# Cause2Cure: Bridging Prediction and Counterfactual Reasoning for Drug Effectiveness Analysis Across Patient Groups

Aditya Hambire[1], Shubham Jainapur[1], Kaustubh Jadhav[1], and Devanand Bathe[1]

K. J. Somaiya Institute of Technology, Mumbai
adityahambiree@gmail.com, shubham.jainapur@somaiya.edu, kaustubh.rj@somaiya.edu, devanandkb@somaiya.edu

**Abstract.** Predicting treatment effectiveness for Type 2 Diabetes Mellitus (T2DM) is challenging due to patient variability and complex longitudinal drug interactions [2,10]. While supervised machine learning offers predictive capabilities in healthcare, these techniques generally function as uninterpretable "black box" models lacking explanatory power [17]. To minimize the gap between prediction and decision support, this paper introduces Cause2Cure, an analytical framework combining three components: (i) ensemble-based treatment prediction using strictly baseline clinical data; (ii) unsupervised K-Means clustering to identify distinct glycaemic and behavioural patient subgroups; and (iii) a scenario-based counterfactual analysis engine to simulate evidence-based treatment modifications and quantify their predicted effectiveness [11, 16]. Evaluated on 4,200 patient records using 14 baseline features, a SMOTE-balanced [3,4,7,9] Random Forest classifier achieved approximately 70% accuracy and an ROC-AUC of 0.73. Ultimately, Cause2Cure provides an explainable, recommendation-focused foundation for Clinical Decision Support Systems (CDSSs) in chronic disease management.

**Keywords:** Drug Effectiveness Prediction · Type 2 Diabetes · Counterfactual Reasoning · Clinical Decision Support

## 1 Introduction

Type 2 Diabetes Mellitus (T2DM) is the most common form of diabetes, impacting over 537 million adults around the world. It is expected that by 2045, this number will rise to more than 783 million people [10]. T2DM treatment significantly impacts global healthcare systems, accounting for a substantial fraction of national healthcare spending [2].

Pharmacotherapies for T2DM include Metformin, Sulfonylureas, DPP-4 inhibitors, SGLT-2 inhibitors, and insulin. Despite clinical guidelines, individual response variability indicates that a one-size-fits-all approach is insufficient for most patients [2, 14].

Machine learning has successfully modeled clinical outcomes, with ensemble methods proving particularly effective for clinical data [3, 13]. However, many models focus on maximizing accuracy rather than providing actionable treatment suggestions for T2DM patients [17]. A further limitation is the conflation of statistical association with causation; ensemble models trained on observational data cannot determine causal relationships between interventions and T2DM status [18, 19, 21].

To minimize the constraints, we propose Cause2Cure. Cause2Cure uses ensemble methods to forecast treatment effectiveness, K-means clustering was used to find subgroups of patients responding best to specific plans, and scenario-based counterfactual methods to model the impact of interventions on treatment success [11, 16]. This approach moves beyond simple prediction to recommend effective treatment actions.

The main contributions of this work are:

- machine learning pipeline using strictly baseline (Visit 1) features, ensuring clinical deployability without temporal data leakage.
- Using SMOTE only on the training data to fix the imbalance between classes, while making sure the evaluation of the test data remains accurate and unaffected [4].
- A KMeans clustering analysis yielding four clinically interpretable patient subgroups characterised by glycaemic severity and adherence behaviour [1,6].
- A counterfactual simulation module implementing three evidence-based treatment modification scenarios with quantified predicted outcome changes [16, 25].
- An integrated interactive web dashboard enabling clinician-facing deployment.

The manuscript proceeds through the following sections: Section 2 provides an overview of existing literature and prior studies. The proposed analytical framework and computational steps are detailed in Section 3 and Section 4 lays out the empirical findings, while Section 5 provides a comprehensive evaluation of their clinical implications. Finally, Section 6 concludes the study and suggests avenues for future investigation.

## 2 Related Work

### 2.1 Machine Learning in Diabetes Management

A lot of research has been done on using computer models in T2DM care, especially in predicting when the condition might start, sorting out which patients are at higher risk, and figuring out how well different treatments might work [13,20]. Random Forest and Gradient Boosting are prominent models from the ensemble learning methods that have shown to be successful in structured clinical data sets because they can manage complex, non-linear relationships and take into consideration different patient features [3,8]. Random Forest models additionally

provide interpretable feature importance measures that facilitate clinical validation and acceptance. Although deep learning approaches have shown promise for longitudinal EHR modelling, their adoption in clinical decision support remains constrained by limited interpretability, high data requirements, and challenges in deployment [17, 20].

### 2.2 Patient Stratification and Personalised Medicine

Unsupervised clustering techniques have increasingly been used to identify clinically meaningful patient subgroups that may exhibit differential disease progression and treatment response. Ahlqvist et al. [1] identified five reproducible adult-onset diabetes subgroups using routinely collected clinical variables and demonstrated substantial differences in complication risk and therapeutic outcomes. Subsequent work by Dennis et al. [6] showed that data-driven subgrouping can improve prediction of treatment response compared with conventional clinical categorisation. More broadly, clustering approaches such as KMeans remain widely adopted for exploratory patient stratification due to their computational efficiency and interpretability [11].

### 2.3 Counterfactual Analysis and Explainability

As AI systems become more widely adopted in clinical settings, interpretability has become a critical need for reliable clinical AI. SHAP-based explanation methods provide feature-level attribution, which measures how much each variable contributes to the prediction [15]. However, such methods primarily explain why a prediction occurred rather than how it might be changed. Counterfactual explanation frameworks address this limitation by asking what minimal modifications to patient characteristics would alter a model prediction [25]. Recent studies have extended counterfactual reasoning toward diverse explanations and algorithmic recourse, enabling more actionable recommendations for decision support [12, 16]. Cause2Cure extends this paradigm by constructing clinically motivated multi-feature intervention scenarios aligned with evidence-based treatment strategies.

### 2.4 Clinical Decision Support Systems

Clinical Decision Support Systems (CDSSs) have evolved from rule based alert systems to data-driven platforms providing personalised recommendations. Roshanov et al. [22] reported that patient-specific recommendations integrated directly into clinical workflows significantly improve process outcomes. Latterly, there has been a growing focus on the role of AI-driven CDSSs in chronic disease management while simultaneously emphasising challenges related to transparency, clinician trust, and implementation [24]. Despite continued advances in predictive modelling, the integration of actionable recommendation mechanisms remains limited in many ML-based CDSSs, motivating the development of frameworks that combine prediction, explanation, and intervention-oriented reasoning.

**Table 1.** Baseline (V1) feature set used for prediction.

| **Category** | **Feature** | **Type** |
|---|---|---|
| Demographic | Age, Gender, BMI | Numeric / Binary |
| Clinical | $\text{HbA1c}_{\text{V1}}$ (%)<br>$\text{Fasting Glucose}_{\text{V1}}$ (mg/dL)<br>$\text{Blood Pressure}_{\text{V1}}$ (mmHg) | Numeric |
| Pharmacological | $\text{Metformin}_{\text{V1}}$<br>$\text{Sulfonylurea}_{\text{V1}}$<br>$\text{DPP-4 Inhibitor}_{\text{V1}}$<br>$\text{SGLT-2 Inhibitor}_{\text{V1}}$<br>$\text{Insulin}_{\text{V1}}$ | Ordinal (0–2) |
| Behavioural | Adherence, Diet, Exercise | Binary |
| **Target** | Treatment Effectiveness | Binary (0/1) |

## 3 Methodology

### 3.1 Preprocessing Pipeline

The dataset consists of 32 features, but the 14 features that are used as baseline features are selected and the outcome label is retained. The baseline feature set used for prediction is summarized in Table 1. The data is then randomly split into 80% training dataset and 20% testing dataset, while ensuring that the class proportions are maintained. Only the training set is used when applying SMOTE [4] and Linear interpolation with k-nearest neighbours in the feature space to create and generate synthetic samples for the minority class. This yields a balanced training set of 4,566 samples, while at the same time ensuring that the test-set evaluation reflects the true class distribution, which has not been modified. All clustering and distance-based operations are done with standardisation.

### 3.2 Predictive Modelling

Two ensemble classifiers are trained and compared as follows:

**Random Forest (RF)** The Gini impurity criterion has been used to build a Random Forest of 200 decision trees with maximum depth of the trees to 8. Class weights will be balanced as a secondary safeguard against imbalance. The feature importance is computed as the average (mean) of the decrease in impurity (MDI) from each individual tree [3].

**Gradient Boosting (GB)** The Gradient Boosting (GB) Classifier with 200 sequential estimators, `max_depth = 4` and `learning_rate = 0.1` is trained as

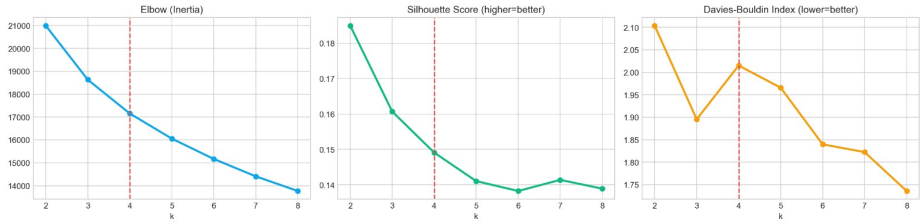


**Fig. 1.** Cluster evaluation metrics: Elbow (Inertia), Silhouette Score, and Davies-Bouldin Index. The optimum number of patient subgroups is found at $k = 4$ with a drop in the inflection..

a comparative baseline [8]. Gradient Boosting(GB) optimises a differentiable loss function through additive stage-wise fitting, often achieving higher bias reduction at the cost of increased training time [5].

### 3.3 Patient Stratification via KMeans Clustering

KMeans clustering [11] is applied to the full 4,200-patient dataset using six clinically relevant features: age, BMI, $\text{HbA1c}_{V1}$, fasting $\text{glucose}_{V1}$, blood $\text{pressure}_{V1}$, and adherence. All features are standardised prior to clustering.

The optimal cluster count is selected via the elbow method, evaluating within-cluster inertia, Silhouette Score, and Davies-Bouldin Index for $k \in \{2, 3, \ldots, 8\}$. Figure 1 shows the resulting curves; a pronounced inflection at $k = 4$ motivates the selection of four clusters. Each cluster is characterised by the within-cluster mean values of all clinical features and the treatment effectiveness rate.

### 3.4 Counterfactual Analysis Engine

The counterfactual module accepts a patient's baseline feature vector $x$ and the trained Random Forest model, generating predicted outcomes under three clinically motivated intervention scenarios $j \in \{1, 2, 3\}$. Let $\hat{p}_{\text{eff}}$ denote the model's estimated probability of treatment effectiveness. The predicted change is:

$$\Delta p^{(j)} = \hat{p}_{\text{eff}}\left(x^{(j)}\right) - \hat{p}_{\text{eff}}(x) \tag{1}$$

where $\hat{p}_{\text{eff}}\left(x^{(j)}\right)$ is the predicted effectiveness probability under intervention scenario $j$, and $\hat{p}_{\text{eff}}(x)$ is the original predicted effectiveness probability. The three scenarios are:

1. **Lifestyle Adherence Improvement:** Adherence, dietary compliance, and exercise frequency are set to 1, reflecting maximised non-pharmacological engagement.
2. **Medication Intensity Escalation:** Metformin and sulfonylurea dosages are incremented by one ordinal level (capped at 2), simulating dose up-titration under physician supervision.

**Table 2.** Comparative evaluation of the classification algorithms

| Predictive Classifier | AUC | F1-Score | Accuracy | Precision | Recall |
|---|---|---|---|---|---|
| Random Forest | 0.734 | 0.539 | 0.699 | 0.529 | 0.550 |
| Gradient Boosting | 0.730 | 0.536 | 0.704 | 0.537 | 0.535 |

3. **Combination Pharmacotherapy:** Metformin is set to high dose (2), DPP-4 and SGLT-2 inhibitors are set to at least low dose (1), and adherence is maximised, reflecting ADA/EASD guidelines for persistent glycaemic dyscontrol [2].

The scenario yielding the maximum $\Delta p$ is flagged as the recommended intervention. All outputs are explicitly framed as model-derived probability estimates rather than causal treatment-effect estimates.

### 3.5 System Architecture

Figure 2 represents the modular, full-stack architecture of Cause2Cure, designed for low-latency clinical decision support [24]. The user interface is a safeguarded, single-page React.js application tailored for medical professionals, featuring an interactive patient-input form restricted to essential baseline clinical data. To ensure usability in high-speed clinical settings, the dashboard dynamically displays prediction outputs, visual confidence indicators, and a separate counterfactual analysis panel that presents the "what if" delta changes in probability across simulated treatments.

Communication between the frontend and the isolated Machine Learning Engine is orchestrated by a REST API. The backend securely handles stateless, asynchronous requests across four main endpoints (`/predict`, `/counterfactual`, `/clusters`, and `/insights`), ensuring patient health information (PHI) is processed in-memory to comply with data privacy standards [23]. Upon receiving a validated JSON payload, the core ML engine performs concurrent inference using pre-trained models (Random Forest and Gradient Boosting) [3, 8]. It determines the patient's phenotype cluster, calculates the calibrated probability of treatment effectiveness, computes counterfactual adjustments, and returns the comprehensive results to the clinician dashboard.

The system architecture (illustrated in Fig. 2) consists of core data processing/prediction subsystems and modules providing interpretability, counterfactual analysis and generating recommendations.

## 4 Results

### 4.1 Classification Performance

The performance of both classifiers on the test dataset of 840 patients is summarized in Table 2. Both models achieve comparable overall accuracy ($\approx 70\%$).

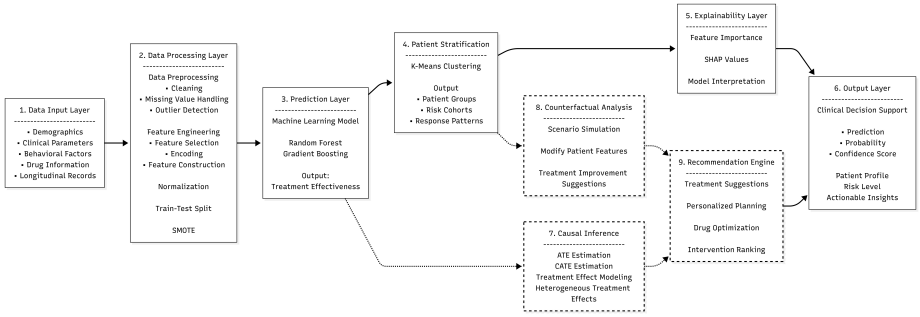


**Fig. 2.** End-to-end software and inference architecture of the Cause2Cure platform.

The Random Forest has a slightly better ROC-AUC of 0.734, which means it has better discriminative power for all classification thresholds, and is chosen as the production model. Gradient Boosting also shows a slight improvement in accuracy but a decrease in the AUC because of a conservative classification rule that prefers the majority class.

The low F1-score for the minority class (ineffective treatment) of 0.54 is due to the limited amount of cases in that class and also the challenge of the prediction task, and reflects the difficulty in the prediction task, consistent with similar published results for similar clinical prediction tasks [13,17]. The Random Forest confusion matrix is shown in Table 3. The model is effective in detecting the effectiveness of treatments, identifying 76.9% of effective treatments and 55.0% of ineffective treatments correctly. The most obvious clinical risk is the 132 false negatives (missed non-responders) which serves as a driving force for further architectural development.

**Table 3.** Confusion matrix for the Random Forest model (test set, $n = 840$).

| | **Pred: Effective** | **Pred: Not Eff.** |
|---|---|---|
| **Actual: Effective** | 439 (TP) | 132 (FN) |
| **Actual: Not Eff.** | 121 (FP) | 148 (TN) |

### 4.2 Feature Importance

The feature importances of the Random Forest model (means decrease in impurity) are displayed in Figure 3. The pre-treatment HbA1c level at 20.4% is the most important single predictor, and this finding is supported by clinical data showing that baseline glycaemic state is one of the most important factors to consider when determining a patient's treatment response [2,14]. Prescribing intensity and adherence are the second and third top modifiable drivers, with

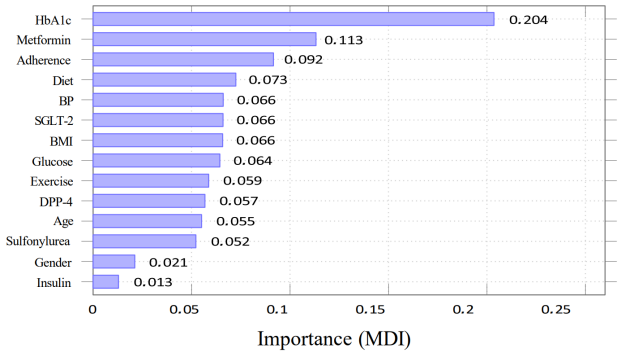


**Fig. 3.** Random Forest feature importance (MDI). Baseline HbA1c dominates, followed by metformin dosage and patient adherence.

**Table 4.** KMeans cluster profiles ($k = 4$). Eff. = effectiveness rate.

| ID | Profile | $n$ | HbA1c | Eff. |
|---|---|---|---|---|
| C0 | Young, Low-HbA1c, Non-Adherent | 925 | 6.22% | 57.1% |
| C1 | Young, Low-HbA1c, Adherent | 1326 | 6.22% | 68.2% |
| C2 | Older, High-HbA1c, Non-Adherent | 820 | 7.12% | 65.2% |
| C3 | Older, High-HbA1c, Adherent | 1129 | 7.18% | 78.6% |

dosages of metformin and adherence to it being the second and third respective drivers. Collectively, behavioural features (diet and exercise), account for approximately 13% of the predictive signal, highlighting the need to consider non-pharmacological factors.

### 4.3 Patient Stratification

The four patient cluster profiles derived by KMeans are summarised in Table 4. The clustering results yield a clinically coherent structure that includes two main dimensions: (i) age and severity of HbA1c and (ii) treatment adherence. Adherence is the single most effective within-stratum clinical lever, regardless of glycaemic severity, consistently improving the within-stratum effectiveness rate by 11–13 percentage points. The highest effectiveness is seen for cluster C3 (older, high HbA1c, adherent) due to the greater intensity of polypharmacy that is seen in this subgroup and the sustained adherence. However, a low level of adherence is associated with significantly poorer outcomes for patients with well controlled baseline glycaemia (Cluster C0: young, low-HbA1c, non-adherent) with the lowest rate (57.1%).

**Table 5.** Counterfactual simulation results ($\hat{p}_{\text{eff}}^{\text{orig}} = 41.2\%$).

| **Scenario** | $\hat{p}_{\text{eff}}^{\text{new}}$ | $\Delta p$ |
|---|---|---|
| S1: Lifestyle Adherence Improvement | 54.7% | +13.5 pp |
| S2: Medication Intensity Escalation | 48.3% | +7.1 pp |
| S3: Combination Pharmacotherapy | 61.9% | +20.7 pp |

### 4.4 Counterfactual Scenario Analysis

Table 5 presents counterfactual outputs for a representative Cluster C0 patient with an original predicted effectiveness probability of 41.2%. Combination pharmacotherapy with maximised adherence (Scenario 3) yields the largest predicted gain (+20.7 percentage points), consistent with contemporary ADA/EASD recommendations for persistent hyperglycaemia management [2].

Notably, lifestyle adherence improvement alone (Scenario 1) outperforms medication escalation (Scenario 2) in this patient profile, reinforcing prior evidence that engagement and adherence-focused interventions may confer benefits comparable to, or greater than, pharmacological intensification in selected patient populations [14].

### 4.5 HbA1c Temporal Trends

Figure 4 compares the mean HbA1c trajectory across Visits 1–3 for patients whose treatment was ultimately effective versus ineffective. Effective cases exhibit a steady decline from 6.81% to 6.44%, while ineffective cases show a more modest improvement from 7.24% to 6.89%, never converging to the effective cohort range.

This divergence underscores the predictive value of baseline HbA1c and further supports its dominant role in the Random Forest model (Figure 3), consistent with established evidence linking baseline glycaemic burden to treatment response and long-term diabetes outcomes [2, 14].

## 5 Discussion

### 5.1 Clinical Interpretation of Findings

The predictive performance of Cause2Cure (ROC-AUC = 0.734) is steady with published standards for baseline-only T2DM outcome prediction and is achieved under the strict constraint of excluding future visit data [13, 17].

The dominance of HbA1c and adherence as predictors is clinically intuitive. Pre-treatment glycaemic status is a well-established marker of disease severity and expected drug response, while adherence directly determines the effective dose delivered over time [2, 14].

The clustering results yield directly actionable insight. The systematic differentiation of subgroups by adherence rather than drug class suggests that, at

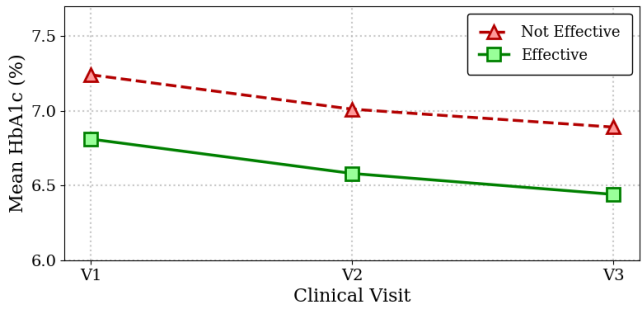


**Fig. 4.** Mean HbA1c trajectory across three clinical visits, stratified by treatment outcome. Effective cases show a steeper decline across all three visits.

the population level, who takes their medication is more predictive of treatment success than what medication is prescribed. This observation aligns with evidence that improvements in adherence may confer greater glycaemic benefit than equivalent pharmacological escalation in primary care settings [14].

### 5.2 The Prediction-to-Causation Gap

A central limitation of all components of Cause2Cure is that they operate on learned associations in observational data rather than true causal mechanisms [18]. The counterfactual engine produces model-derived probability estimates rather than causal effect estimates and is therefore susceptible to confounding and selection biases present in the training data. A patient whose high-adherence record is confounded with socioeconomic status, for example, may exhibit inflated predicted effectiveness that does not generalise to an isolated adherence intervention [19].

This limitation is acknowledged deliberately and motivates the system's positioning as a principled step toward causal inference rather than a complete causal framework. Formal causal integration would be required before counterfactual outputs could be interpreted as unbiased treatment-effect estimates [18, 21].

### 5.3 Comparison with Prior Systems

Cause2Cure distinguishes itself from existing T2DM clinical decision support systems on three axes. First, pharmacological dosage intensity and self-reported behavioural features are included, enriching the feature space beyond the laboratory values and administrative codes used in most EHR-based systems.

Second, the scenario-based counterfactual module moves beyond binary risk flags to provide ranked and quantified intervention recommendations, extending concepts introduced in counterfactual explanation and algorithmic recourse research [12, 16, 25].

Third, the strict enforcement of baseline-only inference directly addresses the data leakage concern that undermines several published predictive models for longitudinal clinical outcomes [20].

### 5.4 Limitations

Several limitations warrant acknowledgment. The F1-score for the ineffective treatment class (0.54) indicates that a substantial fraction of non-responders are missed, posing a safety concern in high-stakes settings without clinician oversight. This limitation may be related to the limited baseline feature set or the individual variation in response to the drug. The inclusion of other biomarkers (lipid profiles, renal function markers and genetic markers) in future studies could enhance recall for minorities.

KMeans assumes approximately spherical and similarly sized clusters, which may not reflect the true topology of the patient space. Better solutions for subgroup discovery are possible with alternative algorithms, such as Gaussian Mixture Models (GMMs) or HDBSCAN [11]. The framework performs well internally, but needs to be further externally validated on independent clinical cohorts in line with best practice for clinical prediction models reporting [23], and the system has not been prospectively evaluated for its impact on glycaemic outcomes. Finally, the dashboard has not yet undergone usability testing with practising clinicians, which remains a prerequisite for responsible real-world deployment [22, 24].

## 6 Conclusion

This paper presented Cause2Cure, an integrated framework for drug effectiveness prediction, patient stratification, and counterfactual treatment simulation in Type 2 Diabetes Mellitus (T2DM). The proposed system combines a SMOTE-balanced Random Forest classifier [3,4] (ROC-AUC = 0.734), KMeans-based patient stratification into four clinically interpretable subgroups [11], and a three-scenario counterfactual engine that quantifies the predicted benefit of evidence-based treatment modifications.

This study has the following key findings:

- Baseline HbA1c and treatment adherence emerged as the two most influential predictors of pharmacotherapeutic effectiveness.
- Adherence was identified as the primary differentiator of treatment outcomes across all glycaemic severity subgroups.
- Combination pharmacotherapy with maximised adherence produced the largest counterfactually predicted improvement in treatment effectiveness.

Overall, the framework establishes a principled and modular foundation for explanation-driven clinical decision support in chronic disease management while maintaining compatibility with baseline-only clinical deployment settings.

### 6.1 Future Work

Future research directions are:

- Doubly robust and Targeted Maximum Likelihood Estimation (TMLE) for formal causal effect estimation [18, 19].
- SHAP-based patient-level explainability integrated within the clinician dashboard [15].
- Prospective clinical validation in collaboration with endocrinology practitioners.
- Extension to multi-class outcome modelling with drug-class-specific effectiveness labels.
- Transformer-based temporal modelling capable of leveraging longitudinal multi-visit information while preserving deployment feasibility.

## References

1. Ahlqvist, E., et al.: Novel subgroups of adult-onset diabetes and their association with outcomes. The Lancet Diabetes & Endocrinology **6**(5), 361–369 (2018). https://doi.org/10.1016/S2213-8587(18)30051-2
2. American Diabetes Association: Standards of care in diabetes—2025. Diabetes Care **48**(Supplement_1) (2025). https://doi.org/10.2337/dc25-S001
3. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
4. Chawla, N.V., et al.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research **16**, 321–357 (2002). https://doi.org/10.1613/jair.953
5. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. ACM (2016). https://doi.org/10.1145/2939672.2939785
6. Dennis, J.M., Shields, B.M., Henley, W.E., et al.: Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features. The Lancet Diabetes & Endocrinology **7**(6), 442–451 (2019). https://doi.org/10.1016/S2213-8587(19)30087-7
7. Fernández, A., Garcia, S., Herrera, F., Chawla, N.V.: Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of Artificial Intelligence Research **61**, 863–905 (2018). https://doi.org/10.1613/jair.1.11192
8. Friedman, J.H.: Greedy function approximation: A gradient boosting machine. The Annals of Statistics **29**(5), 1189–1232 (2001). https://doi.org/10.1214/aos/1013203451
9. He, H., Garcia, E.A.: Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering **21**(9), 1263–1284 (2009). https://doi.org/10.1109/TKDE.2008.239
10. International Diabetes Federation: IDF Diabetes Atlas, 10th ed. IDF, Brussels (2021), https://www.diabetesatlas.org
11. Jain, A.K.: Data clustering: 50 years beyond k-means. Pattern Recognition Letters **31**(8), 651–666 (2010). https://doi.org/10.1016/j.patrec.2009.09.011

12. Karimi, A.H., Schölkopf, B., Valera, I.: Algorithmic recourse: from counterfactual explanations to interventions. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. pp. 353–362. ACM (2021). https://doi.org/10.1145/3442188.3445899
13. Kavakiotis, I., et al.: Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal **15**, 104–116 (2017). https://doi.org/10.1016/j.csbj.2017.01.005
14. Khunti, K., Seidu, S., Kunutsor, S., Davies, M.: Association between adherence to pharmacotherapy and outcomes in type 2 diabetes. Diabetes Care **40**(11), 1588–1596 (2018). https://doi.org/10.2337/dc17-2172
15. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30 (2017), https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
16. Mothilal, R.K., Sharma, A., Chen, C.: Explaining machine learning classifiers through diverse counterfactual explanations. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 607–617. ACM (2020). https://doi.org/10.1145/3351095.3372850
17. Obermeyer, Z., Emanuel, E.J.: Predicting the future — big data, machine learning, and clinical medicine. New England Journal of Medicine **375**(13), 1216–1219 (2016). https://doi.org/10.1056/NEJMp1606181
18. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press, 2nd edn. (2009). https://doi.org/10.1017/CBO9780511803161
19. Prosperi, M., Guo, Y., Sperrin, M., et al.: Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nature Machine Intelligence **2**(7), 369–375 (2020). https://doi.org/10.1038/s42256-020-0197-y
20. Rajkomar, A., Oren, E., Chen, K., et al.: Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine **1**(1), 1–10 (2018). https://doi.org/10.1038/s41746-018-0029-1
21. Richens, J.G., Lee, C.M., Johri, S.: Improving clinicians' diagnostic performance with causal machine learning. Nature Communications **11**(1), 4308 (2020). https://doi.org/10.1038/s41467-020-17419-7
22. Roshanov, P.S., et al.: Features of effective computerised clinical decision support systems. BMJ **346**, f1657 (2013). https://doi.org/10.1136/bmj.f1657
23. Steyerberg, E.W.: Clinical Prediction Models: A Practical Approach to Development, Validation, and Reporting. Springer Nature, Cham, 2nd edn. (2019). https://doi.org/10.1007/978-3-030-16399-0
24. Sutton, R.T., Pincock, D., Baumgart, D.C., et al.: An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digital Medicine **3**(1), 1–10 (2020). https://doi.org/10.1038/s41746-020-0221-y
25. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box. Harvard Journal of Law & Technology **31**(2), 841–887 (2017), https://ssrn.com/abstract=3063289