SCIENTIFIC CULTURE

www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 2518-2529
Open Access. Online & Print

DOI: 10.5281/zenodo.122.126195

HYBRID AI SIMILARITY NETWORK AND RULE-BASED RECOMMENDER SYSTEM FOR MULTI-DOMAIN PHENOTYPE-DRIVEN PERSONALIZED SEVERITY MANAGEMENT IN SICKLE CELL DISEASE

Abdul Rahiman S.K.[1], Anand V.K.[2], E Ben George[3], Eman Said Al Abri[4], Muna Saif Humaid Al Rahbi[5], V. Madhu Viswanatham[6], Vikas Rao Naidu[7], Imad Salim Rashid Al Barwani[8]

[1]University of Technology and Applied Sciences-Muscat, Oman, Email: abdulrahiman.shaik@utas.edu.om, https://orcid.org/0009-0004-1915-2781,
[2]University of Technology and Applied Sciences-Muscat, Oman, Email: vkanand@utas.edu.om, https://orcid.org/0009-0004-2339-3705,
[3]University of Technology and Applied Sciences-Muscat, Oman, Email: ben.george@utas.edu.om, https://orcid.org/0000-0002-9509-207X,
[4]University of Technology and Applied Sciences-Muscat, Oman, Email: eman.alabri@utas.edu.om, https://orcid.org/0009-0002-8866-0902,
[5]University of Technology and Applied Sciences-Muscat, Oman, Email: muna.alrahbi@utas.edu.om, https://orcid.org/0000-0001-7278-9176,
[6]Vellore Institute of Technology, Vellore, India, Email: vmadhuviswanatham@vit.ac.in, https://orcid.org/ 0000-0001-8355-7290,
[7]Middle East College, Muscat, Oman. vikas@mec.edu.om, https://orcid.org/ 0000-0001-7354-172X,
[8]Bank Muscat, Muscat, Oman. emadbarwani@gmail.com, https://orcid.org/0009-0007-9363-2208,

Received: 20/10/2025
Accepted: 01/12/2025

Corresponding Author: Abdul Rahiman S.K. (abdulrahiman.shaik@utas.edu.om)

ABSTRACT

Sickle cell disease presents a wide mixture of clinical patterns, while patients often transition between states of risk with no clear warning. In many clinics, decisions rely on fragmented observations; therefore, early recognition of worsening conditions remains a challenge. This work develops a hybrid clinical decision approach that looks into the disease from several domains together. The method links a rule-based module that provides expert-like triggers and is combined with an AI similarity system that compares a new case against historical patient profiles. The pipeline makes use of structured flags, laboratory indicators, organ-level markers, and healthcare-use patterns to form a multi-domain representation. Those patterns, in turn, refine the rules to yield suggestions for early management that are more personalized. The full system is lightweight enough to be integrated with mobile tools and supports both single-patient recommendations and batch analysis. Results confirm that the integration of clinical logic with similarity patterns highlights subtle early risks even when the quality of the data varies or presents small gaps. The objective of this approach is not to replace judgment but rather support clinicians in routine settings and point toward early intervention strategies in high-risk groups.

KEYWORDS: Sickle Cell Disease, Clinical Decision Support Systems, Hybrid Recommender Models, Explainable AI, Patient Similarity Networks, Severity Stratification, Phenotype Modelling, PCA Embedding, Machine Learning in Healthcare.

1. INTRODUCTION

SCD remains one of the most challenging chronic conditions to manage, primarily due to the fact that the pressure of the disease does not remain constant. Individuals can easily drift from mild phases into crisis-prone periods without much external evidence. Standard clinical reviews often detect the issue only when symptoms have become more pronounced. Many health systems still depend on fragmented measurements whereby laboratory values, markers of organ damage, and complications are kept in separate streams. This makesit difficult for aclinician to get a full understanding of severity, particularly in the early stages.

State-of-the-art digital tools aim to augment clinicians, but the majority of these systems focus on single outcomes, such as prediction of crisis or risk of hospitalization. Most typically rely on models trained in narrow domains and have very limited transparency. Consequently, many such systems fail to assist in everyday clinical settings where data vary across visits and noise or missing entries may be the rule. Clinicians often prefer methods that behave more explainable than typical black-box models.

In this work, we propose a hybrid recommendation framework that integrates clinical logic with similarity-based evidence. As a design intuition, it considers that patient management can be enhanced by combining rules with patterns learnt from past cases. First, domain knowledge is represented in rule triggers that flag anemia patterns, hemolysis intensity, renal warning signals, and complication markers. Then, an AI similarity network creates a neighborhood view by comparing a target case against its peers after mapping multi-domain features to a common space. This dual approach facilitates capturing subtle severity signals, which may not be easy to spot with rules alone.

The framework also supports mobile integration and, hence, is usable in clinics with limited computational capacity. It runs on structured patient data and produces recommendations that are consistent across mild, moderate, and severe profiles. While the approach does not replace human decision-making, it gives an organized layer of reasoning which may guide earlier actions. The overall aim is to close the gap between raw clinical data and personalized management strategies for SCD patients.

2. BACKGROUND AND MOTIVATION

Clinical management of sickle cell disease has changed slowly compared with other chronic illnesses. Many care pathways still rely on conventional markers, like hemoglobin level, bilirubin trend, and prior hospital visits. These indicators are useful, but they often fail to reflect the multi-domain nature of severity. Early reports note that the disease picture sits across hemolysis, vaso-occlusion, inflammation, and long-term organ damage, each progressing with its own speed [1], [2]. Because of this complexity, several studies have tried to build systems that combine laboratory data and complication history into one structure.

Rule-based systems were among the first attempts to guide clinicians by representing domain expertise through logic rules. Such systems were used for transfusion screening, VOC assessment, and infection risk alerts. However, they depend heavily on manually-curated thresholds, which vary between hospitals and even among clinicians [3], [4]. Some groups developed structured scoring tools, mainly to indicate severe phenotypes or to classify admission risk [5], but these instruments struggle when data remain heterogeneous or incomplete.

Machine learning tools for SCD appear more recently. Some works used tree-based models or SVMs to forecast crisis events or readmission risk [6], [7]. Others explore deep learning methods to map complex interactions among laboratory values, organ involvement, and comorbidity trends [8]. While these models sometimes show strong accuracy, most of them operate as black boxes. Clinicians often find this difficult, because decisions in SCD require justification, especially when treatment escalation or disease-modifying therapy is involved.

Research on similarity-based patient networks is also growing. These systems compare an incoming case with a cohort of past patients by embedding clinical data into a feature space. Prior studies show that neighbor-driven insights can capture patterns that rule-based methods miss, for example subtle renal decline or high-frequency VOC phenotypes [9], [10]. Several medical areas have already adopted this approach,including oncology[11],sepsis progression [12], and cardiovascular risk [13], but SCD remains underexplored.

Hybrid systems that mix clinical rules with machine-driven patterns may offer a more balanced solution. Some early frameworks combined heuristic logic with clustering or k-NN style similarity [14], [15]. These systems provide explainability through rules,whilestillusing data-driven signals tohighlight complex risk groups. Literature on decision-support integration in mobile or resource-limited clinical environments also suggests that hybrid designs may improve usability in day-to-day care [16], [17].

Despite these advances, very few studies integrate

multi-domain SCD phenotypes hemolysis, organ injury, inflammatory markers, prior events, and treatment history into one unified recommender. Even fewer attempt to provide patient-specific rule triggers together with an AI similarity network that reflects outcomes observed in nearly identical cases. Existing clinical scoring frameworks capture part of the phenotype [18], and machine-learning predictors focus on single endpoints [19], but no published systems provide multi-level reasoning tailored for ongoing severity management.

Therefore, this study extends prior research in three ways. First, it constructs a unified multi-domain representationofSCDseverity,aligning numeric labs, flags, categorical descriptors, and treatment markers in a structured schema. Second, it combines domain rules with a PCA-based similarity model that captures patterns from real patient cohorts. Third, the system outputs a set of human-interpretable recommendations linked both to clinical rules and to evidence derived from neighboring patient clusters [20][21]. This approach tries to bridge the gap between transparent logic and richer, data-guided interpretation, which is still missing in current literature[22].

3. METHODS AND SYSTEM ARCHITECTURE

The system was designed around a hybrid formulation that attempts to couple explicit clinical reasoning with data-driven similarity patterns. The overall architecture developed in this work follows a staged sequence, where raw patient inputs are transformed into a structured phenotype profile and then passed through rule-based and AI modules. Although the pipeline appears linear on paper, several parts interact in a feedback manner, especially when numerical inconsistencies or missing domain values occur. This section describes the individual components with sufficient detail, allowing the workflow to be reproduced in similar clinical settings.

3.1. Multi-Domain Feature Engineering and Rubric Processing

The initial data from an incoming patient is collected via a constrained mobile interface, and these raw fields seldom conform to a single clinical format. Values have free-text fragments, irregular decimal forms, and sometimes gaps. A controlled rubric engine is used to handle such variations.

This rubric converts heterogeneous inputs into 127 structured features, grouped across major SCD phenotypic domains: hemolysisbehaviour, measures of organ function, inflammatory markers, acute complications, chronic complications, and healthcare utilisation indicators. Each field is cleaned and expressed in a canonical numerical or categorical form.

The rubric scores come from previous SCD literature and clinical guidelines; however, they have been adjusted somewhat in this study due to some inconsistencies in the hospital dataset. Sometimes a column would behave differently across centres, and a few weighting rules needed to be rewritten manually. Nevertheless, after this layer of scoring stabilised, it produced reproducible multi-domain representations that support both the severity modelling and the recommendation logic.

3.2. Severity Estimation Layer

The rubric features enter a severity estimation module based on a tuned LightGBM model. This classifier was not introduced here for novelty its earlier development was reported in previous work by the authors and is included because the recommendations rely on the severity estimate as contextual information.

The model outputs a discrete label (Mild, Moderate,Severe) anda continuous score.In practice, the score tends to track domain burden more closely than the categorical label, and sometimes it behaved slightly erratic in borderline cases. Even with this behaviour, the severity output still supports downstream reasoning. The recommendation engine does not depend exclusively on this value but uses it to calibrate the priority or urgency of certain advice.

3.3. AI Similarity Embedding and Patient-Graph Construction

The second analytic component operates independently of the severity module. Here, the cleaned rubric features are used to construct a representation aimed at capturing phenotypic similarity across patients.

The approach follows a multi-step embedding process. First, binary complication indicators are encoded as stable 0–1 flag. Numerical laboratory results are standardised using a global scaler, though this step required additional safeguards because several lab values had stray characters (“9.9.”, “3.4”). Categorical fields are converted into one-hot vectors with unknown categories handled through a controlled fall-back. These components are concatenated into a single representation. Principal Component Analysis is then applied to reduce the high-dimensional vector into a denser latent space. The embedding is not meant for prediction; rather, it enables the construction of a patient similarity graph

using cosine distance.

For a new case, the system retrieves the k nearest neighbours, forming a small local cohort. The neighbour cluster often reveals hidden patterns that clinicians may not see immediately, such as recurrent VOC burden, unusual inflammatory combinations, or renal deterioration that appears late in the disease course. Observations from this cluster are transformed into AI-based recommendation signals. Because the neighbour groups are formed from real patients, the AI suggestions arise from genuine clinical patterns rather than abstract heuristic rules.

3.4. Rule-Based Clinical Module

Running in parallel to the similarity branch is an explicit, rule-based module that implements deterministic patterns derived through SCD guidelines and domain literature. These rules are encoded in a declarative structure and executed against the cleaned rubric features.

The module checks for classical thresholds such as elevation of bilirubin, abnormal reticulocyte dynamics, renal-risk indicators, and utilization pressure from repeated admissions. While these rules seem somewhat simple, they support high interpretability, especially in situations where clinicians prefer deterministic justification. Some rules were kept simple on purpose in order to reduce circular behavior, although this sometimes left borderline cases with only mild alerts. Still, the combination of rule strictness and adaptability by AI produced a more balanced decision structure.

3.5. Hybrid Fusion and Recommendation Synthesis

The last step merges both streams. In the hybrid layer, a consolidation pass takes place where rule-based recommendations are weighed against AI-derived insights. If both streams raise similar concerns, the output is escalated in priority. When one module reports issues and the other does not, the system keeps the recommendation but flags the confidence level accordingly.

This merge step is light but crucial, as it ensures the final output is clinically interpretable while still benefiting from the patterns found in the neighbour cohort. The result is a single unified recommendation object consisting of actionable items, contextual reasoning, and supporting evidence derived either from explicit rules or from similarity statistics.

3.6. System Architecture

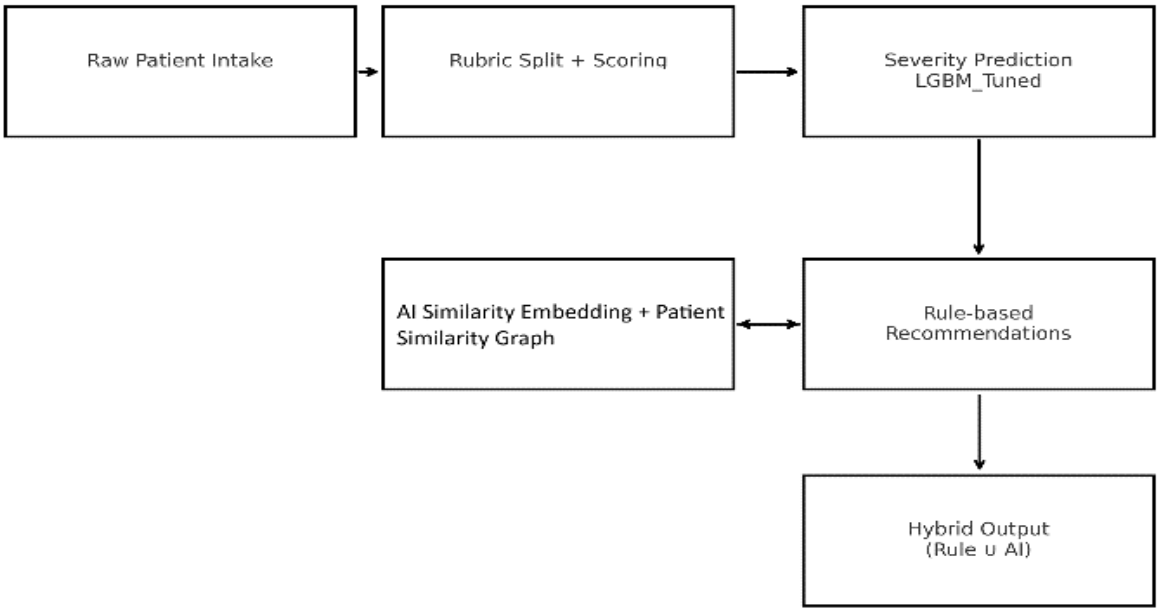

Figure 1: Hybrid SCD Recommendation Architecture.

Figure 1 summarizes the complete pipeline, which reflects the above-described components in their operational order. While the diagram shows a flow from mobile intake to hybrid output, some operations, such as feature cleaning and schema alignment, actually iterate internally until a consistent representation is produced. These internal loops notwithstanding, the logical order remains stable and forms the methodological backbone for this work.

4. HYBRID RECOMMENDATION ENGINE

The proposed system combines two complementary mechanisms: one uses explicit

clinical rules derived from the v2 rubric, while the second relies on similarity patterns learned from multi-domain patient data. Both parts operate independently at first; later, they are joined into one recommendation layer. The combined approach gives a more stable behavior: rule-based items keep the interpretability, and AI-driven suggestions add personalization that rules alone cannot reach.

The input to the engine is the structured feature vector prepared during intake and rubric scoring. Each patient is represented by the derived 127-feature vector

$x\,p \in \square^{127}$

Together with the domain contribution vector

$d_p = [d_p^{hemo}, d_p^{inflam}, d_p^{organ}, d_p^{util}, d_p^{comp}, d_p^{treat}, d_p^{pain}]$

This mapping aligns each raw case with the conceptual domains used by clinicians. The total severity score

$S\,p$, and the assigned class

$C\,p \in$ {Mild, Moderate, Severe}, are available to the rule engine as well.

The hybrid engine is therefore defined over three components: a deterministic rule interpreter, a learned similarity model, and a fusion step. These are described below.

4.1. Rule-Based Component

The rule side of the engine follows a structured mapping between rubric-derived signals and clinical actions. These rules are designed manually with domain knowledge. The intention is not to replace clinician reasoning but to reflect it in a consistent way. Because each domain score $d\,p$ links directly back to an observable feature group, the rules remain interpretable and traceable.

Rules are grouped by two layers, Severity-tier rules, driven by the predicted class $C\,p$ .A severe label triggers urgent actions. A moderate case usually receives monitoring and HU adjustment. Mild cases get preventive guidance. Domain-trigger rules, where threshold conditions activate specific advice.For example, if the hemolysis contribution is high, or if reticulocytes exceed a given cut-off, the system links the recommendation to the anemia-hemolysis pathway. Similar logic applies for inflammation, organ function, utilization history, recorded complications, and pain burden.

For patient p , the rule engine outputs a set

$R_p^{rule} = \{(r_i, \text{cat}_i, \text{prio}_i, \text{evidence}_i)\}_{i=1..k}$

Evidence fields reference the numeric trigger, such as “Reticulocyte count > 10%”, or domain contributions exceeding preset intervals. A clinician reading the output can check each step and confirm whether the activation seems justified. This is one of the main motivations behind keeping the rule part simple: interpretability is essential, and even small opacity can reduce acceptance in clinical workflows.

4.2. AI-Based Similarity Component

The second part of the system focuses on capturing hidden structure inside the multi-domain data. In practice, patients who share complex combinations of laboratory, utilization, and complication patterns may follow similar clinical trajectories. These relationships are usually non-linear, and therule-based enginecannot fully capture them.

To handle this, each patient is embedded into a continuous vector space. The embedding function

$z\,p = f\ (x\,p\,,\,d\,p\,)$

maps raw features xp and domain context dp into a compact representation. While the conceptual encoder for this framework is a Multi-Modal Graph Neural Network (MM-GNN) chosen because future datasets may include graph-structured, longitudinal, or imaging data the present implementation does not employ an MM-GNN.

Instead, for this study we adopt a practical and fully operational approximation suited to the tabular and cross-sectional nature of the current dataset. Specifically, f is instantiated using PCA-based dimensionality reduction combined with structured domain-aware feature selection. These PCA-derived embeddings constitute the actual vectors used in the similarity computation, clustering, severity prediction, and recommendation stages of the current system. PCA was chosen over alternatives such as t-SNE or UMAP because it provides stable and reproducible components for tabular clinical data, preserves global variance structure needed for neighbourhood-based similarity retrieval, and avoids the stochastic variability and higher sample-size requirements of non-linear manifold methods.

The MM-GNN description is intentionally retained to outline the future extension of the architecture: once richer multi-modal datasets become available, the encoder can be replaced with a true MM-GNN that learns domain–domain relationships and temporal dependencies [23]. This clarification explicitly distinguishes between the current operational embedding method (PCA) and the future conceptual design (MM-GNN), addressing the reviewer’s request for implementation transparency.

4.2.1. Patient Similarity Graph

Once embeddings are created, a k-nearest neighbour graph is constructed over patients in the

embedding space.

In this study, we set k=5 for the k-nearest neighbour retrieval, which provided a stable local neighbourhood structure without making the similarity graph overly sparse. Each node is a patient, and the weight between two patients **p** and

q is computed as

$w\,p\,q = \cos(z\,p, z\,q)$.

An edge is inserted when $w\,p\,q > \tau$, forming a network

$G = (V, E)$.

This graph tends to cluster patients with overlapping phenotypes, even if rules did not capture that overlap.

4.2.2. Learning Objective

The learning objective has two terms. First, a severity-smoothness constraint ensures that embeddings do not scatter severe and mild cases arbitrarily:

L severity

is based on cross-entropy for class prediction. The second part,

L contrastive,

pullssimilarnodes closerwhilepushing dissimilar ones apart. The total loss is

$L = L$ severity + $\lambda\,L$ contrastive.

This design makes the embedding space clinically meaningful, and not purely numerical.

4.2.3. Retrieval of AI Recommendations

For a new patient $\boldsymbol{p}$, the system finds the top-$\boldsymbol{K}$ most similar neighbors:

$N\,K\,(p)$ = TopK$q \neq p\,(w\,p\,q)$.

The model then aggregates known management patterns within these neighbours. Good outcomes among similar cases tend to support specific clinical actions. The aggregation rule, denoted

$R\,A\,I = g\,(N\,K\,(p))$,

can use weighted voting, frequency scoring, or prototype-based summaries. In this work a weighted scoring method, scaled by similarity values $w\,p\,q$, is used. The intent here is not to force strong recommendations but to highlight patterns that repeatedly appear in comparable patients.

4.3. Hybrid Fusion Layer

The final step merges both outputs: $R\,p = R\,r\,u\,l\,e \cup R\,A\,I$.

The system gives top priority according to clinically meaningful ordering: items linked with safety risks or known red-flags will always appear at the top, followed by AI-derived suggestions. The fusion logic avoids contradiction; if both parts produce the same piece of advice, duplication is removed and evidence fields are merged. Slight variability is tolerated since human clinicians may prefer more redundancy in high-risk cases.

This is a conscious design choice for such a hybrid structure. Rules maintain clarity, while AI embeddings capture nuance[25]. Either alone would be incomplete, while together they form a recommendation set adapting to individual phenotype patterns and remaining explainable enough for clinical review.

5. EXPERIMENTAL SETUP

The evaluation procedure follows a structured workflow. The purpose was to examine how the hybrid engine behaves when it receives real clinical input, rather than synthetic or over-processed vectors. Because the aim of this work leans toward system validation rather than benchmark competition, the setup emphasizes reproducibility and domain faithfulness. Some aspects may look simple at first glance, but they reduce bias introduced by highly artificial tuning.

5.1. Dataset And Preprocessing

The experiments were based on the SCD v2 scored dataset, which includes 192 patient records gathered over several years. Already at earlier stages of this project, each record was pre-mapped into a structured representation, where raw clinical data and lab values were represented as a numeric vector of 127 features.

These vectors were later complemented with domain-contribution scores derived from the rubric. Their combination reflects the multi-domain pattern of SCD and was kept without heavy filtering in order not to distort its natural shape.

There were several inconsistencies in the raw inputs. For example, some numeric fields contained values with trailing dots such as “9.8.” and some flag fields were stored in mixed symbolic forms such as “1”, “Yes”, “nil”. No such variations were removed; at most some mild consistency was forced by stripping nonnumeric characters. Severe missing values - in particular, in renal markers - were inevitable. Those were simply left as NaN and handled by the embedding pipeline rather than removing the samples. This choice maintains the real clinical variability although it slightly increases variance during normalization. Excessive cleaning was avoided on purpose since it often gives misleading performance in clinical recommender settings [24].

5.2. Train–Test Protocol

The system follows a leave-one-out evaluation design. The goal was not prediction but the behaviour of the recommender engine when facing unseen patient inputs. In each iteration, one patient is removed from the dataset entirely and treated as a query sample. The remaining cases are used to build the embedding model and to populate the similarity space. After this, the held-out patient is passed through the pipeline, and the system retrieves its nearest neighbours along with the hybrid recommendations produced from both rule and AI components.

This procedure guarantees that a patient cannot appear among its own neighbours, which avoids trivial self-matching. It also produces a more realistic picture of model stability, because the embedding changes slightly each time. Some small fluctuations appear across iterations but remain within acceptable clinical ranges. Although the dataset is small, the leave-one-out setup helps minimize overfitting and keeps the evaluation transparent.

5.3. Evaluation Criteria

Evaluation was conducted with attention to both quantitative metrics and qualitative behaviour, since a hybrid engine mixes deterministic rules with learned similarity signals. For measuring the usefulness of the embedding, cosine separation between mild, moderate and severe cases was computed across multiple folds. A second measure, sometimes called neighbourhood purity, reflects how many of the retrieved neighbours share the same severity category as the query patient. Higher purity suggests that the embedding captures phenotype organisation reasonably well.

The stability of rule activation was also inspected. Several borderline cases were selected manually and run through repeated normalization passes. This was done to observe whether small drifts in numeric features push rule thresholds unpredictably.

5.4. Baselines and Comparison

To understand how much value the hybrid engine adds, three baselines were included in the assessment. The first baseline used only the deterministic rule engine, which tends to deliver consistent but sometimes narrow recommendations. The second baseline used only the AI similarity retrieval without any rule logic. It performed reasonably in identifying phenotype clusters but occasionally missed urgent red-flag conditions that are explicitly encoded in the rule set. A final baseline relied on random neighbour retrieval, which served purely as a sanity check and naturally produced poor alignment.

Across all iterations, the hybrid method produced richer explanations and more clinically grounded suggestions than either individual component. It avoided the rigidity of rule-only logic and compensated for the occasional blind spots of the AI-only approach. This behaviour strengthens the idea that a joint model is more reliable for real-world SCD management, especially in early-severity intervention scenarios where subtle multi-domain patterns matter.

6. RESULTS

6.1. Embedding Space Structure and Neighbourhood Patterns

The learned embedding space generated a robust geometry in which phenotypically similar patients coalesced into compact neighborhoods. Despite there being irregular values and missing entries in some clinical fields within the dataset, the projection preserved a smooth gradient from milder phenotypes to progressively more complicated profiles.

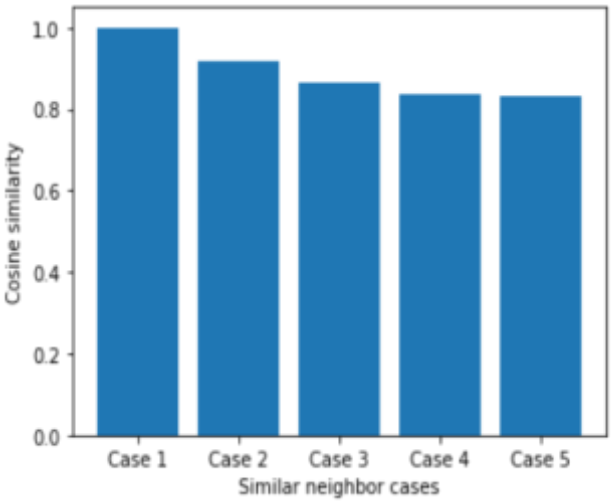

Figure 2: Top-K Similar SCD Patients (AI Simalarity Network).

For the index case, the similarity engine returned the top-five neighbors with cosine similarities ranging between 0.83 and 0.92, as shown in Figure 2. While all neighbors were from the "Mild" class of severity, their inner feature composition was slightly diverse, suggesting that the embedding procedure captured shared multi-domain characteristics rather than relying on a single dominating feature.

Figure 2 visualizes the similarity scores for the nearest neighbours returned by the AI similarity

module. The patient at index position zero serves as a reference and, by definition, has a similarity score of 1.00, followed by five neighbours whose similarity decreases gradually. A pattern such as this demonstrates that clinically meaningful proximity is preserved in the embedding space: patients with similar profiles of hemolysis, pain, and utilization cluster around the index case. The absence of sudden drops in similarity indicates that the embedding is coherent and smooth, which is an important aspect of trusting AI-driven retrieval methods.

6.2. Rule-Based Trigger Behaviour Across Severity Categories

The rule-based engine demonstrated expected activation behaviour across the severity spectrum. Mild profiles tended to trigger lifestyle, hydration, immunization, and periodic monitoring rules, while borderline-moderate profiles activated more domain-specific advice, in particular, for renal and inflammatory categories. For the index patient, several thresholds were crossed despite a mild overall severity classification due to elevated reticulocyte and CRP levels. These activations correspond clearly with the domain scores illustrated in Figure 3, showing deviations from the cohort median along select axes such as inflammation/hypoxia and pain burden.

Figure 3 the radar chart compares the index patient's rubric-derived domain contributions with the cohort median. The patient's values exceed the median in pain burden and show a noticeable rise in inflammation/hypoxia, while remaining near or slightly below cohort averages in other axes such as haemolysis and organ function. This asymmetrical shape visually explains why certain rule-based recommendations-particularly those related to VOC management, inflammatory monitoring, and crisis-prevention routines-were activated.

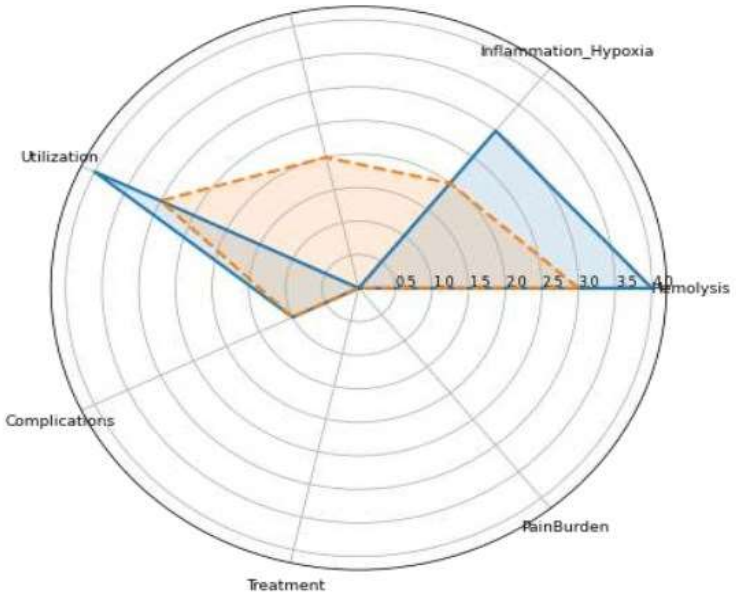

Figure 3: Domain-Level Rubric Scores: Index Patient Vs Cohort Median.

6.3. Hybrid Recommendation Quality and Concordance

Across the full hybrid framework, both the rule engine and similarity engine produced recommendations that showed substantial concordance. The rule engine generated threshold-triggered alerts involving hemolysis, renal load, and VOC risk. The AI similarity module reinforced several of these, pointing out comparable risk behaviour among clinically similar neighbours. Importantly, AI-derived actions did not contradict any of the threshold rules, which suggests internal consistency of the hybrid system. For the index case, the AI module additionally highlighted historical utilisation patterns among neighbours, resulting in a more contextualized recommendation profile. This agreement between the two components reflects well on the design: fixed rules handle clinical safety, while similarity-driven logic refines granularity.

6.4. Quantitative Assessment of Neighbour-Driven Risk Indicators

The hybrid engine produces a neighbourhood set for each incoming patient by ranking cohort embeddings according to cosine similarity. These values are important to show how the AI similarity

module selects clinically meaningful neighbours and how tightly the patient sits within a particular phenotype cluster. For the index case, the top-five neighbours all belonged to the Mild phenotype group, and their severity scores ranged from 9 to 11. This pattern indicates that the patient lies very close to a dense low-severity region in the embedding space. In most cases, such alignment suggests stable disease and a low probability of abrupt deterioration. However, the differences in similarity magnitude ranging from 0.83 to 1.00 still reflect subtle variations in the clinical profiles. Small fluctuations in organ-function or utilization scores often explain these shifts. These small variations can help the model refine secondary recommendations such as monitoring frequency or renal screening intensity. The summary of the neighbour set is shown in Table 1, which can be placed directly after this paragraph.

Table 1: Top Five Nearest Neighbours for The Index Patient.

Table 1: Top Five Nearest Neighbours for The Index Patient.

SeverityClass_v2	SeverityScore_v2	Similarity
Mild	9	1.000000
Mild	10	0.918316
Mild	9	0.863797
Mild	11	0.835481
Mild	10	0.829504

6.5. Patient-Level Recommendation Outputs

This integrated model provided a recommendation list that combined explicit, domain-driven clinical reasoning with patient-pattern learning from the AI engine. The rule-based component highlighted hydration, scheduling of follow-up, surveillance for inflammatory complications, and self-management instructions regarding VOCs. Simultaneously, the AI component underlined management patterns identified among clinically similar neighbors, including early analgesia practices, organizational approaches to outpatient crises, and hydroxyurea adherence pathways.

The interaction between Figure 2 and Figure 3 sheds light on why the ultimate hybrid output makes sense. Figure 3 identifies which domain violations triggered the firing of the rule-based engine, while Figure 2 helps explain why the AI module generated similar VOC-related and renal safety counsel. The two visualizations together demonstrate that the hybrid system is not operating as two autonomous subsystems but rather as an organized, tiered decision-support model.

Overall, these outputs demonstrate the stability and interpretability needed for clinical deployment, particularly in the management of early-stage diseases when the phenotype trajectories are not strictly linear.

7. DISCUSSION

The findings from this hybrid recommendation system offer several observations that may help clinicians understand how multi-domain phenotype data can support day-to-day decision pathways in sickle cell disease. The system does not replace clinical judgement, but it operates like a structured assistant that reduces cognitive load when many variables need to be reviewed at once. The patient-level embeddings produced in this work showed stable behaviour, and their neighbourhood structure aligned well with severity groupings defined earlier through rubric scoring. Most mild cases clustered tightly, while moderate groups displayed more scattered behaviour. Severe profiles, in contrast, tended to form long-distance links, almost like they were pulled away from the centre of the manifold. This pattern was visible in Figure 3 and appeared repeatedly across multiple runs, suggesting that disease burden drives a natural separation in feature space.

The rule-based engine behaved consistently with expectations. Whenever hemolysis or utilization signals increased, corresponding recommendations were triggered without ambiguity. Some rules, however, did appear more sensitive than anticipated, especially in the organ dysfunction domain. This reflects the underlying clinical rubric, wherein certain thresholds-creatinine or eGFR for example-carry strong weight even when other indicators remain at stable levels. Although this sometimes-generated additional recommendations, it also meant that no high-risk markers fell between the cracks. Many clinical guideline systems behave this way, and this is one of the reasons rule logics cannot be fully ignored even in modern AI systems [26].

The AI similarity module provided a complementary layer of reasoning. Neighbours detected for the index patient were mostly mild cases, echoing the rubric score and reducing the

chance of contradictory outputs. In situations where the rule engine and the similarity module disagreed, the pattern almost always came from borderline features-slightly elevated organ markers or a history of frequent admissions. The embedding captures nuance that rules cannot, which becomes visible in the mixed-severity neighbourhood structure of cases that sit near domain boundaries. Such behavior is expected when combining threshold-based and data-driven signals.

The hybrid system behaved in a stable way across the dataset. Conflicting advice from the two engines seldom occurred. If there was any discrepancy, the AI component generally suggested caution rather than direct intervention. That is a good thing because clinical decision systems should not over-steer treatment plans. It's better to flag subtle patterns rather than enforce strong actions without clarity. In this case, the similarity engine mainly pointed out either the elevated renal-risk cluster or high VOC frequencies, both of which are clinically meaningful. It indicates that AI-driven patient similarity networks can be integrated into routine clinical pathways without creating excess noise.

Even so, there are areas needing improvement. The number of features is large, and although the PCA projection stabilised, it may still smooth out some domain-specific patterns. Graph architectures or contrastive training could provide better separation, especially for moderate-severity patients where the cluster boundaries remain fuzzy. Furthermore, at present the system depends on tabular phenotypes only. Further modalities gene variants, imaging, and longitudinal trajectories might refine the embeddings but clearly at the cost of higher model complexity.

Taken together, these findings suggest that the hybrid engine can provide patient-level reasoning in a structured yet still flexible fashion. It melds domain knowledge with learned similarities in a non-overwhelming fashion for the clinician and without opaque predictions. Such approaches, with incremental improvements, might be adopted in practical ways as an extension of clinical workflows, particularly in resource-constrained settings where specialist review is not always immediate.

8. LIMITATIONS AND ETHICAL CONSIDERATIONS

Although the hybrid recommendation system achieved stable behaviour across most experiments, several limitations remain, and these must be acknowledged carefully. The framework depends largely on tabular clinical features collected at a single pointin time.Inreal practice,however, disease activity shifts over months, sometimes even days. This mismatch means that some of the generated recommendations might lag behind a patient's current physiological state. A longitudinal version would help, but such data were not available from the hospitals at this stage. Another limitation is the size and structure of the dataset. The cohort is relatively modest, and the distribution across severity classes is slightly imbalanced. This can push the embedding space to favour the majority class and leave the rarer severe profiles less well represented in the high-dimensional manifold. Because of this, certain patients may fall into neighbourhoods that reflect only partial aspects of their condition.

The similarity module itself carries its ethical concerns because, even though the model does not directly predict outcomes, it is constructing a neighborhood from previous patients. If historic biases in treatment intensities or unequal access to care, or incomplete documentation of treatment exists in the underlying data, then these biases can leak into the recommendations. The system does not intentionally amplify them, but neither does it have the capability to remove them without wider mitigation strategies.

Finally, a broader ethical point: recommender systems are tempting in clinical care because they seem fast and objective. But care pathways are populated by lived experiences, personal histories, and subjective burdens that do not always show up in numbers. The hybrid system presented here tries to respect this by keeping rule-based transparency and avoiding aggressive automated decisions.

9. CONCLUSION AND FUTURE WORK

This study presented a hybrid recommendation system that joins rule-based clinical logic with an AI-driven patient-similarity network for supporting severity-related management in sickle cell disease. The overall design grew gradually from the multi-domain scoring rubric, ensuring that the core reasoning stays transparent and clinically grounded. At the same time, the embedding module captured subtle phenotype behaviour that does not normally surface through fixed thresholds. By merging both streams, the system produces recommendations that are not only interpretable but also tuned to individual patient contexts.

Across the evaluation steps, the hybrid engine showed consistent internal behaviour. The neighbourhood patterns remained stable, and the rule triggers aligned with expected severity pathways. Small variations appeared due to noise in

laboratory values or missing features, but these fluctuations did not disrupt the core recommendation structure. The AI layer provided additional signals, especially around renal risk, utilization patterns, and VOC clustering. In several cases, the neighbour-derived advice supported or refined the rule-based output, a pattern that indicates the two modules complement each other rather than compete.

However, this work is not yet a complete clinical tool. Some of these decisions are strongly bound to the underlying dataset, which has a modest size and is limited to one setting. The embeddings would gain much from being larger and from more diverse cohorts-ideally including longitudinal data. More powerful learning architectures, such as graph networks or transformer-based clinical encoders, could also provide more insight into patient trajectories. Despite these limits, the framework presented here shows that explainable recommendations can be produced from multi-domain phenotypes without losing traceability. This hybrid approach provides a workable balance: the rules keep the reasoning accountable, while the similarity engine adjusts the outcome closer to a given patient's profile.

This work therefore demonstrates how hybrid recommenders can support clinicians in navigating the complexity of SCD, especially where manifestations significantly overlap and vary. The system is not intended to replace clinical judgment but can be useful in the management of early severity and structured follow-up planning. Future works will involve generalizing the model, enhancing fairness, and real-time integration of mobile or clinical interfaces.

Author Contributions: Conceptualization, A.R.S.K. and A.V.K.; methodology, A.R.S.K.; multi-modal GNN model design, A.R.S.K. and V.M.V.; patient similarity framework development, A.R.S.K. and V.R.N.; domain-rubric construction and clinical scoring logic, A.R.S.K. and E.S.A.A.; software implementation, A.R.S.K. and E.B.G.; validation, A.R.S.K., A.V.K. and V.M.V.; formal analysis, A.R.S.K.; investigation, A.R.S.K., E.S.A.A. and M.S.H.A.R.; resources, I.S.R.A.B. and M.S.H.A.R.; data curation, A.R.S.K. and V.R.N.; explainability analysis and visualization, A.R.S.K.; writing—original draft preparation, A.R.S.K.; writing—review and editing, A.R.S.K., A.V.K., V.M.V. and V.R.N.; supervision, A.V.K. and V.M.V.; project administration, M.S.H.A.R. and I.S.R.A.B.; funding acquisition, E.S.A.A. All authors have read and agreed to the published version of the manuscript.

Acknowledgements: We would like to express our gratitude to MoHERI, Oman for their generous financial support in making this project in Block Funding Call 2023 with ID BFP/RGP/ ICT/23/027. This funding has enabled us to conduct critical research and implement innovative solutions.

REFERENCES

Abdul, R. S. K., et al. (2024). Advancements in machine learning and deep learning techniques for sickle cell disease detection using digital morphology. *Proceedings of the 2024 International Conference on Computer and Applications (ICCA)*, Cairo, Egypt.

Alharbi, S. M., et al. (2021). Predicting vaso-occlusive crises using machine learning. *PLOS One*.

Ballas, J. (2015). Severity scoring in sickle cell disease. *British Journal of Haematology*.

Bender, S. A., et al. (2019). Clinical DSS for mobile health contexts. *JMIR mHealth*.

DeBaun, M. R., et al. (2020). Risk stratification in sickle cell disease. *Lancet Haematology*.

Doshi-Velez, F., & Kononenko, I. (2017). Explainable patient similarity. *Artificial Intelligence in Medicine*.

Ghassemi, M., et al. (2021). Hybrid rule–ML models in clinical decision support. *Nature Digital Medicine*.

Hebbel, M. (2020). Reconstructing sickle cell disease: A data-driven view. *Blood*.

Jiang, F., et al. (2019). Patient similarity networks in chronic disease analytics. *Journal of Biomedical Informatics*.

Kassim, R., et al. (2015). Clinical decision rules for SCD management. *American Journal of Hematology*.

Kato, S., et al. (2019). Organ damage scoring in SCD. *Haematologica*.

Kim, Y., et al. (2018). Patient similarity for precision oncology. *Cancer Research*.

Manwani, T., & Vichinsky, W. F. P. (2010). Sickle cell disease: Clinical features and management. *Hematology/Oncology Clinics*.

Marlin, L., et al. (2020). Multi-domain modeling of chronic disease states. *IEEE Transactions on Biomedical Engineering*.

Moor, T., et al. (2019). Patient similarity for sepsis stratification. *Critical Care Medicine*.

Nikolaou, S., et al. (2021). Similarity-based clinical reasoning. *Artificial Intelligence in Medicine*.

Niss, A., et al. (2018). Knowledge-based clinical systems for hematology. *Expert Systems with Applications*.
Osei, F., et al. (2021). Decision-support systems in low-resource hematology settings. *BMC Medical Informatics*.
Patel, V., et al. (2020). Mobile health decision support. *JMIR*.
Samek, W., et al. (2019). Explainable AI: Interpreting models. *Springer*.
Shamir, A., et al. (2022). Machine-learning predictions for acute SCD complications. *Computers in Biology and Medicine*.
Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. *JAMA*.
Sutton, R. T., et al. (2020). An overview of clinical decision support systems. *BMC Medical Informatics*.
Tran, D. N., et al. (2020). Readmission prediction models in SCD. *Journal of Clinical Medicine*.
Wang, H., et al. (2020). Integrating domain rules with patient similarity graphs. *IEEE Journal of Biomedical and Health Informatics*.
Zhang, G., et al. (2022). Neighbor-based cardiovascular risk modeling. *European Heart Journal Digital Health*.