential that the use of these models be carefully scrutinized so as not to negatively impact patient care or violate ethical standards. Additionally, generating and validating a model should be a transparent and systematic process that ensures all relevant information is captured and presented comprehensibly.

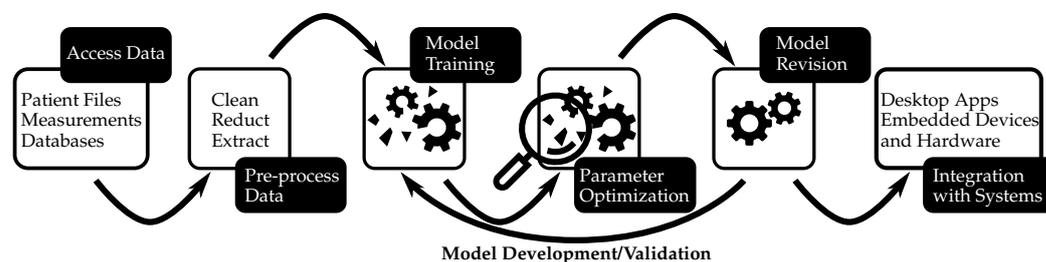
**Table 1.** Glossary summary of common terminology in predictive modeling.

Term	Definition
Predictive modeling	Process of using statistical or computer algorithms to analyze data and make predictions about future outcomes or behaviors.
Equation based model	Type of mathematical model that is used to describe and predict the behavior of a system based on a set of mathematical equations.
Time-series regression model	A type of statistical model used to analyze time-series data, i.e., data are collected over time.
Neural networks	Computer algorithms inspired by the brain to recognize patterns in data.
Bagged decision trees	Ensemble learning, multiple tree models training on subsets, reducing overfitting, and improving accuracy and stability.
Model validation	Process of evaluating and testing a machine learning model to ensure that it is accurate, reliable, and generalizes well on new, unseen data.
Similarity assessment	Process of comparing two or more objects, data points, or patterns to determine how similar or dissimilar they are based on certain criteria of features.
Domain estimation	The process of determining the range of values or categories that a variable can take based on available data.
Conformal prediction	Machine learning framework that provides a probabilistic guarantee of the accuracy of a prediction, based on a given level of confidence.

There are various approaches to developing and validating predictive models. The chosen approach relies on several factors, including the model type developed, data nature, and resource availability. This entry mainly focuses on the development and validation of more complex statistical models. These models differ from mechanistic models that rely on modeling phenomena studied using mathematical equations. Statistical models use empirical equations to capture statistical relationships between different variables instead of relying solely on a modeling approach as mechanistic ones do. This difference also applies to computer models. Additionally, during the predictive modeling process, there is often a coupling between mechanistic and statistical models in the prediction process. Climate models serve as an example since they are based on physical laws (represented by mathematical equations) with their parameters controlled by data and statistical models. However, all of these models share a similar workflow (see Figure 1); the first step involves gathering data for the model followed by developing either a mathematical model or simulation capable of predicting specific outcomes associated with an event. It is crucial to assess the precision of simulations by subjecting them to examinations with datasets that were not employed during the model's construction. This ensures that the model can perform accurately when applied to different datasets. Once validated, it can make predictions about future events based on historical data. These predictions are useful in making decisions regarding patient care, managing hospital resources, or evaluating drug effectiveness. To sum up, the act of creating a model should not be seen as a final objective in itself. Instead, it is crucial to continually assess and confirm its effectiveness through iterative procedures aimed at achieving optimal results.

A clear understanding of the advantages and limitations of this approach is essential for its successful implementation. This entry presents an overview of the key principles of predictive modeling, along with some challenges associated with its use in medicine. It also discusses recent developments in this field and potential future applications. The first section provides an overview of predictive modeling's key principles, followed by a discussion on how it can be applied in medicine. Additionally, the entry highlights major obstacles

related to implementing this technique as well as possible areas for further research that could benefit from using it.



**Figure 1.** There are many different approaches to developing and validating predictive models. However, all models share a similar workflow, an example of which is demonstrated in this diagram.

## 2. Key Principles of Predictive Modeling

Predictive modeling involves the use of mathematical and computational methods to forecast future events or outcomes. These methods encompass various techniques, such as regression analyses, decision trees, random forests, neural networks, and support vector machines. Algorithms are employed by these methods for data analysis and model building that can predict outcomes based on patterns and relationships found in the data. Two primary approaches are utilized to achieve this goal, as outlined below. Table 1 contains a useful glossary that facilitates quick reference to the meanings of the aforementioned terminologies while discussing predictive modeling concepts.

### 2.1. Equation-Based Predictive Modeling

Equation-based models belong to a specific category of models that utilize mathematical equations to describe the relationship between variables. These types of models are commonly utilized in scientific disciplines, such as physics, chemistry, and engineering for predicting the behavior of physical systems. To anticipate future outcomes by projecting changes in input variables mathematically, these models depend on parameters within them that explain how inputs impact the outcome analyzed. Time-series regression frameworks demonstrate this approach effectively through linear regressions used for forecasting airline traffic volume or fuel efficiency based on engine speed versus load adjustments [1,2]. This methodology can prove highly beneficial in forecasting outcomes associated with a specific ailment due to its simplicity in construction and ease of assessment. Therefore, it is often utilized in predictive analysis when copious amounts of information are available, and the outcome variable has previously been comprehensively characterized [3]. However, challenges arise when establishing an association between the variables used for model building and the precise outcome variable under scrutiny. Insufficient availability of data may hinder one's ability to establish a correlation between independent and dependent variables, leading to this complication. Furthermore, identifying the most significant independent variables in a given context can be challenging. Therefore, this approach has limitations when used to analyze complex systems that involve multiple variables contributing to the development of outcome variables. As a result, its application in medicine is not widespread and is unsuitable for situations where the outcome variable fluctuates over time, such as tracking disease incidence within a population relative to time. This issue arises because model parameters would become outdated if they were used to predict future occurrences of said diseases among populations; hence constant updates will be necessary to reflect changes within those communities over time. In equation-based models, a precise comprehension of the interconnections among variables used to develop a model and the outcome variable is indispensable for achieving such an objective. Conversely, machine learning techniques can be more advantageous in such circumstances. In this methodology, datasets are not exploited to create an equation; instead, machine learning algorithms, such as decision trees or neural networks, are employed for understanding correlations. The relationships between variables learned through these methods

depend on the dataset used rather than being predetermined by construction-rendering this approach, which is adaptable across diverse situations.

## 2.2. Computational Predictive Modeling

In contrast to the mathematical approach, computational predictive modeling utilizes models that cannot be easily elucidated by equations. Instead, simulation techniques are necessary for making predictions using this "black box" method. Unlike conventional methods, such as curve and surface fitting or time series regressions that offer insight into how factors relate input to outcomes, computational modeling does not provide such explanations. Machine learning techniques, including neural networks and bagged decision trees, can enable tasks, such as determining a borrower's credit rating [4] or identifying a wine's origin [5]. Numerous industries have effectively utilized these techniques to address a diverse range of problems. However, this method has certain shortcomings, such as its limited applicability when dealing with small datasets and the increased computational complexity it presents for larger ones. The field of insurance and finance has particularly benefited from leveraging computational modeling [6], given its ability to produce accurate predictions that facilitate informed decision-making. Nevertheless, implementing a computational model can be challenging due to factors, such as high costs and difficulties associated with accurately interpreting models; therefore, adopting an appropriate modeling approach requires careful consideration of the various aspects involved. Additionally, combining multiple methods and approaches may offer optimal outcomes by reducing errors, while providing valuable insights into the issue at hand.

Regardless of the approach employed, constructing a predictive model requires adhering to a uniform procedure that is illustrated in Figure 1. The process involves cleansing and refining data for modeling purposes, selecting an appropriate method for prediction, training, and assessing the model using subsets of information, and gauging its performance through goodness-of-fit tests before validating the accuracy on unused datasets. Once satisfied with the results obtained from this sequence of activities, models can be confidently utilized for prognostication. In various applications of predictive modeling, it is possible to repeat the procedure with varying conditions in order to attain optimal performance or enhance model effectiveness. Additionally, refinement of the model can be carried out over time for superior future prediction outcomes. Ultimately, deployment on numerous devices enables predictions whenever needed; however, this entire process may require significant effort and duration. Nevertheless, the advantages derived from an effective predictive model are enormous, thus indicating its increasing utilization in our daily lives as well as its growing popularity.

## 3. Selected Clinical Applications

Studies have demonstrated the use of computational modeling in medicine, which involves creating novel diagnostic tests and determining appropriate treatment plans for patients with specific diseases [7]. An example of such research involved building a computational model that accurately predicted mortality rates among acute myeloid leukemia patients [8]. The utilization of this model could potentially aid in identifying patients who are susceptible to fatality, and enable medical practitioners to devise more efficacious treatment techniques for managing such ailments. Similarly, a study conducted at University College London utilized computational modeling to anticipate the outlook for individuals with lung cancer, while employing similar methodologies. The outcome highlighted significant variances amongst patient prognoses based on specific genetic differences despite having identical forms of malignancy. This discovery may lead to the development of improved interventions tailored according to individual characteristics among those diagnosed with lung cancer [9].

### 3.1. Predictive Modeling in Cardiology

Using predictive modeling has the potential to improve decision-making and personalize healthcare by estimating clinical outcome probabilities. Recently, Peng et al. identified age, smoking status, and blood pressure as primary predictors of cardiovascular disease by developing a predictive model that utilized data from a sizable population-based study. This effective method recognized high-risk individuals [10]. The research underscores the importance of these factors in preventing and managing cardiovascular disease. Additionally, Sajid et al. explored how non-clinical variables could enhance predictive modeling regarding cardiovascular disease [11]. By utilizing machine learning algorithms, the authors determined that incorporating non-clinical factors, such as socioeconomic status and lifestyle behaviors, has resulted in improved efficacy of predictive models. This enhancement holds considerable implications for identifying high-risk individuals who could benefit from targeted interventions in cardiovascular disease prevention and management. These findings emphasize the significance of personalized medicine approaches while highlighting the potential of machine learning to enhance risk assessment capabilities when used in conjunction with data-driven and physics-based methods. Zhang and colleagues conducted a research project with the objective of constructing a predictive computational model to simulate the hyperelastic behavior of ventricular myocardium in 3D [12]. The authors utilized finite-element modeling, neural network techniques, and machine learning algorithms to build an accurate model that could predict how mechanical properties vary under different loading conditions. This study highlights the potential benefits that could result from incorporating predictive modeling, machine learning methods, and computer-based simulations into our understanding of biological tissue mechanics. Furthermore, this work may have significant implications for personalized medicine strategies aimed at managing cardiac ailments by providing more precise treatment options based on individual patient data.

Researchers at the Massachusetts Institute of Technology have developed a computational tool named "RiskCardio" that shows promise in predicting the probability of heart disease among patients who previously suffered from acute coronary syndrome. This software can classify individuals into distinct risk groups based on their raw electrocardiogram (ECG) score within just 15 min [13]. With its ability to assist healthcare professionals in early identification and preventive measures for high-risk patients, this technology holds significant potential as an addition to cardiovascular care. Other research teams have also produced similar models using interpretable ECG data techniques for diagnosing heart disease [14]. These studies demonstrate how computational modeling can enhance healthcare delivery quality and patient care outcomes.

### 3.2. Predictive Modeling of Therapeutic Agents

The utility of computational models extends to facilitating the discovery and development of novel therapeutic agents. In a recent investigation, it was revealed that computational modeling is capable of pinpointing promising drug targets for treating numerous cancers [15]. These discoveries hold promise in enabling more efficacious treatments for such ailments in the future. Moreover, exploiting computational models can foster comprehension of disease pathogenesis at a fundamental level while also advancing treatment modalities [16]. Additionally, researchers have employed computer-based simulation techniques to scrutinize brain function; for instance, creating a model that simulates human brain activity patterns under varying conditions. By utilizing these models, it is possible to formulate therapeutic remedies for various ailments. Several investigations have been conducted to detect novel biomarkers related to Alzheimer's disease [17]. Notably, scientists have developed a computational model that can predict the susceptibility of individuals developing this ailment based on their genetic makeup. Furthermore, such an approach has practical implications in forecasting treatment outcomes for affected patients [18]. These scientific advancements hold promise in promoting more effective interventions toward curing Alzheimer's disease moving forward.

Rapid advances in computer technology have facilitated researchers to develop novel techniques for studying intricate aspects of human physiology and various diseases. By leveraging these cutting-edge tools, healthcare professionals can explore innovative avenues toward developing more potent therapeutic interventions for a wide range of maladies.

### 3.3. Predictive Modeling of Surgery Outcomes

Predictive analytic algorithms possess the ability to identify data patterns and provide accurate forecasts without requiring any hypotheses. This feature enables them to offer individualized patient-specific information that could be helpful in discussing surgical risks with patients. However, only a few studies have been conducted utilizing these techniques within the adult spine surgery literature. Hence, there is an initial implementation of predictive analytics for enhancing outcomes in this domain [19]. The healthcare industry is currently engaged in a thorough investigation of how technology, including but not limited to artificial intelligence and machine learning systems, can be effectively utilized. These support physicians' decision-making abilities while predicting consequences across multiple medical domains. However, despite its popularity in many areas, it remains an emerging field for spine surgery [20]. Those seeking more insights into predictive modeling algorithms can refer to systematic reviews and evaluations of spinal surgeries found in works, such as [21,22].

De Silva and colleagues employed a retrospective methodology to gather information regarding patient characteristics, imaging results, and outcome data. The perioperative CT facilitated the automatic computation of image-based features that were evaluated in conjunction with pre-operative functional indicators as well as pain outcomes at 3- and 12-month post-surgery intervals. The findings revealed that integrating these image-derived parameters into predictive models resulted in superior analysis pertaining to lumbar spine surgery outcomes than could be obtained from conventional demographic data alone. This study underscores the potential value of utilizing advanced technology for scrutinizing surgical outcomes with a view towards improving medical decision-making [23].

To forecast links with nonocclusive mesenteric ischemia (NOMI), a multifactorial statistical analysis was conducted utilizing scored variables from the preoperative, intraoperative, and postoperative stages. The diagnosis of NOMI was determined based on assorted clinical indicators and substantiated through mesenteric angiography. Treatment included catheter retention in the superior mesenteric artery for vasodilator infusion. Ultimately, specific determinants, such as renal inadequacy and transfusion of packed red blood cells, were observed to possess a significant prognostic capacity for NOMI [24].

Cook et al. demonstrated the basics of anticipatory modeling through simulation and applied them in a study related to neurosurgery. Their findings suggest that classical regression can offer perceptive insights into causal mechanisms when relevant variables and confounding factors are thoroughly analyzed, whereas predictive modeling prioritizes prediction accuracy rather than understanding causality by utilizing alternative metrics for assessing model performance. Although combining predictions from multiple models may improve prognostication quality, it does not produce a single risk score [25].

Steimer et al. introduced the distributional clustering framework as a modeling approach for multivariate time series analysis and prediction of surgical outcomes in epilepsy patients based on intracranial EEG data [26]. Their findings revealed the effectiveness of this method in distinguishing between individuals who were seizure-free after surgery from those who were not, showcasing its potential application in clinical settings.

Gaskin et al. utilized a bootstrapped least absolute shrinkage and selection operator model to investigate the preoperative risk factors associated with complications in cataract surgery. They employed random forest classifiers to develop tailored predictive models for each type of complication. Therefore, this approach provides individualized risk assessments for patients based on their distinct attributes that are vital when considering cataract surgery [27].

A research investigation was conducted to assess patients who were registered in a co-management pathway for hip fractures. The Charlson comorbidity index and ASA score were used to estimate patient complexity. Using a multivariate linear regression analysis, with consideration of both individual-specific and system-specific factors, scientists created a projection model predicting the duration of hospitalization. The findings revealed that the prevention of delirium occurrences and reduced surgical wait times proved to be significant prognosticators indicating shorter hospital stays [28].

To reduce healthcare expenses, industry experts have identified total joint replacement surgery as a key area for cost reduction. Since this elective surgical procedure is commonly performed in the United States and has variable costs depending on factors, such as volume, it is crucial to effectively identify and engage patients who may undergo the operation. Qiu et al. conducted an experiment using various machine-learning algorithms and developed an innovative deep-learning strategy that utilizes commercial claims datasets to predict total joint replacement surgeries [29]. The goal of their study was to improve patient identification methods for this type of surgery.

In a study conducted by Passias et al. [30], patients with cervical deformities were categorized into two groups depending on whether they experienced major complications or revision surgery after their initial procedure. The researchers utilized multivariable logistic regressions and decision tree analysis to recognize predictors for these outcomes, revealing that several characteristics significantly influenced the process. Specifically, radiographic parameters emerged as crucial indicators of potential revisions while baseline bone health, surgical traits, and various radiographic factors proved vital in predicting the probability of encountering significant complications post-surgery.

### *3.4. Predictive Modeling of Cancer Characterization*

Frequent visits to the emergency department and hospitalizations can escalate oncology care costs, which negatively impact the quality of life for cancer patients. With value- and quality-based payment models gaining momentum in this sector, preventing such events from occurring has become essential. To overcome this challenge, machine learning algorithms offer a promising solution by precisely identifying high-risk individuals at an early stage and delivering personalized care plans that address their specific requirements. This approach has shown potential in forecasting and avoiding costly incidents, such as emergency department visits or hospital admissions among cancer treatment recipients, resulting in improved outcomes while reducing overall expenses.

Despite the widespread use of predictive modeling for risk stratification in oncology, there are still significant difficulties associated with translating these models into clinical practice. Although numerous predictive models have been developed to forecast treatment-related toxicity or acute care incidents among cancer patients, they lack sufficient validation alongside corresponding interventions and primary consideration for implementation. Osterman et al.'s review clarifies the challenges that must be overcome before such models can become customary instruments within the domain of oncology [31].

The progress in technology and computing capabilities has enabled the comprehension of intricate mechanisms that impact the therapeutic reactions in cancer patients. However, precise prediction requires a refined human-machine interplay ingrained within machine learning design because extensive data are involved. Panja et al.'s discourse on simulating treatment reactions among cancer patients examines various machine learning methods, such as random forests and neural networks, while also considering constraints and alternative methodologies for future investigations [32].

The emergence of personalized medicine and complex medical data has led to the development of numerous prediction models. In particular, there has been a rise in clinical models, such as algorithms, nomograms, and risk-scoring systems, for the categorization of endometrial cancer patients into various subgroups. However, uncertainties still exist regarding optimal surgical staging for lymph node metastasis as well as recurrence and survival outcomes. To address this issue with an emphasis on practical use in real-world

settings, Bendifallah et al. conducted a review focused on existing prognostic and predictive models specific to endometrial cancer while discussing methodological aspects required for integrating these tools into clinical decision-making processes [33].

The classification of glioma and the implementation of predictive models based on artificial intelligence, using multi-modal MRI biomarkers, have shown promise in terms of individualized treatment plans. It has become vital to use hand-crafted or auto-extracted features derived from MRI due to their association with genomic alterations associated with MRI-based phenotypes. A comprehensive survey by Gore et al. aims at integrating all up-to-date work related to molecular diagnosis, prognosis, and monitoring treatment utilizing state-of-the-art radiomics and machine learning solutions for a complete resource on radiogenomic analysis of glioma [34].

In the field of radiation therapy, particularly brachytherapy, predictive modeling can be achieved by using dosimetric and physical parameters obtained from treatment features. To create models that use these measures either individually or in combination, four machine learning methodologies were used. Continued research into the potential applications of such algorithms within this domain may serve as a critical component for improving predictive capabilities [35].

Upon conducting a prognostic investigation on 47,625 individuals with cancer, it was found that natural language processing can proficiently estimate their survival rates through traditional and neural models [36]. The outcomes were as good as or superior to previous research, signifying its viability for practical usage in predicting the endurance of patients with cancer. Furthermore, this methodology eliminates the necessity for supplementary data or training separate models based on specific types of cancer by utilizing initial oncologist consultation documents pertaining to all forms of cancer. This is a promising development in the field of cancer research as improving prognostic accuracy can be instrumental in developing individualized treatment plans and improving patient outcomes.

### *3.5. Predictive Modeling for Drug Discovery*

Predictive modeling for drug discovery is an emerging field that has the potential to revolutionize the discovery of new drugs. It involves developing models capable of predicting the effect of small molecule inhibitors on a target protein in cell culture or animal models. These models can identify promising lead compounds and guide experimental efforts. The use of computational techniques is crucial for visualizing, analyzing, and predicting chemical and biological data. Predictive cheminformatics and bioinformatics rely on statistical methods to extract valuable information from vast databases designed for drug development. However, successful implementation requires careful consideration of factors, such as model validation, similarity assessment, domain estimation, and pre-processing, which are essential for interpreting results from structure-activity landscape models [37].

Machine learning has played an important role in drug discovery for the past two decades. However, its potential impact on clinical trial design and analysis is only beginning to be realized with recent advancements in computational technology and data collection/processing capabilities. The recent COVID-19 pandemic may further accelerate the adoption of artificial intelligence techniques in clinical trials due to increased reliance on digital technologies. As pharmaceutical companies seek greater efficiency and cost savings amid rising drug development expenses, artificial intelligence's automated nature and predictive abilities make it a promising tool for improving drug development processes overall [38]. The utilization of natural products in drug development is crucial, and high-throughput technologies have made it easier to discover their therapeutic effects. However, despite generating a large amount of data, interpreting that data remains challenging. Artificial intelligence techniques show promise as solutions to this issue; however, further exploration is needed for optimal results [39].

Predictive modeling encounters the challenge of accurately quantifying the dependability of model predictions on new objects. However, conformal prediction provides a proven and rigorous framework for in-silico modeling that assures error rates while continuously handling applicability domains associated with machine learning models. Alvarsson et al. have defined types and concepts connected to conformal prediction, which generates valid confidence estimates specific to each predicted object via prediction intervals comprising upper and lower bounds for regression or sets containing none, one, or multiple potential classes for classification purposes when employing predictive models [40].

The high death rate caused by malaria worldwide necessitates the development of new and highly effective drugs against *Plasmodium falciparum*. However, there are challenges, such as resistance to first-line drugs and a lack of suitable animal models for anti-*P. falciparum* assays, along with the complex life cycle of *Plasmodium* that needs to be overcome. Newer approaches in antimalarial drug discovery, including machine learning tools, have emerged with promising results from studies using random forest and support vector machines on limited datasets. The review by Oguike et al. offers insights into these approaches, providing a basis for further research toward developing potent antimalarial compounds [41].

#### 4. Discussion

Healthcare data comprise health-related information sourced from medical records and surveys that pertain to individuals or groups. Healthcare analytics is an essential resource for healthcare professionals, including hospitals, doctors, psychologists, and pharmacists. It can also be used by stakeholders who wish to enhance care quality through informed decisions supported by reliable insights extracted from extensive datasets, such as EHRs and claims-based registries. Predictive modeling, which uses AI, machine learning, and data mining techniques to evaluate past and present information in order to make future predictions forms a key part of data analytics. In healthcare settings specifically, predictive analytics helps analyze both current as well as historical medical records aiding practitioners to improve their decision-making processes while forecasting trends or managing disease outbreaks more efficiently.

The utilization of predictive analysis techniques that take into account medical records, age, social, and economic characteristics, as well as individual anatomy has enabled healthcare organizations to determine the probability of patients developing various lifestyle-related conditions. The application of advanced methods, such as simulation, variable modeling, and predictive analysis has become an essential tool for global healthcare institutions in enhancing their decision-making abilities [42]. These techniques are particularly useful within a fast-paced industry, such as healthcare, where vast amounts of data must be analyzed expeditiously while effectively managing risks. The incorporation of advanced technologies has significantly improved the problem-solving abilities of medical companies, thereby identifying opportunities to enhance global healthcare systems. The adoption of innovative methods in daily operations can greatly enhance work efficiency and reduce stress levels for those responsible for improving the health needs of millions worldwide.

Although there have been advancements in predictive modeling for drug-target interactions over the past twenty years, an important gap still exists when it comes to comprehending how these connections lead to clinical outcomes. Specifically, there is a need for forecasting genome-wide receptor activities and function selectivity that are brought about by new chemicals—particularly with regard to agonist versus antagonist impacts. However, achieving this objective proves challenging due to insufficient data on receptor activity as well as the necessity of training models with diverse shifted distributions suitable for real-world applications [43].

Data analytics is essential in healthcare to transform large amounts of data into valuable insights that improve patient outcomes, enhance operations, optimize resources, and predict disease outbreaks. It also ensures better health results for patients while

optimizing business performance within the sector. Data analytics plays a vital role in revolutionizing the healthcare industry by converting unprocessed health-related information into practical solutions. This revolutionary technology has made significant contributions to various areas of healthcare, such as clinical research, the development of new treatments and drugs, disease prediction and prevention, clinical decision support, and more accurate diagnoses. Additionally, it offers automation for hospital administrative processes that streamline operations while providing high success rates for surgeries and medications that benefit patients directly. Furthermore, its ability to calculate precise health insurance rates ensures accuracy in coverage calculations, making it an indispensable tool in today's modern world.

As technology and research continue to progress at an unprecedented pace, predictive modeling is likely to maintain its crucial position as a powerful tool in enhancing patient outcomes across various fields. The potential applications of this innovative approach seem boundless, paving the way for significant breakthroughs and advances in medical science. With every discovery and development, we edge ever closer to unlocking even greater potentials from predictive modeling with respect to healthcare, revolutionizing our understanding of disease diagnosis, treatment options, and prevention strategies among others.

**Author Contributions:** Conceptualization, M.T. and O.C.W.; methodology, M.T. and O.C.W.; formal analysis, M.T. and O.C.W.; investigation, M.T. and O.C.W.; resources, M.T. and O.C.W.; data curation, M.T. and O.C.W.; writing—original draft preparation, M.T. and O.C.W.; writing—review and editing, M.T. and O.C.W.; visualization, M.T. and O.C.W.; supervision, M.T. and O.C.W.; project administration, M.T. and O.C.W.; funding acquisition, M.T. and O.C.W. 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:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ECG	Electrocardiogram
CT	Computer Tomography
NOMI	Nonocclusive Mesenteric Ischemia
EEG	Electroencephalography
ASA	American Society of Anesthesiology
MRI	Magnetic Resonance Imaging
COVID-19	Coronavirus Disease 2019
EHR	Electronic Health Record

### References

1. Yang, Y.; Gong, N.; Xie, K.; Liu, Q. Predicting gasoline vehicle fuel consumption in energy and environmental impact based on machine learning and multidimensional big data. *Energies* **2022**, *15*, 1602. [[CrossRef](#)]
2. Lei, L.; Wen, Z.; Peng, Z. Prediction of main engine speed and fuel consumption of inland ships based on deep learning. *J. Phys. Conf. Ser.* **2021**, *2025*, 012012. [[CrossRef](#)]
3. Pollok, A.; Klöckner, A.; Zimmer, D. Psychological aspects of equation-based modelling. *Math. Comput. Model. Dyn. Syst.* **2019**, *25*, 115–138. [[CrossRef](#)]
4. Jiang, Y. *A Primer on Machine Learning Methods for Credit Rating Modeling*; IntechOpen: London, UK, 2022.
5. Gupta, U.; Patidar, Y.; Agarwal, A.; Singh, K.P. Wine quality analysis using machine learning algorithms. In *Micro-Electronics and Telecommunication Engineering: Proceedings of 3rd ICMETE 2019*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 11–18.

6. Broby, D. The use of predictive analytics in finance. *J. Financ. Data Sci.* **2022**, *8*, 145–161. [[CrossRef](#)]
7. Golas, S.B.; Nikolova-Simons, M.; Palacholla, R.; op den Buijs, J.; Garberg, G.; Orenstein, A.; Kvedar, J. Predictive analytics and tailored interventions improve clinical outcomes in older adults: A randomized controlled trial. *NPJ Digit. Med.* **2021**, *4*, 97. [[CrossRef](#)]
8. Sorrow, M.L.; Storer, B.E.; Fathi, A.T.; Gerds, A.T.; Medeiros, B.C.; Shami, P.; Brunner, A.M.; Sekeres, M.A.; Mukherjee, S.; Peña, E.; et al. Development and Validation of a Novel Acute Myeloid Leukemia–Composite Model to Estimate Risks of Mortality. *JAMA Oncol.* **2017**, *3*, 1675. [[CrossRef](#)]
9. Yang, Y.; Xu, L.; Sun, L.; Zhang, P.; Farid, S.S. Machine learning application in personalised lung cancer recurrence and survivability prediction. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1811–1820. [[CrossRef](#)]
10. Peng, M.; Hou, F.; Cheng, Z.; Shen, T.; Liu, K.; Zhao, C.; Zheng, W. Prediction of cardiovascular disease risk based on major contributing features. *Sci. Rep.* **2023**, *13*, 4778. [[CrossRef](#)]
11. Sajid, M.R.; Muhammad, N.; Zakaria, R.; Shahbaz, A.; Bukhari, S.A.C.; Kadry, S.; Suresh, A. Nonclinical features in predictive modeling of cardiovascular diseases: A machine learning approach. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*, 201–211. [[CrossRef](#)]
12. Zhang, W.; Li, D.S.; Bui-Thanh, T.; Sacks, M.S. Simulation of the 3D hyperelastic behavior of ventricular myocardium using a finite-element based neural-network approach. *Comput. Methods Appl. Mech. Eng.* **2022**, *394*, 114871. [[CrossRef](#)]
13. Shanmugam, D.; Blalock, D.W.; Gong, J.J.; Guttig, J.V. Multiple Instance Learning for ECG Risk Stratification. *arXiv* **2018**, arXiv:1812.00475. <https://doi.org/10.48550/arXiv.1812.00475>.
14. Ayano, Y.M.; Schwenker, F.; Dufera, B.D.; Debele, T.G. Interpretable Machine Learning Techniques in ECG-Based Heart Disease Classification: A Systematic Review. *Diagnostics* **2022**, *13*, 111. [[CrossRef](#)] [[PubMed](#)]
15. Sinha, N.; Sinha, S.; Valero, C.; Schäffer, A.A.; Aldape, K.; Litchfield, K.; Chan, T.A.; Morris, L.G.; Ruppig, E. Immune Determinants of the Association between Tumor Mutational Burden and Immunotherapy Response across Cancer Types. *Cancer Res.* **2022**, *82*, 2076–2083. [[CrossRef](#)] [[PubMed](#)]
16. Toma, M.; Singh-Gryzbon, S.; Frankini, E.; Wei, Z.A.; Yoganathan, A.P. Clinical Impact of Computational Heart Valve Models. *Materials* **2022**, *15*, 3302. [[CrossRef](#)]
17. Lustig-Barzelay, Y.; Sher, I.; Sharvit-Ginon, I.; Feldman, Y.; Mrejen, M.; Dallaseh, S.; Livny, A.; Beeri, M.S.; Weller, A.; Ravona-Springer, R.; et al. Machine learning for comprehensive prediction of high risk for Alzheimer’s disease based on chromatic pupilloperimetry. *Sci. Rep.* **2022**, *12*, 9945. [[CrossRef](#)]
18. Timmers, P.R.; Mounier, N.; Lall, K.; Fischer, K.; Ning, Z.; Feng, X.; Bretherick, A.D.; Clark, D.W.; Agbessi, M.; Ahsan, H.; et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **2019**, *8*, e39856. [[CrossRef](#)]
19. Osorio, J.A.; Scheer, J.K.; Ames, C.P. Predictive modeling of complications. *Curr. Rev. Musculoskelet. Med.* **2016**, *9*, 333–337. [[CrossRef](#)]
20. Malik, A.T.; Khan, S.N. Predictive modeling in spine surgery. *Ann. Transl. Med.* **2019**, *7*, S173. [[CrossRef](#)]
21. Romiyi, P.; Ding, K.; Dejam, D.; Franks, A.; Ng, E.; Preet, K.; Tucker, A.M.; Niu, T.; Nagasawa, D.T.; Rahman, S.; et al. Systematic review and evaluation of predictive modeling algorithms in spinal surgeries. *J. Neurol. Sci.* **2021**, *420*, 117184. [[CrossRef](#)]
22. Joshi, R.S.; Lau, D.; Scheer, J.K.; Serra-Burriel, M.; Vila-Casademunt, A.; Bess, S.; Smith, J.S.; Pellise, F.; Ames, C.P. State-of-the-art reviews predictive modeling in adult spinal deformity: Applications of advanced analytics. *Spine Deform.* **2021**, *9*, 1223–1239. [[CrossRef](#)]
23. Silva, T.D.; Vedula, S.S.; Perdomo-Pantoja, A.; Vijayan, R.; Doerr, S.A.; Uneri, A.; Han, R.; Ketcha, M.D.; Skolasky, R.L.; Witham, T.; et al. SpineCloud: Image analytics for predictive modeling of spine surgery outcomes. *J. Med. Imaging* **2020**, *7*, 1. [[CrossRef](#)] [[PubMed](#)]
24. Morris, B.N.; Sheehan, M.K.; Royster, R.L. Predictive Modeling for Nonocclusive Mesenteric Ischemia. *J. Cardiothorac. Vasc. Anesth.* **2019**, *33*, 1298–1300. [[CrossRef](#)]
25. Cook, R.J.; Lee, K.A.; Lo, B.W.; Macdonald, R.L. Classical Regression and Predictive Modeling. *World Neurosurg.* **2022**, *161*, 251–264. [[CrossRef](#)] [[PubMed](#)]
26. Steimer, A.; Müller, M.; Schindler, K. Predictive modeling of EEG time series for evaluating surgery targets in epilepsy patients. *Hum. Brain Mapp.* **2017**, *38*, 2509–2531. [[CrossRef](#)]
27. Gaskin, G.L.; Pershing, S.; Cole, T.S.; Shah, N.H. Predictive Modeling of Risk Factors and Complications of Cataract Surgery. *Eur. J. Ophthalmol.* **2015**, *26*, 328–337. [[CrossRef](#)] [[PubMed](#)]
28. Hecht, G.; Slee, C.A.; Goodell, P.B.; Taylor, S.L.; Wolinsky, P.R. Predictive Modeling for Geriatric Hip Fracture Patients. *J. Am. Acad. Orthop. Surg.* **2019**, *27*, e293–e300. [[CrossRef](#)] [[PubMed](#)]
29. Qiu, R.; Jia, Y.; Wang, F.; Divakarmurthy, P.; Vinod, S.; Sabir, B.; Hadzikadic, M. Predictive modeling of the total joint replacement surgery risk: A deep learning based approach with claims data. *AMIA Summits Transl. Sci. Proc.* **2019**, *2019*, 562–571.
30. Passias, P.G.; Ahmad, W.; Oh, C.; Imbo, B.; Naessig, S.; Pierce, K.; Lafage, V.; Lafage, R.; Hamilton, D.K.; Protosaltis, T.S.; et al. Development of Risk Stratification Predictive Models for Cervical Deformity Surgery. *Neurosurgery* **2022**, *91*, 928–935. [[CrossRef](#)]
31. Osterman, C.K.; Sanoff, H.K.; Wood, W.A.; Fasold, M.; Lafata, J.E. Predictive Modeling for Adverse Events and Risk Stratification Programs for People Receiving Cancer Treatment. *JCO Oncol. Pract.* **2022**, *18*, 127–136. [[CrossRef](#)]
32. Panja, S.; Rahem, S.; Chu, C.J.; Mitrofanova, A. Big Data to Knowledge: Application of Machine Learning to Predictive Modeling of Therapeutic Response in Cancer. *Curr. Genom.* **2021**, *22*, 244–266. [[CrossRef](#)]

33. Bendifallah, S.; Darai, E.; Ballester, M. Predictive Modeling: A New Paradigm for Managing Endometrial Cancer. *Ann. Surg. Oncol.* **2015**, *23*, 975–988. [[CrossRef](#)]
34. Gore, S.; Chougule, T.; Jagtap, J.; Saini, J.; Ingahalikar, M. A Review of Radiomics and Deep Predictive Modeling in Glioma Characterization. *Acad. Radiol.* **2021**, *28*, 1599–1621. [[CrossRef](#)]
35. Abdalvand, N.; Sadeghi, M.; Mahdavi, S.R.; Abdollahi, H.; Qasempour, Y.; Mohammadian, F.; Birgani, M.J.T.; Hosseini, K. Brachytherapy outcome modeling in cervical cancer patients: A predictive machine learning study on patient-specific clinical, physical and dosimetric parameters. *Brachytherapy* **2022**, *21*, 769–782. [[CrossRef](#)] [[PubMed](#)]
36. Nunez, J.J.; Leung, B.; Ho, C.; Bates, A.T.; Ng, R.T. Predicting the Survival of Patients With Cancer From Their Initial Oncology Consultation Document Using Natural Language Processing. *JAMA Netw. Open* **2023**, *6*, e230813. [[CrossRef](#)] [[PubMed](#)]
37. Chen, W.; Liu, X.; Zhang, S.; Chen, S. Artificial intelligence for drug discovery: Resources, methods, and applications. *Mol. Ther.-Nucleic Acids* **2023**, *31*, 691–702. [[CrossRef](#)]
38. Kolluri, S.; Lin, J.; Liu, R.; Zhang, Y.; Zhang, W. Machine Learning and Artificial Intelligence in Pharmaceutical Research and Development: A Review. *AAPS J.* **2022**, *24*, 19. [[CrossRef](#)]
39. Xue, H.T.; Stanley-Baker, M.; Kong, A.W.K.; Li, H.L.; Goh, W.W.B. Data considerations for predictive modeling applied to the discovery of bioactive natural products. *Drug Discov. Today* **2022**, *27*, 2235–2243. [[CrossRef](#)]
40. Alvarsson, J.; McShane, S.A.; Norinder, U.; Spjuth, O. Predicting with Confidence: Using Conformal Prediction in Drug Discovery. *J. Pharm. Sci.* **2021**, *110*, 42–49. [[CrossRef](#)]
41. Oguike, O.E.; Ugwuishiwu, C.H.; Asogwa, C.N.; Nnadi, C.O.; Obonga, W.O.; Attama, A.A. Systematic review on the application of machine learning to quantitative structure–activity relationship modeling against *Plasmodium falciparum*. *Mol. Divers.* **2022**, *26*, 3447–3462. [[CrossRef](#)] [[PubMed](#)]
42. Kovalchuk, S.V.; Kopanitsa, G.D.; Derevitskii, I.V.; Matveev, G.A.; Savitskaya, D.A. Three-stage intelligent support of clinical decision making for higher trust, validity, and explainability. *J. Biomed. Inform.* **2022**, *127*, 104013. [[CrossRef](#)] [[PubMed](#)]
43. Cai, T.; Abbu, K.A.; Liu, Y.; Xie, L. DeepREAL: A deep learning powered multi-scale modeling framework for predicting out-of-distribution ligand-induced GPCR activity. *Bioinformatics* **2022**, *38*, 2561–2570. [[CrossRef](#)] [[PubMed](#)]

**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.