JOURNAL OF DATA SCIENCE | Vol.2026, Issue 3, No.6
eISSN:2805-5160

Mining Healthcare Pathways Through Sequential Pattern Analysis for Chronic Disease Progression Prediction

Achmad Chusnun Ni'am[1], Tutuk Indriyani[2], Elmar B. Noche[3]*

[1]Environmental Engineering Department, Faculty of Civil Engineering and Planning, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia
[2]Informatics Engineering Department, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia
[3]Department of Information Technology, Pangasinan State University–Lingayen Campus, Lingayen, Pangasinan, Philippines

***Email:** elmarbnoche_ms@psu.edu.ph[3]

Abstract

Chronic disease progression is often shaped by a series of temporally ordered clinical events, including diagnoses, medication changes, laboratory abnormalities, hospital encounters, and treatment adjustments. However, many conventional predictive models represent electronic health records as static feature sets, thereby limiting their ability to capture sequential relationships that may indicate worsening health conditions. This study proposes a sequential pattern mining framework for identifying frequent healthcare pathways associated with chronic disease progression and incorporating these pathways into predictive modelling. Longitudinal electronic health records are first transformed into patient-level event sequences. PrefixSpan is then applied to discover frequent temporal patterns, while sequential association rule mining is used to examine pathway-outcome relationships through support, confidence, and lift. The extracted pathway patterns are subsequently converted into predictive features and combined with conventional clinical variables to estimate disease progression risk. The proposed framework is designed to be evaluated using publicly available healthcare datasets, with performance assessed through accuracy, precision, recall, F1-score, and rule-quality measures. By integrating temporal pathway discovery with predictive analytics, this study addresses the limitations of static feature-based prediction and enhances the interpretability of chronic disease progression modelling. The framework contributes to healthcare data mining by producing clinically meaningful pathway knowledge that can support risk stratification, early intervention planning, and decision-support development.

Keywords

Sequential pattern mining; Healthcare analytics; Disease progression; Association rules; Electronic health records; Human disease

Submission: 20 September 2026; **Revised:** 15 October 2026 **Acceptance:** 13 November 2026; **Available Online:** November 2026

Copyright: © 2026. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

Introduction

Chronic diseases present a significant burden to modern health care services (Hacker, 2024). The progression of chronic diseases relies heavily on multiple consultations, medication regimen alterations, laboratory and diagnostic studies, and other procedural interventions. This is in contrast to the progression of acute diseases which often relies on a single clinical event. The early identification of these gradual pathways will allow improvement in the planning of care and the personalization of intervention strategies.

EHRs describe clinical pathways in great detail and allow researchers to study the progression of clinical states (Pellegrini et al., 2025). Unfortunately, the vast majority of predictive models rely on historical data which are reduced to static variables. The number of a patient's diagnoses, medications, or laboratory results may be of some assistance, but often they lose significant meaning. The same events may be triggered by different clinical scenarios if the order of the events is ignored.

A patient whose laboratory tests remain abnormal and who subsequently receives intensified medication may have a different disease progression than a patient whose laboratory tests are relatively normal and who receives the same medication change after a routine follow-up. The order of clinical events is especially important for the analysis of disease progression (D'Amico et al., 2022). Predictive models which fail to account for clinical event order will likely fail to identify significant clinical patterns.

Sequential pattern mining is equipped for this challenge (Jamshed et al., 2024). Given ordered data, it identifiesfrequent subsequences. The technique can identify recurrent pathways among patients having the same clinical patterns. In the context of healthcare analytics, some examples of recurrent sequences that may be discovered include a diagnosis followed by a medication adjustment, a lab deterioration that is followed by an admission, or an escalation in the outpatient treatment that is followed by a rise in treatment. These patterns of sequences can provide clinically actionable insights that can aid clinicians in making decisions.

Herein, a sequential pattern mining framework for chronic disease progression pathways is introduced (Assadi Shalmani et al., 2026). Using an event sequence mining approach, electronic health records are converted to patient-level event sequences. A PrefixSpan algorithm and sequential pattern mining are used. The derived patterns are then used as predictive features in machine learning and exposed to models for estimation of the risk of chronic disease progression. The authors hope that the simultaneous incorporation of the discovery of temporal patterns and the estimation of disease progression risk will result in an improvement of the predictive and interpretative dimensions.

This focus area builds on the use of machine learning and computational intelligence in biomedical engineering. There is an increased reliance on data-driven methods for prediction and support for diagnosis and clinical decision-making (Mohammad et al., 2026). The objective of the research presented in this paper has at its core the prediction of human disease progression using the temporal mining of healthcare pathways. Unlike other papers, this objective is not a side comment, but an analysis of how ordered healthcare events can convey progression risk in an interpretable manner.

This paper helps advance data science technologies in the healthcare field in three specific manners (Ahmed et al., 2023): (1) creating a new approach to temporal sequential pattern mining in the context of preserving temporal order in medical event sequences recorded in electronic health records (EHRs); (2) incorporating the behavioral healthcare pathway patterns into predictive modeling for estimating the risk of progressive chronic illness; and (3) enhancing the interpretability of chronic illness progression prediction by delineating pathway-based rules that are verifiable, comparable, and translatable into clinical decision-support rules.

The rest of the paper is organized as follows: Section 2 provides an overview of relevant studies and frameworks related to healthcare data mining, the prediction of chronic disease progression, sequential pattern mining, and clinical interpretability. Section 3 provides an overview of the proposed research framework. The design and execution of the research are addressed in Section 4, with Section 5 describing the approach to evaluation. Implications and anticipated contributions are discussed in Section 6. The final section provides a summary and future research.

Literature Review

Healthcare Data Mining and Electronic Health Records

The goal of healthcare data mining is to produce knowledge by analyzing data contained in medical and clinical records (Guo&Chen, 2023). The use of electronic health records (EHRs) has led many healthcare organizations to collect and store sizable amounts of patient data which can be analyzed to make predictions, aid in the decision-making process, and enhance the health care system. EHRs contain longitudinal data, which consist of information regarding patients' diagnoses, prescriptions, laboratory tests, and medical procedures as well as documentation of the patients' admissions and discharges along with follow-up visits.

EHRs can provide support for stratification of patients, development of early warning systems, design of individualized treatment strategies, and prediction of treatment outcomes in chronic disease management (Alam et al., 2024). Unfortunately, they are often characterized as complex, heterogeneous, temporally dependent and incomplete. Patient histories may consist of missing data, visit patterns of varying frequency, redundancy, and varying coding systems. Therefore, the complexity and temporality of clinical data structure must be addressed in order to facilitate data mining in this domain.

In many current predictive models, demographic and static patient-level information are employed (Morid et al., 2023). Static patient-level information include features such as comorbidity counts, medication history, laboratory summaries, and previous hospital visits. These features, while informative, may be deficient in their ability to capture disease progression and the related risk. Temporal order and the occurrence of clinical events may also be informative in assessing a patient’s risk. Thus, health-care data science is increasingly focusing on the analysis of temporal patterns.

Chronic Disease Progression Prediction

The goal of predicting the progression of chronic illness is to determine whether or not a patient is at risk of worsening health status (Nenova&Shang, 2022). This can be in the form of disease complications, hospitalizations, more intensive treatment, or a progression of the illness to a higher severity. The value of an accurate prediction is that high-risk patients can be identified prior to adverse outcomes. This supports the use of more intensive treatments and the most effective use of healthcare services.

The traditional methods of prediction tend to treat the progression of illness as a series of classifications that rely on a set of clinical variables (Eaton et al., 2023). This is a useful practice, but it can tell very little about the sequence of progression. An example of this might be a prediction model that shows a patient is at a high risk of something, but does not delineate if this high risk is due to a series of abnormal test results, a delay in treatment, a series of emergency room visits, or a combination of all of the above. In the practice of medicine, such a model would be of limited use.

The utilization of a pathway-based model is one of the solutions to this problem (Wieder et al., 2024). Pathway-based models tend to classify pathways of clinical practice, noting the ones that are most frequently seen prior to the progression of a chronic illness. This model treats records of patients as a set of variables that are not structured, but considers the development of chronic disease as a logical sequence. This model in particular is useful for chronic diseases that require long term outcomes to be assessed due to the cumulative effect of treatment, as well as the cumulative effect of symptoms and diagnoses.

Sequential Pattern Mining

Sequential pattern mining looks for specific sequences that repeat in data records sorted in chronological order (Mazumdar&Sarma, 2025). Examples of its application are in web mining, behavioral analysis, and in the analysis of biological sequences and health care pathways. In the field of health care, a patient can be represented as a sequence of clinical events ordered in time. Analysis focuses on the patterns of clinical events of different patients or subsets of patients.

One of the most well-known algorithms for sequential pattern mining is PrefixSpan (Bechini et al., 2023). A frequent prefix is used to project the sequence data for further pattern discovery. PrefixSpan does not rely on the generation of candidates. Instead, it reduces the search space for the dataset, which can be advantageous as a dataset grows in size and in the number of sequences in the dataset. This algorithm can definitely be used for health care data for patient histories.

Sequential association rule mining is sequential pattern mining with the addition of rule-based associations for clinical events that precede and succeed other clinical events (Floricel et al., 2024). Support, confidence, and lift can be used as metrics to quantify these rules. Support relates to the frequency of a given sequence, confidence quantifies the probability of the succeeding clinical event, and lift measures the strength of the rule-based associations. These are the most common metrics for analysis as they allow for evaluation of both the frequency and clinical significance of a rule.

Machine Learning in Biomedical Engineering

The surge of machine learning in healthcare analytics and biomedical engineering is not difficult to understand. There are many ML applications that support medical imaging, disease prediction, and diagnostics, target identification, and even assist in the making of clinical decisions and support. In this context of literature, we consider the work of Mohammad et al. (2026) to substantiate the position of increasing impacts of machine learning applications, trends, and directions on research in biomedical engineering. Therefore, the computational methods applicable to healthcare predictions described are justified, yet the scope of this research is limited to the temporal mining of electronic health record pathways.

A completely black-box ML system is one that lacks transparency and understanding(Şahin et al., 2025). Our system, on the other hand, is focused on the interpretable discovery of temporal pathways and is not considered a black-box system. In developing prediction models for healthcare, it is essential that not only the correct prediction is made, but that the working of the model is understandable by the healthcare practitioner and can be reviewed. Sequential pattern mining aids in the discovery of pathway patterns and development of association rules that describe the pathway and the clinical events that lead to the progression of a particular disease.

Clinical Interpretability and Decision-Support Value

Clinical interpretability is an essential factor in healthcare predictive analytics. Prediction outputs must be understandable and serve a purpose to the healthcare professional (Yang, 2022). Models that produce only a risk score are likely not useful to clinicians if they cannot appreciate the events or pathways that contributed to the assessed risk. In the management of chronic disease, the need for interpretability of the model is further amplified because of the need for continued assessment, and the need to modify and personalize the care of the patient.

Interpretability of the model is enhanced by the ability of sequential pathway mining to capture event sequences that can be viewed as clinical pathways (Assadi Shalmani et al., 2026). A discovered pathway may capture the progression that is associated with the repeated occurrence of abnormal test results, and the subsequent escalation of the patient's medication, and increasing numbers of emergency visits. Pathways that reflect the manner in which risk evolves are more meaningful to a clinician than the abstraction of a static.

Integration of prediction and explanation is the foundation for decision support in the proposed methodology (Kostopoulos et al., 2024). Frequent patterns and association rules capture the pathways and the event sequences, respectively, that are of greatest interest to clinicians. This allows clinicians to target efforts toward the most important patients, the most necessary visits, and to identify opportunities for earlier intervention in the care of the patient. The proposed framework optimally balances the needs of assessment and the application of the resultant analyses to the practice of healthcare.

Proposed Research Framework

This research introduces a novel framework for healthcare pathway mining that incorporates preprocessing of electronic health records, methods for developing event sequences, algorithms for mining sequential patterns and generating sequential association rules, and methods for pattern-based feature construction and for predictive modeling. The framework aims to enhance both the predictive output and the interpretability of clinical insights.

The framework operates on structured electronic health records. Relevant clinical data are extracted and categorized. The event sequences are built from each patient's clinical data organized in a chronological manner. PrefixSpan is used to discover the most common healthcare pathways, and sequential association rule mining is employed to uncover rule-based relations of earlier clinical events to later clinical outcomes of disease progression. The patterns are then transformed to features for machine learning frameworks.

Prior to a visual representation of the framework, the main elements of the framework are outlined in Table 1. The table helps understand the function of each step of the framework along with the expected outcome.

Table 1. Main Components of the Proposed Sequential Healthcare Pathway Mining Framework

Component	Description	Expected Output
Electronic health record extraction	Collect longitudinal patient records from publicly available healthcare datasets	Structured clinical records
Event standardization	Convert diagnoses, treatments, laboratory results, procedures, and admissions into event categories	Standardized healthcare events
Sequence construction	Arrange each patient's events according to time or encounter order	Patient-level event sequences
Sequential pattern mining	Apply PrefixSpan to discover frequent temporal healthcare pathways	Frequent sequential patterns
Sequential association rule mining	Generate pathway rules using support, confidence, and lift	Interpretable progression-related rules
Pattern-based feature engineering	Transform discovered patterns into model-ready variables	Pattern-enhanced feature matrix
Predictive modelling	Estimate chronic disease progression risk using machine learning models	Disease progression prediction
Evaluation and interpretation	Assess predictive performance and rule quality	Accuracy, F1-score, support, confidence, and lift

Table 1 indicates the benefits of the proposed framework beyond classification accuracy. This framework emphasizes the importance of interpretable pathway discovery, which is crucial for clinical decision support. Stage-wise construction of the proposed framework helps to convert typical healthcare records into interpretable temporal knowledge that aids in predicting the progression of chronic diseases.

The general structure of the proposed framework is shown in Figure 1. It shows the overall process of longitudinal healthcare data and how it is used for risk prediction after going through the various stages of processing, mining, and transformation.

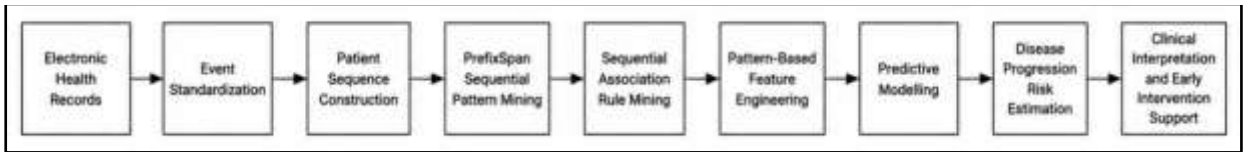


Figure 1
Proposed Framework for Sequential Healthcare Pathway Mining and Chronic Disease Progression Prediction

The entire research procedure is illustrated in Figure 1. Starting with electronic health records and culminating in elucidated estimations of disease progression risk, this framework is useful to both domains of data science and healthcare. At its core, this framework combines the identification of evolving patterns and the construction of predictive models.

In order to not burden the methodology section with too many equations, the core mathematics of this study is represented in a schematic diagramin Figure 2. To enhance the clarity of the technical procedures described, this diagram may be converted into JPG format and embedded as a formula framework.

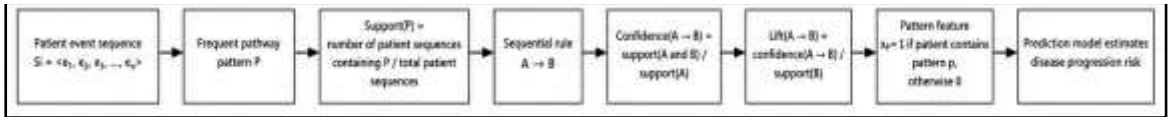


Figure 2
Formula Framework for Sequential Pattern Mining and Disease Progression Prediction

Figure 2 shows the basic structure of the mathematics needed for the study. Rather than writing out many separate equations, the figure puts together sequence representation, support, confidence, lift, pattern-based features, and prediction into one clear formula workflow.

Methodology

Research Design

Using a quantitative data mining research design, this study attempts to identify common health care pathways within longitudinal electronic health records and assess their potential for predicting the progression of chronic diseases. The research design incorporates both unsupervised pattern discovery and supervised predictive modeling. For this study, sequential pattern mining is applied to identify recurrent temporal pathways, and the progression risk is modeled with machine learning.

Dataset Selection

This study focuses on healthcare datasets that are publicly available and contain longitudinal patient records. Datasets may contain diagnosis codes, medication histories, laboratory results, records of procedures, admissions and discharges, and encounters and follow-ups. The datasets need to include some sort of timestamp to allow patient histories to be sorted in a sequential manner.

The dataset needs to satisfy some specific criteria. First, healthcare events should be multiple for every individual. Second, all events should be timestamped or structured in order of encounters. Third, the dataset should present some sort of outcome that indicates progression, such as complications, hospitalizations, escalation in the level of treatment, or worsening stage of a disease. Fourth, the dataset should include a sufficient number of records for every individual to allow sequential pattern mining and the development of predictive models.

Data Preprocessing

Data preprocessing makes electronic health records suitable for the type of analysis we wish to carry out. This involves eliminating duplicates, correcting inconsistent timestamps (if possible), filling missing values based on the variable's type and significance, and grouping clinical events to optimize classification and reduce data sparsity. This grouping can include therapeutic class grouping for medication records, grouping diagnosis codes by disease categories, and grouping lab results by normal/abnormal.

After preprocessing, each patient is represented as an ordered sequence of events. Rather than presenting a long mathematical derivation in the text, this study treats each patient record as a chronological event sequence. This representation allows the mining algorithm to identify whether particular event patterns occur before disease progression.

Disease Progression Outcome

Chronic disease progression is the dependent variable. Based on the dataset used, disease progression may be defined as a hospitalization, development of a complication, deterioration of clinical stage, laboratory deterioration, or treatment intensification and escalation. The outcome must be clinically relevant and uniformly applied across all patient records.
In predictive modelling, disease progression is described as a binary outcome. For this study, patients who undergo the defined progression event are classified as progression cases, while patients who do not undergo the event are classified as non-progression cases. The study aims to evaluate if the use of sequential pathway features improves the prediction of progression cases in comparison to conventional static features.

Sequential Pattern Mining Using PrefixSpan

PrefixSpan is designed to mine frequent healthcare pathways based on patient event sequences. A pattern is considered frequent if it appears in at least a given percentage of patient sequences. The minimum support threshold should be determined carefully as it will affect both the discovery of patterns and how interpretable they are.

When the threshold is too high, clinically important but infrequent patterns may be lost. When the threshold is too low, the pattern discovery may become so extensive that the results become uninterpretable. Because of this, the support threshold should consider the clinical importance and interpretability, as well as the feasibility of the computations.

Sequential Association Rule Mining

After discovering frequent patterns, the next step is generating sequential association rules. These rules evaluate the relationship of consequent outcomes to the preceding pathways and their progression. A sequential rule relates a prior sequence of events to a subsequent event or disease progression outcome. Evaluating each rule focuses on support, confidence, and lift.

Support indicates rules that pertain to pathways common among subjects, confidence indicates the rules enforce strong and consistent pathway-outcome relationships, and lift shows the rules have relationships that are greater than random co-occurrence, meaning they’re truly significant. These indices are illustrated in Figure 2 so they aren't repeated in the text as multiple individual equations.

Pattern-Based Feature Engineering

Identified patterns are of a sequential nature and can be transformed into predictive features. For example, a binary variable may be defined for each patient. The variable would indicate if a specific pattern was present in that patient’s sequence. Alternatively, features based on frequency may be defined to show how often a specific pattern was recorded for a patient.

In conjunction with age group, gender, number of diagnoses and medications, past admissions, and the number of lab abnormalities, the aforementioned variables create a hybrid feature set. This also includes static clinical information combined with temporal pathway data.

Predictive Modelling

This research analyzes standard feature-based models and models that build upon sequential patterns. Baseline models account only for static patient-level variables. Proposed models account for sequential pathway features. Suitable algorithms consist of logistic regression, decision trees, random forests, gradient boosting, and support vector machines. The choice of the final model is based on the predictive power and interpretability.
Before assessing the models, Table 2 lays out the expected design for model comparison. This model comparison is critical, as it assesses if the temporal pathway features add extra predictive power.

Table 2
Planned Model Comparison for Chronic Disease Progression Prediction

Model	Input Features	Purpose
Baseline model	Demographic and conventional clinical variables	Measures standard prediction performance
Sequential pattern model	Frequent healthcare pathway indicators	Tests the predictive value of temporal pathway patterns
Hybrid model	Conventional variables and sequential pattern features	Evaluates combined predictive improvement
Rule-based interpretation model	High-support, high-confidence, and high-lift rules	Provides clinically interpretable pathway knowledge

As presented in Table 2, the experiment will assess not only if the proposed model has superior predictive performance, but also if it provides better explanations of the disease progression pathways. Due to the integration of conventional clinical predictors alongside temporal pathway information, the hybrid model is anticipated to yield the best tradeoff among predictive performance and interpretability.

Validation Strategy

The data is split into training and test sets. To avoid data leakage, only patterns found in the training data are kept. These patterns are then used as prediction features for both the training and test data. Model parameters and robustness may be evaluated and refined through the use of cross-validation on the training data. The generalization performance of the model is assessed by testing it on data that has not been seen.

Evaluation Metrics

The evaluation consists of two components: predictive model evaluation and sequential rule evaluation. Predictive model evaluation assesses how accurately the models forecast the risk of disease progression. Sequential rule evaluation assesses the quality and interpretability of the identified healthcare pathways.

Predictive Performance Metrics

Preditive model assessment includes a number of metrics such as accuracy, precision, recall, and F1-score. Accuracy shows the total number of correct predictions. Precision shows the total number of predicted positive progression cases that are actually positive progression cases.

Recall shows the total number of actual positive progression cases that were predicted. F1-score represents the balance between precision and recall. The F1-score is crucial for the assessment of a model when the data is class imbalance.

For the sake of simplicity, classification metrics using mathematical expression are not included in the bulk of the text. This study provides these metrics in the result tables and explains their meaning in a clinical context. Since predicting cases of progression is one of the most important functions of a healthcare prediction model, the study focused on recall.

Sequential Rule Metrics

The quality of sequential rules is assessed using support, confidence, and lift. Support measures how many times a sequence occurs in a dataset. Confidence measures the likelihood of the consequent in the case of the antecedent sequence. Lift measures if the association of a rule is stronger than the association by chance.

A rule with strong support means that the rule is common, a rule with strong confidence means that the rule is reliable, and a rule with lift greater than one indicates a significant association. In healthcare, the most significant rules do not have to be the most frequent rules. Some rules that are infrequent may strongly suggest the progression of a disease and be clinically significant.

This study provides a set of criteria against which result interpretation can be expected. A summary of the evaluation metrics used is shown in Table 3.

Table 3
Evaluation Metrics for Predictive Modelling and Sequential Rule Analysis

Metric	Used For	Interpretation
Accuracy	Predictive model evaluation	Overall proportion of correct predictions
Precision	Predictive model evaluation	Reliability of predicted progression cases
Recall	Predictive model evaluation	Ability to identify actual progression cases
F1-score	Predictive model evaluation	Balance between precision and recall
Support	Sequential rule evaluation	Frequency of a pathway in patient sequences
Confidence	Sequential rule evaluation	Probability of consequent event after antecedent sequence
Lift	Sequential rule evaluation	Strength of association beyond random occurrence

Table 3 exhibits that the evaluation strategy encompasses both the predictive and interpretive aspects of the study. This is relevant as the proposed framework is expected to enhance prediction while simultaneously developing understandable healthcare pathway knowledge.

Expected Results and Discussion

Discovery of Frequent Healthcare Pathways

The sequential pattern mining stage will yield frequent healthcare pathways related to the progression of chronic diseases, which may demonstrate sequences of abnormal laboratory findings, progressively intensified medication, emergency visits, numerous specialist referrals, or hospitalizations. These pathways illustrate the sequences of clinical emergencies that patients experience and may explain how patients usually develop more advanced negative outcomes from early clinical warning signs.

The interpretability of the pathways is their most important feature. Instead of only stating that a patient is high risk, the framework can suggest what may have led to that risk.

This is beneficial to clinicians because the risk pathways suggest a mechanism by which they may be able to intervene in a clinical situation.

Interpretation of Sequential Association Rules

Sequential association rules focus on the discovery of event patterns associated with specific progression outcomes. For example, one rule could state that patients who have abnormal lab results that are followed by an escalation in treatment are at a higher risk of being hospitalized. Another rule may state that patients who have constant visits to the emergency room combined with a progression of treatment have a higher risk of progression.

These rules can be ranked based on support, confidence, and lift. Rules with high lift are the most important because they show stronger than random association. Clinical interpretation is needed, however, because not every statistically rigorous association is clinically important, and so, they should be regarded more as knowledge for decision support than clinical conclusions.

Predictive Improvement from Sequential Features

It is anticipated that the hybrid model will do better than the baseline model due to the addition of temporal pathway features in addition to standard clinical variables. If the hybrid model registers an improved F1-score or recall, this could be a case where the power of the temporal pathways is leveraged to assist in the prediction of progression cases. In the context of healthcare, the recall metric is critical, as neglecting to predict that a patient will become high risk will result in an avoidable delay in patient care.

Improving a prediction will always be a priority, but the implications of that improved prediction must be carefully assessed. If a model becomes more accurate, but far more complex to interpret, that model is likely not ideal for a healthcare context. This study focuses on both the performance and the explanation. The model proposed in this study is a good model because it permits the investigation of healthcare pathway patterns and allows them to be linked more directly to the prediction outcome.

Contribution to Healthcare Data Science

This research develops data science in healthcare by combining predictive modeling with sequential pattern mining to circumvent the constraints of static feature-based models by capturing the sequence of clinical events. It generates pathway rules which can be evaluated by researchers and clinicians, contributing to transparent healthcare analytics.

Data mining in chronic disease management is also advanced by this research. The framework provides a more accurate depiction of patient care pathways by converting electronic health records to a sequence of temporal events. This model framework is conducive to the development of early warning systems and the design of patient stratification and tailored interventions.

Practical Implications for Clinical Decision Support

This proposed framework may enhance clinical decision support systems by offering risk estimations and interpretable healthcare pathways. In managing chronic diseases, clinicians need to know if a patient is at risk and, if so, the reason. This can be answered by pathway patterns as sequential pathway patterns can illustrate the reasoning behind risk by indicating clinical events that lead to risk of progression.

The framework supports the prioritization of patients. High risk pathway patterns in patients may lead to more careful monitoring, follow-ups and a review of the treatment prescribed. This enables healthcare teams to focus their resources more effectively in environments with a constrained ability to deliver clinical services due to the size of the patient population.

Yet another example of a practical utility of the framework is that pathway-based rules may provide an empirical basis to support phase retrospective quality improvement. Progression pathways can be explored to examine opportunities to close a gap in care, prompt an intervention, or address frequent predictable utilization. Hence, the framework is not purely a predictive tool. It offers opportunities to understand care and improve the management of chronic conditions.

Conclusion

This paper introduces a new methodology that applies sequential pattern mining to the estimation of chronic illness progression from electronic health records. Patient histories are first converted to event records that are temporally arranged. Then, frequent healthcare pathways are discovered, and sequential association rules are drawn. The identified healthcare pathways are used for model forecasting. By maintaining the temporal arrangement of health events, the proposed methodology overcomes one of the key drawbacks of standard feature-based chronic illness progression forecasting models.

This work combines the discovery of pathways, the mining of interpretable rules, and the estimation of chronic illness progression. This work is aimed at enhancing the clinical applicability of forecasting models by offering pathway-based explanations for the model's output. The methodology is anticipated to harmonize forecasting accuracy with the level of interpretive knowledge required by clinicians, health care investigators, and the developers of decision-support systems.

The methodology can be further expanded by substantiating it with other chronic illnesses, implementing comparative studies of sequential mining methods, incorporating time lags between events, and involving clinical experts to verify the proposed pathways. Advanced studies can also apply deep learning on sequences to interpret results through rule-based reasoning.

References

Ahmed, A., Xi, R., Hou, M., Shah, S. A.,&Hameed, S. (2023). Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. IEEE Access, 11, 112891–112928. https://doi.org/10.1109/ACCESS.2023.3323574

Alam, M. A., Sohel, A., Uddin, M. M.,&Siddiki, A. (2024). Big data and chronic disease management through patient monitoring and treatment with data analytics. Academic Journal on Artificial Intelligence, Machine Learning, Data Science and Management Information Systems, 1(1), 77–94. https://doi.org/10.69593/ajaimldsmis.v1i01.133

Assadi Shalmani, M., Khani, M., Taleban, A., Yi, Z., Fink, J. T., Weber, C. E., Lu, Q.,&Luo, J. (2026). Sequential pattern transformer (SPT): A generative and interpretable framework for predicting disease trajectories. Neural Computing and Applications, 38(3), Article 28. https://doi.org/10.1007/s00521-025-11695-4

Bechini, A., Bondielli, A., Dell'Oglio, P.,&Marcelloni, F. (2023). From basic approaches to novel challenges and applications in sequential pattern mining. Applied Computing and Intelligence, 3(1), 44–78. https://doi.org/10.3934/aci.2023004

D'Amico, G., Bernardi, M.,&Angeli, P. (2022). Towards a new definition of decompensated cirrhosis. Journal of Hepatology, 76(1), 202–207. https://doi.org/10.1016/j.jhep.2021.06.018

Eaton, N. R., Bringmann, L. F., Elmer, T., Fried, E. I., Forbes, M. K., Greene, A. L., Krueger, R. F., Kotov, R., McGorry, P. D., Mei, C.,&Waszczuk, M. A. (2023). A review of approaches and models in psychopathology conceptualization research. Nature Reviews Psychology, 2(10), 622–636. https://doi.org/10.1038/s44159-023-00218-4

Floricel, C., Wentzel, A., Mohamed, A., Fuller, C. D., Canahuate, G.,&Marai, G. E. (2024). Roses have thorns: Understanding the downside of oncological care delivery through visual analytics and sequential rule mining. IEEE Transactions on Visualization and Computer Graphics, 30(1), 1227–1237. https://doi.org/10.1109/TVCG.2023.3326939

Guo, C.,&Chen, J. (2023). Big data analytics in healthcare. In Y. Nakamori (Ed.), Knowledge technology and systems: Toward establishing knowledge systems science (Vol. 34, pp. 27–70). Springer. https://doi.org/10.1007/978-981-99-1075-5_2

Hacker, K. (2024). The burden of chronic disease. Mayo Clinic Proceedings: Innovations, Quality&Outcomes, 8(1), 112–119. https://doi.org/10.1016/j.mayocpiqo.2023.08.005

Jamshed, A., Mallick, B.,&Bharti, R. K. (2024). A systematic review on sequential pattern mining-types, algorithms and applications. Wireless Personal Communications, 138(4), 2371–2405. https://doi.org/10.1007/s11277-024-11605-2

Kostopoulos, G., Davrazos, G.,&Kotsiantis, S. (2024). Explainable artificial intelligence-based decision support systems: A recent review. Electronics, 13(14), Article 2842. https://doi.org/10.3390/electronics13142842

Mazumdar, N.,&Sarma, P. K. D. (2025). Sequential pattern mining algorithms and their applications: A technical review. International Journal of Data Science and Analytics, 20(3), 1683–1726. https://doi.org/10.1007/s41060-024-00659-x

Mohammad, A. A., Alshdaifat, N., Mohammad, S., Al-Daoud, K., Vasudevan, A., Ling, A. W. P., Devi, S., Mcxin, T.,&Mohammad, A. I. (2026). Machine learning in biomedical engineering: A bibliometric analysis of trends, applications, and future directions. Studies in Systems, Decision and Control. https://doi.org/10.1007/978-3-032-10016-0_8

Morid, M. A., Sheng, O. R. L.,&Dunbar, J. (2023). Time series prediction using deep learning methods in healthcare. ACM Transactions on Management Information Systems, 14(1), Article 2, 1–29. https://doi.org/10.1145/3531326

Nenova, Z.,&Shang, J. (2022). Chronic disease progression prediction: Leveraging case-based reasoning and big data analytics. Production and Operations Management, 31(1), 259–280. https://doi.org/10.1111/poms.13532

Pellegrini, C., Özsoy, E., Bani-Harouni, D., Keicher, M.,&Navab, N. (2025). From EHRs to patient pathways: Scalable modeling of longitudinal health trajectories with LLMs [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2506.04831

Şahin, E., Arslan, N. N.,&Özdemir, D. (2025). Unlocking the black box: An in-depth review on interpretability, explainability, and reliability in deep learning. Neural Computing and Applications, 37(2), 859–965. https://doi.org/10.1007/s00521-024-10437-2

Wieder, C., Cooke, J., Frainay, C., Poupin, N., Bowler, R., Jourdan, F., Kechris, K. J., Lai, R. P. J.,&Ebbels, T. (2024). PathIntegrate: Multivariate modelling approaches for pathway-based multi-omics data integration. PLOS Computational Biology, 20(3), Article e1011814. https://doi.org/10.1371/journal.pcbi.1011814

Yang, C. C. (2022). Explainable artificial intelligence for predictive modeling in healthcare. Journal of Healthcare Informatics Research, 6(2), 228–239. https://doi.org/10.1007/s41666-022-00114-1