Prediction Models in Nursing: A Survey on Risk Assessment and Clinical Decision Support

www.surveyx.cn

Abstract

This survey paper explores the transformative role of prediction models in nursing, focusing on their application in risk assessment and clinical decision support. By integrating advanced machine learning techniques with traditional statistical methods, these models enhance predictive accuracy and interpretability, thereby improving patient care. The paper underscores the importance of data integration and quality, highlighting efforts to combine diverse sources such as electronic health records, genomic data, and imaging modalities. These integrations provide a comprehensive view of patient health, leading to precise risk assessments and personalized interventions. Methodological advancements, including dynamic and adaptive modeling, ensure models remain relevant in evolving healthcare environments by responding to real-time data and adapting to patient conditions. The survey identifies challenges such as data quality, model interpretability, and integration into healthcare systems, emphasizing the need for robust methodologies and real-world validation. Future directions include leveraging emerging technologies like natural language processing and enhancing data integration for more comprehensive risk assessments. Overall, prediction models are pivotal in advancing nursing practice, supporting data-driven clinical decision-making, and ultimately improving patient outcomes and healthcare delivery efficiency.

1 Introduction

1.1 Significance of Prediction Models in Nursing

Prediction models are vital in nursing, facilitating precise risk assessments and informed clinical decision-making to enhance patient care. These models estimate risks associated with clinical events, enabling tailored interventions that improve outcomes. For example, forecasting unplanned patient readmission risks exemplifies their role in optimizing nursing interventions and resource allocation [1]. Furthermore, integrating machine learning with electronic health records (EHRs) has shown potential in reducing healthcare costs and enhancing care quality [2].

In cancer prognosis, prediction models are critical for forecasting therapeutic responses and stratifying patients for personalized treatment plans [3]. The challenges posed by current predictive models for castration-resistant prostate cancer (CRPC), particularly their complexity and poor generalization capabilities, underscore the need for refined methodologies [4]. Additionally, combining recursive feature elimination with gradient boosting has improved heart disease prediction, showcasing advanced machine learning techniques' potential for enhancing model accuracy [5].

The increasing prevalence of prediction models in healthcare necessitates benchmarks for their evaluation, particularly for mortality predictions in COVID-19 patients, to ensure reliability and effectiveness in clinical decision-making. Developing comprehensive methodologies for designing risk assessment services that incorporate prediction models is essential for providing structured development roadmaps for healthcare practitioners [6].

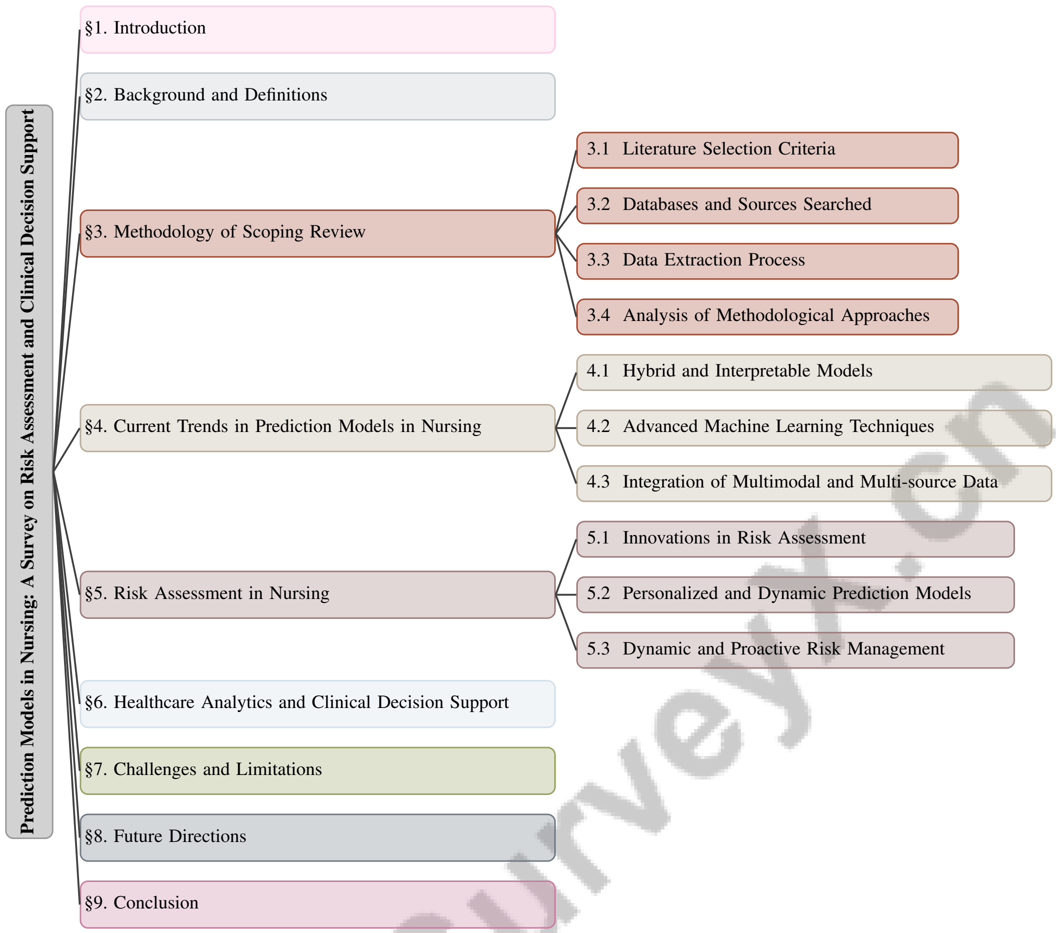


Figure 1: chapter structure

Prediction models are indispensable in nursing practice, supporting risk assessment and enhancing clinical decision-making, ultimately improving patient outcomes across diverse healthcare settings. By employing advanced machine learning techniques, such as patient representation models inspired by natural language processing and stratified sampling designs, healthcare providers can create tailored interventions that enhance both effectiveness and efficiency. These models utilize comprehensive data from EHRs to improve clinical prediction accuracy, especially in scenarios with limited patient records. Moreover, integrating personalized treatment decision-making strategies, evaluated through decision curve analysis, allows for precise matching of interventions to individual patient needs, leading to significant improvements in healthcare delivery and outcomes [2, 7, 8, 9].

1.2 Relevance of Healthcare Analytics

Healthcare analytics are pivotal in refining clinical decision-making and enhancing patient outcomes. Advanced data mining techniques are crucial for identifying subtle patterns in clinical indicators that traditional methods may overlook. For instance, data mining facilitates early detection of chronic kidney disease (CKD) by uncovering nuanced patterns in CKD indicators [10]. Similarly, machine learning applications analyzing biomarkers from blood samples have shown promise in predicting mortality risk for COVID-19 patients, aiding effective triage and management.

Innovative ensemble feature selection models that integrate statistical, deep, and optimally selected features have demonstrated enhanced predictive power in disease prediction, further highlighting healthcare analytics' transformative potential [11]. Incorporating various data modalities into survival prediction models also enhances their predictive capabilities, benefiting both clinical research and practice by providing comprehensive insights into patient health [3].

AI-generated, for reference only.

In addition to improving predictive accuracy, healthcare analytics facilitate rigorous performance evaluations of prediction models. Establishing benchmarks for model evaluation is essential for ensuring reliability and effectiveness in clinical settings, as evidenced by the need for robust methodologies in COVID-19 mortality prediction [12]. The complexity of measuring trust in prediction models, influenced by factors such as domain expertise and the interpretation of visualized data, underscores the importance of analytics in fostering confidence in these tools [13].

Healthcare analytics significantly enhance contemporary clinical practice by providing advanced tools and methodologies that improve decision-making processes. These analytics support the development of predictive models, such as those for clinical workload in patient-centered medical homes (PCMH) and readmission risks for psychiatric patients, optimizing resource allocation and improving care continuity. By integrating complex data from EHRs and employing sophisticated statistical techniques, healthcare analytics yield more precise predictions of patient needs and outcomes, ultimately enhancing patient care and reducing costs associated with readmissions and inefficiencies [14, 9]. By enabling precise risk assessments and personalized interventions, healthcare analytics ensure that clinical decisions are data-driven and patient-centered.

1.3 Structure of the Survey

This survey is systematically structured to comprehensively explore prediction models in nursing, focusing on their application in risk assessment and clinical decision support. It begins with an introduction emphasizing the significance of prediction models in enhancing nursing practice and patient care, followed by a discussion on healthcare analytics' pivotal role in refining clinical decision-making and improving patient outcomes. The background section provides a detailed overview of key concepts, including prediction models, nursing, risk assessment, healthcare analytics, and clinical decision support, elucidating their interrelations and importance in the nursing context.

The methodology section outlines the scoping review process, detailing literature selection criteria, databases searched, and data extraction and analysis procedures. This is followed by an examination of current trends in prediction models within nursing, highlighting advancements in hybrid and interpretable models, the use of advanced machine learning techniques, and the integration of multimodal and multi-source data for enhanced prediction accuracy.

The survey explores the critical role of prediction models in risk assessment, highlighting innovative methodologies and tools that enhance their effectiveness. It examines the integration of personalized and dynamic models tailored to individual risk profiles and strategies for implementing dynamic and proactive risk management approaches. Additionally, it discusses developing a comprehensive roadmap for creating risk assessment services, including stages such as requirements analysis, model validation, and iterative deployment, ensuring that models are accurate and trusted by users, influenced by factors like domain expertise and individual experience with predictive modeling [6, 13]. The integration of healthcare analytics with clinical decision support systems is explored, focusing on incorporating EHR data, applying machine learning models in clinical settings, and the role of visual data in enhancing predictive models.

Implementing prediction models in nursing faces significant challenges and limitations, including data quality and availability issues that hinder model training and accuracy. The complexity and interpretability of these models pose barriers to practical application in clinical settings, as healthcare providers may struggle to understand and trust their predictions. Furthermore, integrating these models into existing healthcare systems is complicated by ethical considerations, such as ensuring patient autonomy and informed consent, alongside the necessity for robust calibration to prevent misleading predictions that could adversely affect patient outcomes [2, 15, 16, 17, 14]. The survey concludes by suggesting potential future directions for research and development, discussing emerging technologies, data integration and quality enhancement, and methodological advancements that could further enhance prediction models in nursing.The following sections are organized as shown in Figure 1.

2 Background and Definitions

2.1 Interrelation of Key Concepts

The integration of prediction models, nursing, risk assessment, healthcare analytics, and clinical decision support systems is fundamental to advancing patient care and optimizing healthcare delivery. In nursing, prediction models are pivotal tools for estimating clinical risks and guiding decisions essential for personalized care. These models primarily rely on electronic health records (EHRs) as data sources; however, issues like data discontinuity and bias necessitate robust methodologies to ensure their reliability in clinical contexts [18].

Prediction models in risk assessment enable the identification and stratification of patient risks, facilitating proactive interventions. Nevertheless, the development of accurate classification models for rare clinical outcomes is often challenged by insufficient labeled data for training machine learning algorithms [8]. Moreover, varying data distributions between source and target populations complicate the direct application of these models, underscoring the need for benchmarks to validate model applicability across diverse clinical settings.

Healthcare analytics enhance the predictive power of these models by employing advanced data analysis techniques to uncover patterns within clinical and genomic data. Integrating heterogeneous data sources, such as clinical indicators and genomic information, is crucial for improving disease prediction accuracy and addressing limitations of traditional methods that may oversimplify data through pooling [3]. Causal inference frameworks, including directed acyclic graphs, offer a theoretical foundation for understanding causal relationships within clinical data, thus enhancing the interpretability and effectiveness of prediction models [19].

In clinical decision support systems, prediction models provide evidence-based insights that inform clinical decisions and improve patient outcomes. The interpretability challenge of complex machine learning models is significant in nursing, where understanding decision-making processes is vital for effective patient care [20]. Factors like model misspecification and variations in patient populations impact a model's discriminative ability, typically assessed by the Area Under the Curve (AUC), emphasizing the necessity for models that are both accurate and generalizable across diverse patient cohorts [21].

Predicting short-span Regular Health Behaviors (RHBs), such as medication adherence, presents challenges with current methods that struggle to scale for longer prediction horizons, crucial for timely interventions [22]. Thus, the interrelation of these key concepts underscores their collective importance in nursing, where the integration of prediction models, risk assessment, healthcare analytics, and clinical decision support systems fosters a data-driven approach to patient care, ultimately enhancing healthcare outcomes.

3 Methodology of Scoping Review

Category	Feature	Method
Literature Selection Criteria	Diversity and Modality Focus	PRRPM[1], D-RM[23], SynDI[24], ATRS[25]
Databases and Sources Searched	Collaborative Data Integration	MHBF[26], AG[27]
Data Extraction Process	Temporal Data Processing	DPSM[28]
	Imaging and High-Dimensional Analysis	CNN[29], EOCSA[30]
	Data Source Integration	IMDPM[31], KML[32], DL-DRAM[33], SR[34]
	Sampling and Stratification	SGS[8], INDT-ML[35]
Analysis of Methodological Approaches	Reliability and Calibration	ASC[36]
	Integration and Synergy Techniques	SPAPPM[37], CM-MMF[3]
	Generalization and Robustness	Multi-DANN[38], TPM[39]
	Model Simplification and Optimization	VSS[40]

Table 1: This table presents a comprehensive overview of the methodological approaches analyzed in the scoping review for prediction models in nursing and healthcare analytics. It categorizes the strategies into literature selection criteria, databases and sources searched, data extraction processes, and analysis of methodological approaches, highlighting specific features and methods applied in each category. The table serves as a detailed reference for understanding the diverse techniques employed to enhance predictive accuracy and clinical relevance in healthcare settings.

In conducting a scoping review, a structured framework is vital for systematically evaluating the existing literature. This framework defines the review's scope and establishes criteria for selecting pertinent studies, ensuring a comprehensive representation of prediction models in nursing, with a focus on risk assessment and healthcare analytics. Table 1 provides a detailed summary of the methodological approaches used in the scoping review, categorizing them into key areas such as literature selection criteria, data sources, data extraction processes, and analysis of methodological approaches. Additionally, Table 2 provides a comprehensive summary of the methodological frameworks utilized in the scoping review, detailing the literature selection criteria, databases searched, and data extraction processes, which are pivotal in evaluating prediction models in nursing and healthcare analytics. **??** illustrates the structured methodology of a scoping review, highlighting key areas such as literature selection criteria, databases and sources searched, data extraction processes, and analysis of methodological approaches. Each section is further detailed with specific techniques and strategies employed to enhance prediction models in nursing and healthcare analytics, thereby reinforcing the significance of a well-defined framework in guiding the review process.

3.1 Literature Selection Criteria

The selection criteria for this scoping review were crafted to include a broad spectrum of studies on prediction models in nursing, emphasizing risk assessment and healthcare analytics. A key criterion involved studies using diversity-based selection techniques to address class imbalance in clinical datasets, ensuring minority classes are accurately represented while maintaining majority class characteristics [23]. The review also prioritized studies integrating synthetic data from external models to enhance regression estimates and predictions, as demonstrated by the SynDI framework [24]. This highlights the importance of leveraging both internal and external data for model robustness.

As illustrated in Figure 2, the literature selection criteria encompass these domains, focusing on diversity-based selection, data integration techniques, and hybrid multimodal approaches. Additionally, studies employing hybrid approaches like the Adaptive Treatment Recommendation System (ATRS), which utilizes cluster-classification for personalized treatment, were included to illustrate patient data integration for tailored healthcare solutions [25]. The selection also embraced literature on cross-modality attention-based multimodal fusion techniques, combining image and RNA-seq data to improve patient survival predictions [3].

The criteria emphasized retrospective data from clinical environments, such as COVID-19 care centers, which collected comprehensive clinical parameters [41]. Benchmarks for determining the minimum sample size required for precise evaluation in external studies were also considered to ensure model generalizability and reliability [12]. Studies utilizing extensive datasets from multiple hospitals, focusing on diverse patient demographics, were included to provide a comprehensive understanding of prediction model effectiveness across varied healthcare settings [1]. This thorough set of criteria facilitated the inclusion of literature that advances theoretical knowledge while enhancing the practical application of prediction models in nursing and healthcare analytics.

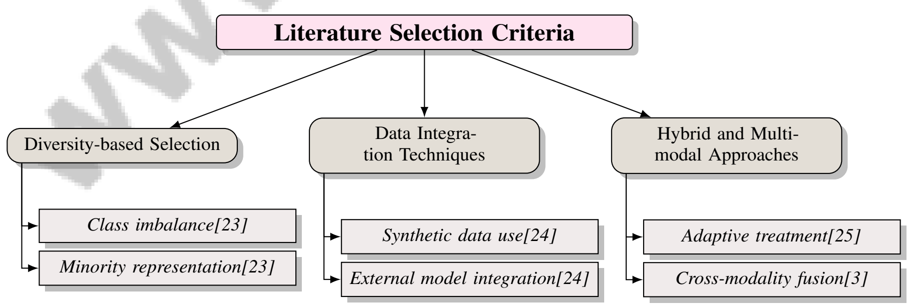


Figure 2: This figure illustrates the literature selection criteria for prediction models in nursing, focusing on diversity-based selection, data integration techniques, and hybrid multimodal approaches.

3.2 Databases and Sources Searched

The review employed a comprehensive search strategy across multiple databases to ensure thorough literature collection on prediction models in nursing. Key databases included PubMed, CINAHL, and Scopus, chosen for their extensive coverage of healthcare and nursing research, supporting the development and validation of clinical prediction models and enhancing risk factor understanding

[14, 16, 40]. These databases were instrumental in identifying studies focusing on integrating prediction models with clinical decision support systems.

Additionally, searches in IEEE Xplore and ACM Digital Library captured literature on healthcare analytics and machine learning applications in nursing, offering insights into variable selection strategies and treatment considerations crucial for enhancing prediction models' accuracy [40, 42]. The review process was enriched by examining datasets from multi-institutional collaborations, such as outpatient data from 888 facilities, which facilitated a comprehensive understanding of prediction model performance across diverse healthcare environments [26].

Incorporating datasets from various sources, including 155 videos from 89 participants across 11 medical sites, underscored the importance of multi-institutional collaboration in data collection for prediction models [27]. This diverse dataset contributed to the robustness and generalizability of the review findings, highlighting varied data sources' critical role in advancing prediction model research. The careful selection of diverse databases and sources facilitated a comprehensive literature review, effectively encompassing the extensive range of current research on prediction models in nursing, particularly their applications in risk assessment and clinical decision support [14, 43, 40, 44].

3.3 Data Extraction Process

The data extraction process for this scoping review employed a systematic approach to gather relevant information from selected literature. Initiating with data cleaning and standardization, techniques like Chi-square and Principal Component Analysis were used for feature selection, enhancing data robustness [35]. The process integrated diverse data sources, such as combining ClinVar clinical data with gene expression data to train a k-Nearest Neighbours classifier for disease outcome prediction [31].

Surrogate-guided sampling (SGS) was implemented to stratify cohorts based on surrogate variables, maximizing true outcome case abstraction and improving data representativeness [8]. In applications like ICU delirium prediction, time-series data from environmental sensors were utilized, demonstrating non-traditional data sources' utility in enhancing model predictions [33]. For COVID-19 studies, daily case data and demographics from the Johns Hopkins University repository provided a comprehensive dataset for pandemic modeling [32].

Advanced imaging data analysis, such as extracting patches from whole slide images and clustering them using K-means for training the DeepConvAttentionSurv model, highlighted spatial data's importance in prognosis prediction [30]. Similarly, pre-treatment DCE-MRI images analyzed using a multi-input convolutional neural network underscored high-dimensional imaging data's role in clinical predictions [29].

Moreover, modeling increments of prevalent ICU cases as a function of past observations, optimized via gradient descent, illustrated time-series modeling's integration in prediction frameworks [28]. Throughout the extraction process, data on model performance, methodology, and risk of bias were meticulously documented, ensuring rigorous evaluation of predictive models [34]. This comprehensive data extraction approach enabled in-depth analysis of prediction models in nursing, ensuring that the extracted data were representative and suitable for robust evaluation and application, enhancing predictive accuracy and clinical relevance [43, 14, 2, 40].

3.4 Analysis of Methodological Approaches

The scoping review explores diverse methodological approaches significantly contributing to developing and refining prediction models within nursing and healthcare analytics. Cross-modality attention-based techniques, which integrate various data modalities to construct a unified representation, enhance survival prediction accuracy by leveraging multiple data sources' strengths, such as clinical indicators and genomic data [3].

The combination of logistic regression with multilayer perceptron (MLP) algorithms exemplifies the synergy between linear and nonlinear modeling techniques, improving predictive accuracy for conditions like non-specific low back pain (NSLBP) by capturing intricate relationships within clinical datasets [37]. Deep transfer learning methods, like the Multi-DANN approach, demonstrate adversarial training's potential to enhance model generalizability across diverse clinical environments, such as predicting COVID-19 patient outcomes in emergency room settings [38].

Inverse-odds weighting addresses covariate distribution differences to improve prediction models' transportability, ensuring accuracy and reliability across heterogeneous patient groups [39]. Variable selection strategies optimize model performance while minimizing complexity, resulting in parsimonious models that retain predictive power [40]. The Adaptive Score-based CUSUM method detects calibration issues within machine learning algorithms, ensuring models are well-calibrated and reliable across different patient cohorts [36].

Projection predictive inference offers a structured workflow for model selection and validation, facilitating systematic evaluation and refinement of prediction models [45]. The integration of advanced methodological approaches, such as variable selection strategies, multinomial prediction models, and innovative patient representation techniques, signifies substantial progress in developing and applying prediction models in nursing, enhancing their accuracy and effectiveness in identifying at-risk patients and informing clinical decision-making [2, 40, 43, 44, 14]. By integrating diverse techniques and leveraging cutting-edge technologies, these methodologies contribute to developing robust, reliable, and clinically relevant prediction models, ultimately enhancing risk assessment and clinical decision support in healthcare settings.

Feature	Literature Selection Criteria	Databases and Sources Searched	Data Extraction Process
Data Sources	Clinical Datasets	Multiple Databases	Diverse Data Sources
Selection Technique	Diversity-based Selection	Comprehensive Search Strategy	Chi-square, Pca
Integration Approach	Hybrid Multimodal Approaches	Multi-institutional Collaboration	Surrogate-guided Sampling

Table 2: This table presents a comparative analysis of the methodologies employed in a scoping review focused on prediction models in nursing. It categorizes the approaches into literature selection criteria, data sources, data extraction processes, and integration techniques, highlighting diverse strategies used to enhance prediction model development and application in healthcare analytics.

4 Current Trends in Prediction Models in Nursing

4.1 Hybrid and Interpretable Models

Recent progress in hybrid and interpretable models has significantly enhanced the predictive power and clinical utility of nursing prediction models. Hybrid models, which integrate diverse data sources and methodologies, are increasingly pivotal in healthcare analytics. The SynDI framework, for instance, generates synthetic data from multiple external models to accommodate heterogeneous covariate effects, thereby bolstering prediction model robustness [24]. Similarly, the Adaptive Treatment Recommendation System (ATRS) uses a hybrid cluster-classification approach to tailor treatment plans based on real-world patient data, highlighting hybrid models' potential in personalized healthcare [25].

In cancer prognosis, the PR-NET model optimizes network architecture to predict prostate cancer outcomes by focusing on key genetic loci, illustrating pathway-refined models' ability to enhance clinical predictions [4]. Moreover, integrating patient history, demographics, and clinical characteristics into predictive risk models, as demonstrated by Choudhury et al., shows advancements in assessing patient readmission risks [1].

Interpretable models have also advanced, emphasizing transparency and clinician trust. The Meta-ANOVA method enhances the interpretability of complex models, making it suitable for nursing applications requiring clarity [20]. Explainable AI (XAI) methods further elucidate socioeconomic factors' impact on COVID-19 mortality, reinforcing interpretability's role in fostering model reliability and clinician adoption [46].

Innovations in uncertainty quantification further strengthen model robustness. The adaptive CUSUM test, combined with sample-splitting and cross-validation, enhances power while controlling Type I errors, showcasing an innovative approach to ensuring model reliability across diverse populations [36]. The ActSafe model exemplifies the transformative potential of hybrid models by integrating multimodal health data to predict medical treatment compliance violations, significantly improving prediction accuracy [22].

These advancements demonstrate the transformative potential of hybrid and interpretable models in nursing, leveraging innovations such as patient representation learning from electronic health records (EHRs), multitask learning with multimodal data, and enhanced sampling designs for rare outcomes. These approaches improve clinical prediction accuracy, showing up to a 19% performance improvement in data-scarce scenarios, and facilitate personalized patient care strategies by enabling comprehensive health status assessments and optimizing risk predictions for outcomes like hospital readmissions [2, 8, 47, 1, 14]. By integrating diverse data sources and leveraging cutting-edge technologies, these models enhance risk assessment and clinical decision support, ultimately improving patient outcomes.

4.2 Advanced Machine Learning Techniques

Advanced machine learning techniques have revolutionized nursing prediction models by enhancing predictive accuracy and interpretability. Multi-input convolutional neural networks have improved prediction models for breast cancer treatment outcomes, as demonstrated by Braman et al., by utilizing high-dimensional imaging data to refine treatment predictions, showcasing deep learning's potential in clinical applications [29].

The use of clinical language models to create fixed-length representations of patient data has shown promise in enhancing clinical prediction model accuracy [2]. This method standardizes data inputs, facilitating patient data reuse across diverse predictive models and improving robustness and consistency.

Multinomial logistic regression (MLR) effectively handles complex relationships between predictors and categorical outcomes, underscoring traditional statistical techniques' utility in modern predictive modeling [44]. Incorporating diversity-based selection methods atop common re-sampling techniques has enhanced the detection of minority events in mortality risk models, addressing class imbalance issues that often hinder predictive accuracy [23].

In cardiovascular disease prediction, the Recursive Feature Elimination with Gradient Boosting (RFE-GB) method has demonstrated superior accuracy and consistency compared to existing algorithms, illustrating the potential of combining feature selection with ensemble learning techniques [5]. Optimizing the input layer of deep learning models, as seen in the PR-NET approach, further exemplifies advancements in model efficiency and accuracy [4].

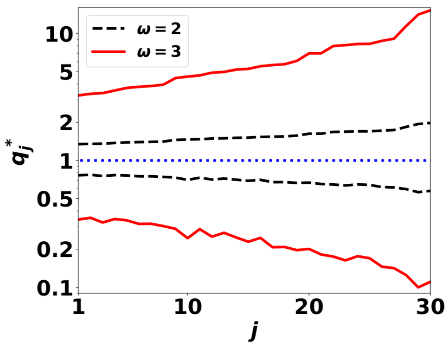
Enhanced ResNet architectures with squeeze-and-excitation blocks have improved modeling of spatial relationships in ECG data, enabling more accurate predictions of cardiac conditions [48]. The MTNet model, employing a multi-task learning framework, combines classification with auxiliary anomaly ranking tasks to improve prediction accuracy for Major Depressive Disorder [49].

These advancements in machine learning techniques, particularly through patient representation learning inspired by natural language processing, significantly enhance nursing models' predictive capabilities. Evidence suggests an average improvement of 3.5% in AUROC across various prediction tasks, with up to a 19% enhancement when training data is limited, underscoring machine learning's potential to refine clinical prediction models beyond traditional logistic regression, despite some studies indicating no performance advantage in certain contexts [2, 34].

As illustrated in Figure 3, advanced machine learning techniques are pivotal in exploring current trends in nursing prediction models. The first figure highlights the nuanced adjustments in prediction accuracy achievable through parameter tuning, while the second figure depicts the intricate architecture of neural networks utilized in medical predictions. Together, these figures provide a compelling overview of how advanced machine learning techniques are revolutionizing predictive modeling in nursing, enabling informed decision-making and improved patient outcomes.

4.3 Integration of Multimodal and Multi-source Data

The integration of multimodal and multi-source data marks a significant advancement in nursing prediction models, enhancing predictive accuracy and clinical applicability. By incorporating various data types, including electronic health records (EHRs) and clinical text data, these models create a nuanced understanding of patient health. This multifaceted approach improves clinical prediction model accuracy, evidenced by a mean increase of 3.5% in AUROC across five prediction tasks, and leads to more personalized predictions, particularly when training data is limited. Stratified sampling techniques facilitate the efficient identification of rare clinical outcomes, refining predictive capabilities and ensuring effective and resource-efficient training processes [2, 8].



(a) The image shows a graph with two lines representing different values of omega () and their corresponding qj values.[50]

(b) A Graphical Representation of a Neural Network[51]

Figure 3: Examples of Advanced Machine Learning Techniques

A key innovation in this domain is the HgbNet framework, which utilizes a NanDense layer for processing missing data and multiple attention mechanisms to capture data irregularities effectively [52]. This underscores the importance of robust data handling techniques in ensuring prediction models' reliability and accuracy, particularly with incomplete or irregular datasets.

Moreover, integrating multimodal data—combining clinical indicators, imaging data, and genomic information—allows for a more nuanced understanding of patient conditions. This comprehensive approach enables constructing models that are not only more accurate but also adaptable to real-world clinical complexities. By merging data from diverse sources, including individual-level data and external summary information, these models can uncover intricate patterns and correlations that may be overlooked when relying on a single data type. This enhances risk assessment accuracy and improves clinical decision support by leveraging insights from various predictors and populations, providing a more nuanced understanding of patient outcomes and optimizing predictive modeling efforts [14, 8, 24].

Integrating multi-source data fosters the development of robust predictive models that demonstrate consistent high performance across varied patient demographics and healthcare environments. Advanced techniques such as multitask learning leverage correlations among clinical tasks, while innovative patient representation methods drawn from natural language processing allow for effective utilization of incomplete data and improved accuracy in clinical predictions, even with limited patient records [2, 47]. This capability is crucial for ensuring prediction models remain applicable and effective in diverse clinical contexts, ultimately leading to improved patient outcomes and more efficient healthcare delivery.

5 Risk Assessment in Nursing

5.1 Innovations in Risk Assessment

Recent advancements in nursing risk assessment leverage predictive models and analytical techniques to enhance accuracy and reliability. The integration of multi-phase Dynamic Contrast-Enhanced Magnetic Resonance Imaging with deep learning has improved treatment response predictions in HER2-positive breast cancer patients [29]. Hybrid calibration techniques address individual biometric variations, emphasizing personalized risk predictions [53]. Moreover, novel methods for assessing moderate calibration enhance prediction reliability through advanced statistical approaches [54].

AI-generated, for reference only.

The Taylor method refines risk prediction by addressing under-correction in odds-ratio adjustments, improving statistical inference in clinical settings [55]. Cross-Modality Attention-Based Multimodal Fusion effectively integrates diverse data, significantly boosting prediction accuracy [3]. The MTNet model, utilizing longitudinal socio-demographic data, addresses challenges such as high within-class variance, underscoring the importance of longitudinal data for accurate risk assessments [49]. Integration of demographic features with deep learning models further enhances classification accuracy [48].

Calibration methods for discrete subdistribution hazard models show strong performance, even in high censoring scenarios, validating their effectiveness in risk assessments [56]. Advanced risk prediction models improve clinical decision-making through personalized data integration, enhancing disease prognosis and resource allocation. Localized risk models using regional EHR repositories have improved predictive performance, such as a 10.7% increase in ASCVD risk prediction accuracy for diabetic patients. Genetic Algorithms optimize readmission risk models, illustrating better forecasting of unplanned admissions, contributing to improved patient outcomes through timely interventions [57, 1, 16, 58, 6].

5.2 Personalized and Dynamic Prediction Models

Personalized and dynamic prediction models in nursing are transforming risk prediction by offering tailored healthcare solutions through individualized data integration and adaptive methodologies. Bayesian logistic regression (BLR) and its Markov variant (MarBLR) facilitate dynamic model updates, maintaining accuracy in changing clinical environments [59]. Multimodal graph learning techniques, such as MMGL, effectively utilize multi-modal data to improve disease prediction through enhanced model generalization [60]. Sequential prediction models align with clinical workflows, providing insights for intervention decisions [57].

Innovations in calibration and inference techniques enhance personalized prediction models. A unified 'bridge' test for calibration assessments offers robust model calibration tools [54]. An approximate equation for recovering conditional odds-ratio enhances calibration accuracy [55]. The PR-NET model focuses on relevant genetic loci, streamlining model complexity and improving efficiency [4]. CLMBR, a clinical language model, captures complex EHR data relationships, enhancing prediction accuracy even with limited training data [2].

Two-stage prediction models enable personalized treatment effect predictions based on individual baseline risk, optimizing treatment outcomes [61]. These advancements highlight the transformative potential of personalized and dynamic models in nursing, fostering individualized risk prediction and patient care through cutting-edge technologies and methodologies, ultimately improving patient outcomes.

5.3 Dynamic and Proactive Risk Management

Dynamic and proactive risk management in nursing relies on sophisticated prediction models and analytical frameworks, empowering healthcare providers to anticipate and mitigate risks effectively. Integrating real-time data analytics with predictive modeling enables continuous monitoring and timely interventions. Ambient intelligence applications, such as predicting ICU delirium incidence using environmental sensor data, illustrate the potential for dynamic risk assessment and proactive management strategies [33].

Adaptive machine learning models, like Bayesian logistic regression, dynamically adjust to new data inputs, maintaining accuracy in evolving clinical scenarios [59]. Sequential prediction frameworks, providing risk estimates at multiple decision points, align with clinical workflows, supporting timely interventions [57]. Advanced calibration techniques enhance prediction model reliability across diverse populations, with bridge tests improving risk prediction robustness [54]. Causal inference frameworks, such as directed acyclic graphs, enhance the interpretability and effectiveness of proactive risk management strategies [19].

Integrating multimodal and multi-source data into prediction models provides a comprehensive understanding of patient health, enabling precise risk assessments and timely interventions [3]. Leveraging diverse data types allows for dynamic risk management strategies responsive to patient care complexities, improving patient outcomes and healthcare delivery efficiency. Data-driven approaches to patient care enhance providers' capacity to predict and address risks, ensuring effective interventions while facilitating ongoing evaluations and decision adjustments as new data becomes available, optimizing patient outcomes over time [57, 2].

6 Healthcare Analytics and Clinical Decision Support

The integration of electronic health record (EHR) data into clinical decision support systems is pivotal for navigating healthcare analytics, enhancing predictive accuracy, and improving patient outcomes. This section delves into methodologies that leverage EHR data to bolster predictive models, highlighting frameworks and approaches that illustrate this transformative potential.

6.1 Integrating EHR Data for Enhanced Prediction

Incorporating EHR data into predictive models is crucial for advancing clinical decision support. EHRs provide comprehensive longitudinal data, enriching predictive accuracy and model applicability. The OneFlorida+ network, aggregating EHR and claims data from over 20 million individuals, exemplifies the power of large datasets in predictive analytics [62]. Federated learning methodologies facilitate collaborative model development across decentralized EHR datasets, enhancing model robustness while safeguarding patient privacy [16, 2, 63, 64]. By addressing privacy and scalability, federated learning boosts clinical decision-making accuracy.

Incorporating treatment effects into models is essential for clinical relevance, providing actionable insights for decision-making [42]. The K-Means-LSTM model leverages historical data patterns and demographics, enhancing EHR-integrated model predictions [32]. Techniques like modal-attentional feature fusion (MaFF) and adaptive graph learning (AGL) allow comprehensive EHR data analysis, improving disease prediction accuracy [60]. The SynDI framework integrates diverse external models with EHR data, enhancing prediction accuracy and enabling statistical inference across populations [24]. The Adaptive Treatment Recommendation System (ATRS) emphasizes the inclusion of disease-specific data to advance treatment recommendations [25].

6.2 Machine Learning Models in Clinical Settings

Machine learning models have revolutionized clinical decision-making by providing data-driven insights that enhance patient outcomes. These models offer real-time applications across medical domains, improving healthcare precision and efficiency. For instance, the Classification and Regression Trees (CART) model predicts hypoglycemic events in diabetes management, enabling timely interventions [65]. In COPD risk prediction, Bayesian Logistic Regression (BLR) and its Markov variant (MarBLR) outperform static models, showcasing dynamic learning advantages [59].

Integrating EHR data into machine learning models further enhances predictive capabilities. The Clinical Language Model for Biomedical Research (CLMBR) shows superior performance using de-identified EHR data, highlighting the potential of comprehensive patient data in refining predictive models [2]. Predictive models assessing readmission risk optimize resource allocation and improve patient management [1]. These models provide actionable insights through metrics like accuracy, sensitivity, and specificity, informing clinical decisions and enhancing care.

These examples demonstrate the transformative potential of machine learning models in clinical environments, enhancing predictive accuracy through advanced patient representation techniques and utilizing multimodal data for simultaneous predictions across healthcare tasks. This integration supports precise and efficient clinical decision-making [2, 15, 34, 8, 47]. By facilitating personalized, data-driven decision-making, these models contribute to improved patient outcomes and healthcare efficiency.

6.3 Integration of Visual Data and Predictive Models

Integrating visual data into predictive models enhances accuracy and applicability in clinical settings by leveraging multimodal EHR data and advanced patient representation techniques. This integration accommodates incomplete data inputs and identifies correlations across clinical tasks for simultaneous outcome predictions. Recent advancements, like the FlexCare model, demonstrate asynchronous single-task predictions and task-guided multimodal fusion, significantly improving performance.

Patient representation methods inspired by natural language processing achieve mean AUROC improvements of 3.5

Key advancements include deep learning techniques that extract features from high-dimensional visual data, such as medical images. The DeepConvAttentionSurv model uses whole slide images for prognosis prediction, integrating spatial data into predictive frameworks [30]. By leveraging CNNs and attention mechanisms, this approach captures intricate patterns within visual data, improving prediction accuracy.

The application of multi-input CNNs to pre-treatment DCE-MRI images refines treatment response predictions, supporting personalized plans [29]. Dynamic visual data integration, such as video analysis, offers opportunities for enhancing predictive models. Analyzing sensor data and video feeds through ambient intelligence shows promise in predicting ICU delirium risk, enabling proactive risk management [33]. This highlights the potential of combining visual data with other modalities for a comprehensive patient health view.

Developing methodologies that incorporate visual data aligns with the trend of multimodal data integration. By combining visual data with genomic and clinical information, predictive models enhance accuracy and robustness, improving clinical decision support and patient outcomes. Machine learning techniques applied to EMRs effectively classify rare outcomes, especially with enriched sampling strategies targeting auxiliary variables. Integrating genomics data into clinical predictions enhances multiclass disease prediction accuracy, achieving up to 73

The integration of visual data into predictive models represents a significant advancement in healthcare analytics, enhancing capabilities for risk assessment and clinical decision-making. Utilizing advanced machine learning techniques, particularly those inspired by natural language processing, these models improve the accuracy and comprehensiveness of patient health assessments derived from EHRs. This approach facilitates knowledge transfer from broader patient populations to improve predictions for specific clinical outcomes and supports personalized care strategies. Patient representation schemes have shown up to a 19

7 Challenges and Limitations

7.1 Data Quality and Availability

Predictive models in nursing heavily depend on data quality and availability, posing significant challenges. Historical data may not accurately reflect current patient conditions or healthcare practices, limiting model adaptability in dynamic clinical settings [1]. Insufficient clinical parameter collection, especially in mortality prediction models, further impairs accuracy, underscoring the need for comprehensive data acquisition [41].

Models like MTNet face challenges due to high within-class variance in depression samples and limited labeled data, affecting generalization across diverse populations [49]. Small training datasets compromise model robustness, necessitating larger, more diverse datasets for improved accuracy. Benchmark assumptions in simulations may not hold true in practice, affecting reliability by overlooking clinical setting variations [12]. The ActSafe model's lack of real-world deployment and user studies limits validation across diverse populations, emphasizing the need for practical testing to ensure efficacy [22].

Addressing these challenges requires enhancing data collection methodologies, improving interoperability, and ensuring high-quality dataset availability. Overcoming limitations posed by restricted patient data and the necessity for localized risk models enables healthcare providers to harness advanced machine learning techniques, such as knowledge-enhanced patient representation and stratified sampling designs, leading to more accurate predictive models and improved clinical outcomes. Employing natural language processing in electronic health records (EHRs) has resulted in significant accuracy improvements, with mean AUROC increases of up to 19

7.2 Model Interpretability and Complexity

Model interpretability and complexity present significant barriers to deploying predictive models in nursing. The 'black box' nature of many machine learning models, especially deep learning frameworks, complicates clinical decision-making due to a lack of transparency in prediction derivation

[66]. This opacity is problematic in healthcare, where understanding causal relationships between input variables and outcomes is crucial for clinician trust and adoption. Integrating multiple data modalities for survival predictions often requires substantial computational resources, which may not be readily available in all clinical settings. The computational intensity of processes like cluster selection can impact overall results and limit practical applicability [30].

Reliance on available data and the potential for biased estimates, particularly in single-arm studies, significantly limit model interpretability. These biases can skew results and undermine prediction reliability, necessitating robust methodologies that account for confounding factors [7]. The computational demands of implementing projection predictive inference in large datasets complicate matters, particularly when a well-defined reference model is required [45]. Observational data biases can obscure causality, necessitating careful consideration of confounding factors to ensure reliable outcomes [67]. Derived regret bounds that scale linearly with the number of variables may also limit certain methods' effectiveness in high-dimensional settings, underscoring the need for efficient computational techniques [59]. Models may fail to account for all variables affecting outcomes, such as underlying health conditions not included in the dataset, impacting mortality predictions [68].

Addressing these challenges requires developing models that balance complexity with interpretability, ensuring applicability and reliability across diverse clinical settings. Meta-ANOVA offers a promising approach by providing interpretable models without significant predictive performance loss, addressing some interpretability challenges [20]. Enhancing transparency and reducing computational demands can render prediction models more accessible and useful for clinical decision support in nursing.

7.3 Integration into Healthcare Systems

Integrating prediction models into existing healthcare systems presents several challenges. Achieving seamless interoperability between predictive models and diverse healthcare IT infrastructures, which often vary across institutions, is a primary difficulty. This lack of standardization can impede the efficient transfer and utilization of predictive insights, limiting the models' potential to enhance clinical decision-making [26].

The integration process is further complicated by the need to align prediction models with existing clinical workflows, often requiring substantial adjustments to accommodate new technologies. Clinician adaptation to these changes is essential, as successful implementation heavily relies on acceptance and trust. The complexity of these models, often perceived as 'black boxes', can hinder adoption, as clinicians may hesitate to rely on systems they do not fully understand [66].

Another challenge is integrating prediction models with EHRs, a critical data source for these models. Variability in EHR data formats and structures complicates extraction and integration processes, necessitating sophisticated pre-processing and transformation techniques to ensure compatibility and accuracy [62]. The requirement for real-time data processing to deliver timely insights complicates integration, demanding robust computational infrastructure and efficient data handling capabilities.

Data privacy and security issues must also be addressed during model integration, particularly concerning sensitive patient information. Compliance with regulatory standards, such as the Health Insurance Portability and Accountability Act (HIPAA), is crucial to protect patient confidentiality and maintain trust in predictive technologies [64].

Successful integration of prediction models into healthcare systems necessitates addressing these technological, organizational, and regulatory challenges. Promoting collaboration among technology developers, healthcare providers, and policymakers can significantly enhance the integration of localized risk prediction models. This collaborative approach addresses the limitations of existing national models when applied to specific populations, as evidenced by the development of a knowledge-enhanced ASCVD risk model that improved prediction accuracy from an AUC of 0.653 to 0.723. Leveraging advanced patient representation techniques from natural language processing has improved model performance by an average of 3.5

8 Future Directions

8.1 Emerging Technologies and Innovations

Emerging technologies in machine learning and artificial intelligence are poised to enhance predictive models in nursing by improving their accuracy, adaptability, and clinical utility. Advanced patient representation techniques using natural language processing (NLP) have shown significant performance improvements in clinical prediction models, achieving up to 19

Future research should explore the application of the PR-NET model across various cancer types and optimize its clinical performance [4]. Enhancing deep learning models for cardiovascular disease risk prediction by incorporating larger datasets and additional clinical parameters remains essential for achieving comprehensive assessments [5]. Developing dynamic models that adapt to changes in patient populations and healthcare practices will refine calibration methodologies, ensuring predictions remain accurate over time [12].

Innovations like cross-modality attention-based multimodal fusion techniques and additional loss functions could enhance model robustness with complex data inputs. Investigating clinical language models' efficacy for predicting diverse outcomes, particularly with limited patient records, is vital. Strategies to minimize computational expenses during model training, such as stratified sampling and federated learning, can address challenges posed by rare outcomes in EHRs [2, 34, 8, 64, 14].

Improving data labeling and exploring domain adaptation techniques will enhance model robustness across various datasets [48]. Future studies should examine alternative variable selection techniques and incorporate healthcare cost measures to improve predictive capabilities [1]. Conducting user studies and expanding frameworks to predict medication violations are essential for advancing healthcare prediction models [22].

Integrating advanced methodologies, including multitask learning frameworks like FlexCare and novel ensemble feature selection algorithms, significantly enhances nursing prediction models' accuracy and effectiveness. These advancements promise to transform patient care and healthcare delivery by enabling more personalized and timely interventions [11, 2, 1, 47]. By leveraging cutting-edge technologies, these models can provide actionable insights, supporting more effective and personalized care strategies.

8.2 Data Integration and Quality Enhancement

Enhancing data integration and quality is crucial for improving nursing models' predictive capabilities across diverse clinical settings. Future research should develop methodologies to integrate disparate data sources, including EHRs, genomic data, and imaging modalities, to create a comprehensive view of patient health. Such integration is vital for increasing prediction models' accuracy and reliability, facilitating a holistic understanding of patient conditions and outcomes [41].

Refining variable selection methods is essential, particularly in high-dimensional data contexts where missing values and complex interactions can impact model performance. Combining clinical expertise with statistical methods will enhance models' robustness and predictive accuracy, aiding their integration into routine practice [44]. Developing methods to accurately estimate variance in predicted risks, such as applying the Taylor method across different models, is critical for precise risk assessments [55].

External validation is crucial before clinical application, ensuring machine learning models are accurate and generalizable across diverse populations and environments. Establishing robust sample size criteria for external validation, especially with threshold-based performance measures, and refining estimation strategies for sequential predictions are essential for aligning estimands with clinical needs [57, 69, 12, 44].

Improving data collection methods and addressing population-specific factors are vital for developing robust mortality prediction models in varied settings [41]. Incorporating clinical expertise and enhancing data quality will improve predictive capabilities and treatment recommendations of adaptive systems [25].

Refining calibration methods for subdistribution hazard models and applying them in complex survival analysis scenarios is another promising area for future exploration, enhancing prediction model reliability [56].

Collectively, these initiatives in data integration and quality enhancement are essential for advancing clinical prediction models in nursing. They ensure models deliver accurate insights, facilitating informed clinical decision-making and ultimately leading to improved patient outcomes. By employing effective variable selection strategies and leveraging EHR data continuity, these models can better identify at-risk patients and tailor interventions, enhancing healthcare delivery effectiveness [2, 16, 40, 62].

8.3 Methodological Advancements

Future research in nursing prediction models should prioritize methodological advancements that enhance model accuracy, interpretability, and applicability across varied healthcare settings. Developing more efficient computational techniques is crucial for managing the increasing complexity and volume of healthcare data. Extending bootstrap-based optimism correction methods, as explored by Noma et al., to other approaches can provide a robust framework for improving model reliability and reducing bias [70].

The expansion of model-based Receiver Operating Characteristic (mROC) methodology to various data types, including categorical and time-to-event data, presents significant opportunities for advancing model evaluation techniques. Developing non-parametric methods for assessing model calibration, as suggested by Sadatsafavi et al., could further enhance prediction model robustness, ensuring reliability across diverse clinical scenarios [71].

Integrating advanced machine learning techniques, such as deep learning and ensemble learning, with traditional statistical methods can create hybrid models that enhance predictive accuracy and interpretability. This is particularly important in healthcare, where accountability is crucial. For example, methods like Meta-ANOVA transform complex black-box models into interpretable functional ANOVA models, allowing for higher-order interactions while maintaining computational efficiency. Additionally, innovative frameworks such as Heckman-FA facilitate appropriate prediction feature selection, addressing sample selection bias challenges. By leveraging the strengths of both machine learning and traditional statistical approaches, these hybrid models can provide robust, interpretable, and accurate predictions across various applications [11, 43, 20, 45, 72]. Such models benefit from the interpretability of statistical methods while harnessing machine learning's predictive power, ultimately improving risk assessment and clinical decision support.

Exploring dynamic and adaptive modeling techniques is crucial for developing prediction models responsive to changing patient data and healthcare environments. By integrating real-time data inputs and feedback mechanisms, these advanced predictive models can deliver timely and personalized forecasts tailored to individual patient needs, significantly enhancing their effectiveness in clinical practice. This approach utilizes bipartite graph representations of patient-clinician relationships, stratified sampling for rare outcomes, and sophisticated patient representation methods derived from NLP, contributing to more accurate and relevant predictions in healthcare settings [2, 8, 40, 73, 42].

Recent methodological advancements in prediction modeling, particularly through multinomial prediction models (MPMs) and innovative patient representation techniques derived from NLP, hold substantial promise for enhancing clinical predictions' accuracy and interpretability in nursing. These advancements enable healthcare providers to derive actionable insights from complex data, ultimately leading to improved patient care and outcomes. MPMs effectively address clinical scenarios with multiple outcome categories, while advanced patient representation methods significantly boost model performance, especially in cases with limited patient records. By adopting these methodologies, nursing professionals can better tailor their interventions and decision-making processes to meet their patient populations' diverse needs [2, 44].

9 Conclusion

This survey highlights the pivotal influence of prediction models in nursing, demonstrating their transformative potential in risk assessment and clinical decision-making processes. By enabling precise risk predictions and informed clinical choices, these models significantly elevate the quality of patient care. The adoption of advanced machine learning techniques, such as deep learning and ensemble learning, alongside traditional statistical methods, has led to the creation of hybrid models that leverage the strengths of both paradigms. This integration results in models that are not only more accurate but also more interpretable, thereby enhancing their practical utility in clinical settings.

The survey emphasizes the critical need for data integration and quality enhancement, underscoring the importance of comprehensive and robust datasets in improving model accuracy and reliability. The incorporation of diverse data sources—including electronic health records, genomic data, and imaging modalities—is essential for obtaining a holistic understanding of patient health, which in turn facilitates more precise risk assessments and tailored interventions.

Methodological advancements, particularly through dynamic and adaptive modeling techniques, are crucial for maintaining the relevance and effectiveness of prediction models in the rapidly evolving landscape of healthcare. These advancements enable models to process real-time data inputs and adjust to changing patient conditions, thereby enhancing their clinical applicability and impact.

In conclusion, the survey underscores the indispensable role of prediction models in nursing, highlighting their capacity to enhance patient outcomes through improved risk assessment and clinical decision support. By fostering a data-driven approach to patient care, these models contribute to more effective and efficient healthcare delivery, ultimately leading to better patient outcomes and increased healthcare efficiency.

AI-generated, for reference only.

References

[1] Avishek Choudhury and Christopher M Greene. Evaluating patient readmission risk: A predictive analytics approach, 2019.

[2] Ethan Steinberg, Ken Jung, Jason A. Fries, Conor K. Corbin, Stephen R. Pfohl, and Nigam H. Shah. Language models are an effective patient representation learning technique for electronic health record data, 2020.

[3] Ruining Deng, Nazim Shaikh, Gareth Shannon, and Yao Nie. Cross-modality attention-based multimodal fusion for non-small cell lung cancer (nsclc) patient survival prediction, 2024.

[4] R. Li, J. Liu, X. L. Deng, X. Liu, J. C. Guo, W. Y. Wu, and L. Yang. Pr-net: Leveraging pathway refined network structures for prostate cancer patient condition prediction, 2024.

[5] Prasannavenkatesan Theerthagiri and Vidya J. Cardiovascular disease prediction using recursive feature elimination and gradient boosting classification techniques, 2021.

[6] Eryu Xia, Yiqin Yu, Enliang Xu, Jing Mei, and Wen Sun. From risk prediction models to risk assessment service: A formulation of development paradigm, 2019.

[7] Konstantina Chalkou, Andrew J. Vickers, Fabio Pellegrini, Andrea Manca, and Georgia Salanti. Decision curve analysis for personalized treatment choice between multiple options, 2023.

[8] W. Katherine Tan and Patrick J. Heagerty. Surrogate-guided sampling designs for classification of rare outcomes from electronic medical records data, 2020.

[9] Saeede Ajorlou, Issac Shams, and Kai Yang. An analytics approach to designing patient centered medical homes, 2014.

[10] Pedro A. Moreno-Sanchez. Development and evaluation of an explainable prediction model for chronic kidney disease patients based on ensemble trees, 2022.

[11] D. Dhinakaran, S. Edwin Raja, M. Thiyagarajan, J. Jeno Jasmine, and P. Raghavan. Optimizing disease prediction with artificial intelligence driven feature selection and attention networks, 2024.

[12] Rebecca Whittle, Joie Ensor, Lucinda Archer, Gary S. Collins, Paula Dhiman, Alastair Denniston, Joseph Alderman, Amardeep Legha, Maarten van Smeden, Karel G. Moons, Jean-Baptiste Cazier, Richard D. Riley, and Kym I. E. Snell. Extended sample size calculations for evaluation of prediction models using a threshold for classification, 2024.

[13] Jeroen Ooge and Katrien Verbert. Trust in prediction models: a mixed-methods pilot study on the impact of domain expertise, 2021.

[14] Eben Holderness, Nicholas Miller, Philip Cawkwell, Kirsten Bolton, James Pustejovsky, Marie Meteer, and Mei-Hua Hall. Analysis of risk factor domains in psychosis patient health records, 2018.

[15] Ben Van Calster, David J McLernon, Maarten Van Smeden, Laure Wynants, Ewout W Steyerberg, Topic Group 'Evaluating diagnostic tests, and prediction models' of the STRATOS initiative Bossuyt Patrick Collins Gary S. Macaskill Petra McLernon David J. Moons Karel GM Steyerberg Ewout W. Van Calster Ben van Smeden Maarten Vickers Andrew J. Calibration: the achilles heel of predictive analytics. *BMC medicine*, 17(1):230, 2019.

[16] Jing Mei, Eryu Xia, Xiang Li, and Guotong Xie. Developing knowledge-enhanced chronic disease risk prediction models from regional ehr repositories, 2017.

[17] Louis Chislett, Louis JM Aslett, Alisha R Davies, Catalina A Vallejos, and James Liley. Ethical considerations of use of hold-out sets in clinical prediction model management, 2024.

[18] Anirudha Rayasam and Nagamma Patil. Robust self-healing prediction model for high dimensional data, 2022.

[19] Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M. Glass, Jeffrey Spaeder, Justin Romberg, Jimeng Sun, and Cao Xiao. Stan: Spatio-temporal attention network for pandemic prediction using real world evidence, 2020.

[20] Yongchan Choi, Seokhun Park, Chanmoo Park, Dongha Kim, and Yongdai Kim. Meta-anova: Screening interactions for interpretable machine learning, 2024.

[21] Florian D. van Leeuwen, Ewout W. Steyerberg, David van Klaveren, Ben Wessler, David M. Kent, and Erik W. van Zwet. Empirical evidence that there is no such thing as a validated prediction model, 2024.

[22] Parker Seegmiller, Joseph Gatto, Abdullah Mamun, Hassan Ghasemzadeh, Diane Cook, John Stankovic, and Sarah Masud Preum. Actsafe: Predicting violations of medical temporal constraints for medication adherence, 2023.

[23] Yuxuan, Yang, Hadi Akbarzadeh Khorshidi, Uwe Aickelin, Aditi Nevgi, and Elif Ekinci. On the importance of diversity in re-sampling for imbalanced data and rare events in mortality risk models, 2020.

[24] Tian Gu, Jeremy M. G. Taylor, and Bhramar Mukherjee. A synthetic data integration framework to leverage external summary-level information from heterogeneous populations, 2022.

[25] Xue Teng, Fuad Gwadry, Haley McConkey, Scott Ernst, and Femida Gwadry-Sridhar. An adaptive treatment recommendation and outcome prediction model for metastatic melanoma, 2018.

[26] Issac Shams, Saeede Ajorlou, and Kai Yang. A multivariate hierarchical bayesian framework for healthcare predictions with application to medical home study in the department of veteran affairs, 2014.

[27] Wasifur Rahman, Masum Hasan, Md Saiful Islam, Titilayo Olubajo, Jeet Thaker, Abdelrahman Abdelkader, Phillip Yang, Tetsuo Ashizawa, and Ehsan Hoque. Auto-gait: Automatic ataxia risk assessment with computer vision on gait task videos, 2022.

[28] Maren Hackenberg, Marlon Grodd, Clemens Kreutz, Martina Fischer, Janina Esins, Linus Grabenhenrich, Christian Karagiannidis, and Harald Binder. Using differentiable programming for flexible statistical modeling, 2021.

[29] Nathaniel Braman, Mohammed El Adoui, Manasa Vulchi, Paulette Turk, Maryam Etesami, Pingfu Fu, Kaustav Bera, Stylianos Drisis, Vinay Varadan, Donna Plecha, Mohammed Benjelloun, Jame Abraham, and Anant Madabhushi. Deep learning-based prediction of response to her2-targeted neoadjuvant chemotherapy from pre-treatment dynamic breast mri: A multi-institutional validation study, 2020.

[30] Tianling Liu, Ran Su, Changming Sun, Xiuting Li, and Leyi Wei. Eocsa: Predicting prognosis of epithelial ovarian cancer with whole slide histopathological images, 2022.

[31] Moeez M. Subhani and Ashiq Anjum. Multiclass disease predictions based on integrated clinical and genomics datasets, 2020.

[32] Shashank Reddy Vadyala, Sai Nethra Betgeri, Eric A. Sherer, and Amod Amritphale. Prediction of the number of covid-19 confirmed cases based on k-means-lstm, 2020.

[33] Sabyasachi Bandyopadhyay, Ahna Cecil, Jessica Sena, Andrea Davidson, Ziyuan Guan, Subhash Nerella, Jiaqing Zhang, Kia Khezeli, Brooke Armfield, Azra Bihorac, and Parisa Rashidi. Predicting risk of delirium from ambient noise and light information in the icu, 2023.

[34] Evangelia Christodoulou, Jie Ma, Gary S Collins, Ewout W Steyerberg, Jan Y Verbakel, and Ben Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of clinical epidemiology*, 110:12–22, 2019.

[35] Trapti Shrivastava, Harshal Chaudhari, and Vrijendra Singh. Leveraging machine learning for early autism detection via indt-asd indian database, 2024.

[36] Jean Feng, Alexej Gossmann, Romain Pirracchio, Nicholas Petrick, Gene Pennello, and Berkman Sahiner. Is this model reliable for everyone? testing for strong calibration, 2023.

[37] Hua Cheng. Spinopelvic anatomic parameters prediction model of nslbp based on data mining, 2020.

[38] Yuelyu Ji, Yuhe Gao, Runxue Bao, Qi Li, Disheng Liu, Yiming Sun, and Ye Ye. Prediction of covid-19 patients' emergency room revisit using multi-source transfer learning, 2023.

[39] Jon A. Steingrimsson, Constantine Gatsonis, and Issa J. Dahabreh. Transporting a prediction model for use in a new target population, 2021.

[40] Mohammad Ziaul Islam Chowdhury and Tanvir C Turin. Variable selection strategies and its importance in clinical prediction modelling. *Family medicine and community health*, 8(1):e000262, 2020.

[41] Yukti Makhija, Samarth Bhatia, Shalendra Singh, Sneha Kumar Jayaswal, Prabhat Singh Malik, Pallavi Gupta, Shreyas N. Samaga, Shreya Johri, Sri Krishna Venigalla, Rabi Narayan Hota, Surinder Singh Bhatia, and Ishaan Gupta. Challenges in the application of a mortality prediction model for covid-19 patients on an indian cohort, 2021.

[42] Nan van Geloven, Sonja Swanson, Chava Ramspek, Kim Luijken, Merel van Diepen, Tim Morris, Rolf Groenwold, Hans van Houwelingen, Hein Putter, and Saskia le Cessie. Prediction meets causal inference: the role of treatment in clinical prediction models, 2020.

[43] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177):1–81, 2019.

[44] Celina K Gehringer, Glen P Martin, Ben Van Calster, Kimme L Hyrich, Suzanne M M Verstappen, and Jamie C Sergeant. How to develop, externally validate, and update multinomial prediction models, 2023.

[45] Yann McLatchie, Sölvi Rögnvaldsson, Frank Weber, and Aki Vehtari. Advances in projection predictive inference, 2024.

[46] Li Shi, Redoan Rahman, Esther Melamed, Jacek Gwizdka, Justin F. Rousseau, and Ying Ding. Using explainable ai to cross-validate socio-economic disparities among covid-19 patient mortality, 2023.

[47] Muhao Xu, Zhenfeng Zhu, Youru Li, Shuai Zheng, Yawei Zhao, Kunlun He, and Yao Zhao. Flexcare: Leveraging cross-task synergy for flexible multimodal healthcare prediction, 2024.

[48] Zhibin Zhao, Darcy Murphy, Hugh Gifford, Stefan Williams, Annie Darlington, Samuel D. Relton, Hui Fang, and David C. Wong. Analysis of an adaptive lead weighted resnet for multiclass classification of 12-lead ecgs, 2021.

[49] Guansong Pang, Ngoc Thien Anh Pham, Emma Baker, Rebecca Bentley, and Anton van den Hengel. Deep depression prediction on longitudinal data via joint anomaly ranking and classification, 2022.

[50] Rohit Kannan, Güzin Bayraksan, and James R. Luedtke. Data-driven sample average approximation with covariate information, 2022.

[51] Yutong Chen, Jiandong Gao, and Ji Wu. Dynamic feature selection in medical predictive monitoring by reinforcement learning, 2024.

[52] Zhuo Zhi, Moe Elbadawi, Adam Daneshmend, Mine Orlu, Abdul Basit, Andreas Demosthenous, and Miguel Rodrigues. Hgbnet: predicting hemoglobin level/anemia degree from ehr data, 2024.

[53] Kizito Nkurikiyeyezu, Anna Yokokubo, and Guillaume Lopez. The effect of person-specific biometrics in improving generic stress predictive models, 2019.

[54] Mohsen Sadatsafavi and John Petkau. Non-parametric inference on calibration of predicted risks, 2024.

[55] Mohsen Sadatsafavi, Hamid Tavakoli, and Abdollah Safari. Minding non-collapsibility of odds ratios when recalibrating risk prediction models, 2021.

[56] Moritz Berger and Matthias Schmid. Assessing the calibration of subdistribution hazard models in discrete time, 2020.

[57] Kim Luijken, Paweł Morzywołek, Wouter van Amsterdam, Giovanni Cinà, Jeroen Hoogland, Ruth Keogh, Jesse Krijthe, Sara Magliacane, Thijs van Ommen, Niels Peek, Hein Putter, Maarten van Smeden, Matthew Sperrin, Junfeng Wang, Daniala Weir, Vanessa Didelez, and Nan van Geloven. Risk-based decision making: estimands for sequential prediction under interventions, 2023.

[58] Stephanie F. Chan, Jue Hou, Xuan Wang, and Tianxi Cai. Risk prediction with imperfect survival outcome information from electronic health records, 2021.

[59] Jean Feng, Alexej Gossmann, Berkman Sahiner, and Romain Pirracchio. Bayesian logistic regression for online recalibration and revision of risk prediction models with performance guarantees, 2021.

[60] Shuai Zheng, Zhenfeng Zhu, Zhizhe Liu, Zhenyu Guo, Yang Liu, and Yao Zhao. Multi-modal graph learning for disease prediction, 2021.

[61] Konstantina Chalkou, Ewout Steyerberg, Matthias Egger, Andrea Manca, Fabio Pellegrini, and Georgia Salanti. A two-stage prediction model for heterogeneous effects of many treatment options: application to drugs for multiple sclerosis, 2022.

[62] Yu Huang, Jingchuan Guo, Zhaoyi Chen, Jie Xu, William T Donahoo, Olveen Carasquillo, Hrushyang Adloori, Jiang Bian, and Elizabeth A Shenkman. The impact of electronic health records (ehr) data continuity on prediction model fairness and racial-ethnic disparities, 2023.

[63] Theodora S Brisimi, Ruidi Chen, Theofanie Mela, Alex Olshevsky, Ioannis Ch Paschalidis, and Wei Shi. Federated learning of predictive models from federated electronic health records. *International journal of medical informatics*, 112:59–67, 2018.

[64] Ofir Ben Shoham and Nadav Rappoport. Federated learning of medical concepts embedding using behrt, 2023.

[65] Miyeon Jung, You-Bin Lee, Sang-Man Jin, and Sung-Min Park. Prediction of daytime hypoglycemic events using continuous glucose monitoring data and classification technique, 2017.

[66] Zhusi Zhong, Jie Li, Zhuoqi Ma, Scott Collins, Harrison Bai, Paul Zhang, Terrance Healey, Xinbo Gao, Michael K. Atalay, and Zhicheng Jiao. Region-specific risk quantification for interpretable prognosis of covid-19, 2024.

[67] Alexandros Rekkas, David van Klaveren, Patrick B. Ryan, Ewout W. Steyerberg, David M. Kent, and Peter R. Rijnbeek. A standardized framework for risk-based assessment of treatment effect heterogeneity in observational healthcare databases, 2022.

[68] Quazi Adibur Rahman Adib, Sidratul Tanzila Tasmi, Md. Shahriar Islam Bhuiyan, Md. Mohsin Sarker Raihan, and Abdullah Bin Shams. Prediction model for mortality analysis of pregnant women affected with covid-19, 2021.

[69] Mohsen Sadatsafavi, Andrew J Vickers, Tae Yoon Lee, Paul Gustafson, and Laure Wynants. The expected value of sample information calculations for external validation of risk prediction models, 2024.

[70] Hisashi Noma, Tomohiro Shinozaki, Katsuhiro Iba, Satoshi Teramukai, and Toshi A. Furukawa. Confidence intervals of prediction accuracy measures for multivariable prediction models based on the bootstrap-based optimism correction methods, 2021.

[71] Mohsen Sadatsafavi, Paramita Saha-Chaudhuri, and John Petkau. Model-based roc (mroc) curve: examining the effect of case-mix and model calibration on the roc plot, 2021.

[72] Huy Mai and Xintao Wu. On prediction feature assignment in the heckman selection model, 2024.

[73] Fan Zhang, Tong Wu, Yunlong Wang, Yong Cai, Cao Xiao, Emily Zhao, Lucas Glass, and Jimeng Sun. Predicting treatment initiation from clinical time series data via graph-augmented time-sensitive model, 2019.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.