

Development and Validation of a Predictive Model for Glycemic Response to Dorzagliatin in Type 2 Diabetes: A Retrospective Real-World Study

<https://doi.org/10.63174/xdi.DBXS5758>

Yue Liu¹, Jia Liu¹, Zhongqing Yan¹, Ying Chen¹, Liping Han^{1,*}

Received: 06 Oct 2025

Accepted: 24 Oct 2025

Published: 1 Nov 2025

Open Access



Abstract: Patient response to dorzagliatin, a novel glucokinase activator for type 2 diabetes mellitus (T2DM), exhibits significant heterogeneity in clinical practice, posing a challenge for precision medicine. This study aimed to develop and validate a predictive model for identifying optimal responders prior to treatment initiation. We conducted a retrospective cohort study involving 300 T2DM patients hospitalized between April 2024 and April 2025. Patients were categorized into a responder group (n=146, fasting/postprandial glucose reduction >0.5 mmol/L) and a non-responder group (n=154, reduction <0.3 mmol/L). Multivariable logistic regression and random forest machine learning algorithms were employed. Key predictors identified included diabetes duration, HOMA- β index, and use of GLP-1 receptor agonists and metformin. The logistic regression model demonstrated high predictive accuracy (87.0%), while the random forest model achieved an AUC of 0.99 in the test set. A comprehensive model integrating the most significant variables yielded a superior AUC of 0.900. This clinically applicable prediction model facilitates the pre-treatment identification of patients most likely to benefit from dorzagliatin, thereby advancing personalized T2DM management.

1. Introduction

Type 2 diabetes mellitus (T2DM) represents a formidable global health challenge, characterized by a complex interplay of insulin resistance and progressive β -cell dysfunction. Its escalating prevalence, driven by demographic shifts and the obesity pandemic, underscores the urgent need for innovative therapeutic strategies that move beyond mere glycemic reduction to address the underlying pathophysiology of the disease^[1]. While metformin remains the cornerstone first-line therapy, a significant proportion of patients fail to achieve or sustain glycemic targets with conventional treatments, creating a substantial unmet medical need for agents with novel mechanisms of action^{[2],[3]}.

The pursuit of physiological glucose regulation has led to the development of glucokinase activators (GKAs), which target the central glucose sensor in both pancreatic β -cells and hepatocytes. Dorzagliatin, a first-in-class dual-acting GKA, has emerged as a promising therapeutic option by enhancing glucokinase activity in a glucose-dependent manner, thereby restoring glucose-stimulated insulin secretion and improving hepatic glucose metabolism^{[1],[4]}. Pivotal phase III trials have demonstrated its robust efficacy in both drug-naïve patients and those inadequately controlled on metformin monotherapy, with significant reductions in HbA1c and favorable safety profiles^{[2],[4]}. Furthermore, the drug's compatibility with other glucose-lowering agents, such as sitagliptin, without clinically meaningful pharmacokinetic interactions, enhances its potential for combination therapy^[5].

Despite these encouraging results, real-world clinical implementation faces a critical challenge: the substantial interindividual variability in treatment response. Clinical observations reveal that while some patients experience profound glycemic improvements with dorzagliatin, others demonstrate suboptimal responses, potentially related to heterogeneous factors including residual β -cell function, diabetes duration, presence of comorbidities, and concomitant medications^[3]. This therapeutic heterogeneity underscores the limitations of a one-size-fits-all approach and highlights the imperative for precision medicine in T2DM management. Currently, no validated predictive tools exist to identify optimal responders before treatment initiation, potentially limiting the drug's cost-effective deployment and clinical utility.

The integration of machine learning (ML) methodologies into clinical prediction represents a paradigm shift in addressing such challenges. While tradi-

tional statistical approaches like logistic regression have demonstrated robust performance in various medical contexts, including diabetes classification^[6], ensemble methods such as Random Forest (RF) have shown superior predictive accuracy in complex clinical scenarios. RF algorithms have excelled in predicting diverse outcomes including sepsis mortality^[7], postoperative pain^[8], vertebral fracture risk^[9], and breast cancer events^[10]. In diabetes care specifically, RF has outperformed conventional models in predicting carotid intima-media thickness^[11], Parkinson's disease progression in diabetic cohorts^[12], and nocturnal hypoglycemia^[13]. However, the optimal model selection remains context-dependent, with some studies demonstrating comparable performance between simpler logistic regression and more complex ML algorithms^{[6],[14]}.

This study aims to bridge the critical gap between dorzagliatin's proven efficacy and its precision application by developing and validating a comprehensive predictive model for glycemic response. Leveraging a retrospective cohort of T2DM patients initiating dorzagliatin therapy, we will employ a dual-model approach integrating multivariate logistic regression with Random Forest machine learning. This methodological synergy capitalizes on the interpretability of traditional regression while harnessing the pattern-recognition capabilities of ensemble learning to identify key clinical, biochemical, and treatment-related predictors of therapeutic response. The resulting predictive tool seeks to facilitate personalized treatment selection, optimize resource allocation, and ultimately enhance therapeutic efficacy in real-world clinical practice, moving toward a more nuanced, patient-centric approach to T2DM management.

2. Method

2.1. Study Design

Single-center retrospective cohort of type 2 diabetes patients receiving dorzagliatin (April 2024-April 2025).

2.1.1. Grouping Criteria

Based on 3-day glycemic response: 1. Responders: FPG reduction ≥ 1.0 mmol/L or PPG reduction ≥ 2.0 mmol/L. 2. Non-responders: Failed to meet above criteria.

¹NHC Key Lab of Hormones and Development and Tianjin Key Lab of Metabolic Diseases Tianjin Medical University Chu Hsien-I Memorial Hospital & Institute of Endocrinology, Tianjin 300134, China.

* Corresponding Author: E-mail address: hlp710816@163.com

Table 1. Clinical characteristics of the two groups.

Variables	A(n=146)	B(n=154)	P value
Age(years)	57.28±7.5	58.17±7.5	0.434
Sex n(%)			
Male	66 (45.2)	69 (44.8)	0.859
Duration(years)	6.3±2.8	8.8±3.0	<0.001
BMI(kg/m ²)	27.1±3.4	27.1±3.4	0.966
Smoke n(%)			
Yes	63 (43.2)	60 (39.0)	0.461
Drink n(%)			
Yes	44 (30.1)	48 (31.2)	0.846
Comorbidity n(%)			
Fatty liver	86 (58.9)	121 (78.6)	<0.001
Coronary heart disease	82 (56.2)	79 (51.3)	0.398
Hypertension	76 (52.1)	74 (48.1)	0.488
Cerebral infarction	46 (31.5)	53 (34.4)	0.592
Combination medications n(%)			
Metformin	80 (54.8)	67 (43.5)	0.049
DPP-4i	39 (26.7)	41 (26.6)	0.986
SGLT-2i	81 (55.5)	73 (47.4)	0.162
GLP-1RA	107 (73.3)	36 (23.4)	<0.001
TZD	83 (56.8)	96 (62.3)	0.333
AGI	74 (50.7)	75 (48.7)	0.731
insulin	77 (52.7)	78 (50.6)	0.717

*BMI:body mass index, DPP-4i:dipeptidyl peptidase-4 inhibitors, SGLT-2i:sodium-glucose cotransporter-2 inhibitors, AGI:α-glucosidase inhibitors,Pio pioglitazone, GLP-1RA: glucagon-like peptide-1 receptor agonists

Table 2. Comparisons of the parameters of islet β-cell function.

Variables	A(n=146)	B(n=154)	P value
Fasting plasma glucose (mmol/L)	10.2±2.2	10.0±1.8	0.754
Baseline of HbA1c(%)	9.5±1.9	9.7±1.9	0.444
Fasting insulin(μU/mL)	18.1±5.5	16.5±6.0	0.001
Fasting C-peptide (ng/mL)	1.7±0.4	1.6±0.3	0.982
HOMA-IR	3.6±0.4	3.5±0.4	0.369
HOMA-β	116.3±20.5	91.0±23.4	<0.001

*HbA1c: Hemoglobin A1c; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; HOMA-β: Homeostatic Model Assessment of Beta-cell Function

2.1.2 Data Collection

Collected demographics, comorbidities, laboratory parameters (glucose metabolism, liver/kidney function, lipids, hematology), and medications.

2.2. Statistical Analysis

SPSS 22.0 for conventional statistics; R 4.5.1 for random forest analysis. A rigorous validation strategy was used to ensure model robustness and prevent overfitting. The dataset was randomly split into a 70% training set for model development and a 30% internal test set for unbiased evaluation. During training, 10-fold cross-validation was applied to optimize parameters and obtain reliable performance estimates. Model performance was ultimately validated using the held-out test set. Continuous variables analyzed by t-test or Mann-Whitney U test; categorical by χ^2 test. Significant univariate predictors entered multivariate logistic regression.

2.3. Eligibility

Included treatment-naive dorzagliatin users meeting 2025 ADA criteria. Excluded patients with severe hepatic/renal impairment, acute complications, malignancy, or heart failure.

2.4. Ethical Approval and Consent Statement

This retrospective study was conducted in accordance with the ethical principles of the Declaration of Helsinki and received approval from the Institutional Ethics Committee of [Tianjin Medical University Chu Hsien-I Memorial Hospital]. The committee granted a waiver of informed consent for this study. All patient data were anonymized and maintained with strict confidentiality throughout the research process.

3. Results

3.1. Baseline Characteristics and Univariate Analysis

This study included 300 type 2 diabetes patients treated with dorzagliatin, comprising 146 responders (Group A) and 154 non-responders (Group B). The

two groups were comparable in basic demographic characteristics including age, gender, and smoking history ($P>0.05$) (Table 1).

Group B had significantly longer diabetes duration (8.8±3.0 vs. 6.3±2.3 years, $P<0.001$) and higher prevalence of fatty liver (78.6% vs. 58.9%, $P<0.001$) (Table 1, Figure. 1). Regarding medication use, Group A showed higher utilization of GLP-1 receptor agonists (73.3% vs. 23.4%, $P<0.001$) and metformin (54.8% vs. 43.5%, $P=0.049$). Metabolic parameters revealed that Group A had significantly higher HOMA-β (116.3±20.5 vs. 91.0±23.4, $P<0.001$) and fasting insulin levels (18.1±5.5 vs. 16.5±6.0 μU/mL, $P=0.001$) (Table 2, Figure. 1), while Group B had elevated triglyceride levels (3.3±1.6 vs. 3.1±1.4 mmol/L, $P=0.024$) (Table 3, Figure. 1). No significant differences were observed in core glycemic parameters including HbA1c, fasting plasma glucose, HOMA-IR, or fasting C-peptide.

3.2. Multivariate Logistic Regression Analysis

Multivariate logistic regression identified four independent predictors (Table 4). Metformin use (OR=4.927, 95%CI: 4.512-7.123, $P=0.001$) and diabetes duration (OR=1.37, 95%CI: 1.21-1.55, $P<0.001$) were independent positive correlates for treatment efficacy, while GLP-1 use (OR=0.005, 95%CI: 0.001-0.042, $P<0.001$) and HOMA-β levels (OR=0.95, 95%CI: 0.94-0.97, $P<0.001$) were independent negative correlates. The model demonstrated significant statistical power ($\chi^2=204.67$, $P<0.001$) with good fit (Nagelkerke $R^2=0.66$) and 87.0% prediction accuracy.

3.3. Random Forest Model Analysis

The random forest model exhibited excellent predictive performance with 99% accuracy in the test set and AUC of 0.99 (95%CI: 0.97-1.00). Variable importance ranking identified diabetes duration (Mean Decrease Gini=28.09) and HOMA-β index (22.65) as the most important predictors (Figure. 2).

3.4. ROC Curve Analysis

Univariate ROC analysis demonstrated that HOMA-β provided the best predictive performance (AUC=0.847), followed by GLP-1RA use (AUC=0.746),

Table 3. Comparisons of other results of laboratory examination.

Variables	A(n=146)	B(n=154)	P value
CBC			
RBC($\times 10^{12}/L$)	5.4 \pm 1.2	5.5 \pm 1.2	0.587
WBC($\times 10^9/L$)	7.9 \pm 1.84	8.2 \pm 1.85	0.331
Plt($\times 10^9/L$)	245.0 \pm 85.0	245.9 \pm 86.3	0.662
Hb(g/L)	18.7 \pm 13.0	119.4 \pm 12.1	0.814
Liver function			
ALB(g/L)	36.5 \pm 1.9	36.7 \pm 1.9	0.477
TBIL(μ mol/L)	16.5 \pm 1.9	16.7 \pm 1.9	0.477
ALT(U/L)	46.6 \pm 19.6	46.2 \pm 20.0	0.870
AST(U/L)	38.0 \pm 21.5	36.3 \pm 21.1	0.414
Renal function			
Scr(μ mol/L)	95.8 \pm 53.9	98.6 \pm 58.4	0.925
eGFR($mL/min/1.73m^2$)	74.7 \pm 11.2	76.38 \pm 11.51	0.199
UA(μ mol/L)	384.3 \pm 41.8	379.5 \pm 43.3	0.498
ACR(mg/g)	126.93 \pm 19.1	139.9 \pm 20.1	0.892
Electrolytes			
K+(mmol/L)	4.0 \pm 0.8	4.0 \pm 1.1	0.491
Na+(mmol/L)	140.20 \pm 1.98	139.96 \pm 1.93	0.274
Ca2+(mmol/L)	2.2 \pm 0.4	2.0 \pm 0.6	0.114
serum lipid			
TC(mmol/L)	5.4 \pm 2.5	5.3 \pm 2.1	0.730
TG(mmol/L)	3.1 \pm 1.4	3.3 \pm 1.6	0.024
HDL-C(mmol/L)	2.0 \pm 1.0	1.8 \pm 0.9	0.370
LDL-C(mmol/L)	3.0 \pm 1.4	2.8 \pm 1.3	0.726
Thyroid ultrasound			
Abnormal n(%)	103 (70.5)	110 (71.4)	0.859
Lower extremity arterial disease n(%)			
0	18 (12.0%)	13 (8.7%)	0.505
1	90 (60.0%)	88 (58.7%)	
2	42 (28.0%)	49 (32.7%)	
Carotid artery lesion n(%)			
0	12 (8.0%)	17 (11.3%)	0.488
1	90 (60.0%)	92 (61.3%)	
2	48 (32.0%)	41 (27.3%)	

*RBC: Red Blood Cell Count; WBC: White Blood Cell Count; Plt: Platelet Count; Hb: Hemoglobin; ALB: Albumin; TBIL: Total Bilirubin; ALT: Alanine Aminotransferase; AST: Aspartate Aminotransferase; Scr: Serum Creatinine; eGFR: Estimated Glomerular Filtration Rate; UA: Uric Acid; ACR: Albumin-to-Creatinine Ratio; TC: Total Cholesterol; TG: Triglycerides; HDL-C: High-Density Lipoprotein Cholesterol; LDL-C: Low-Density Lipoprotein Cholesterol

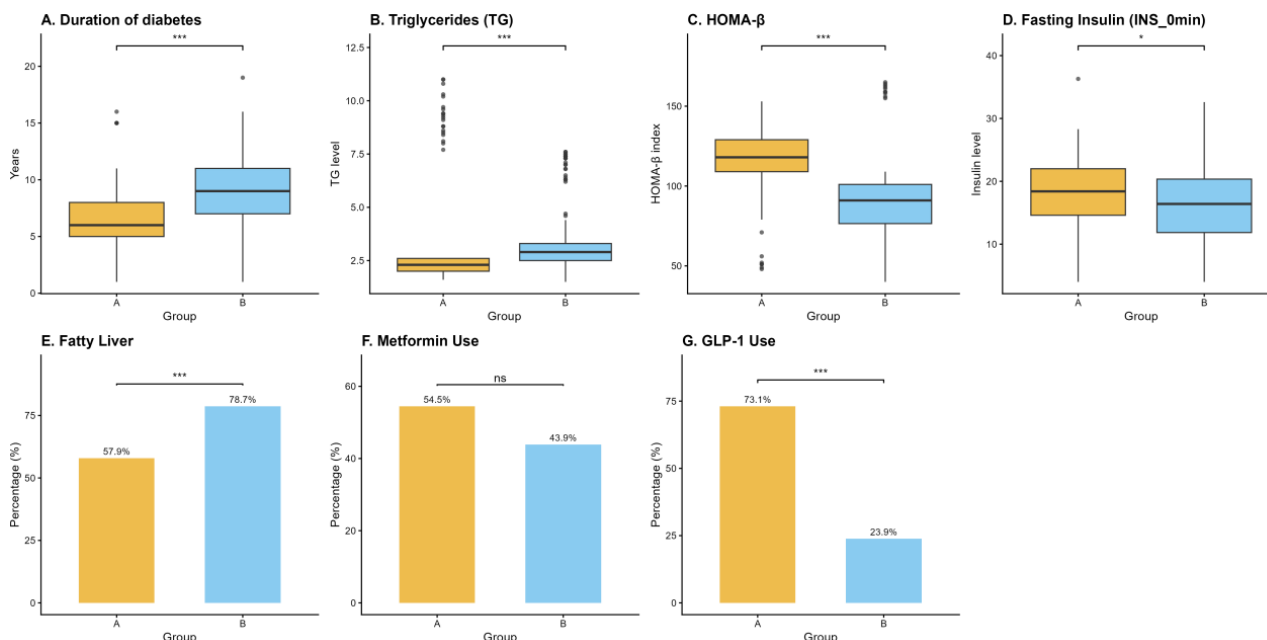


Figure 1. Comparisons of diabetes duration, TG, HOMA-β, fasting insulin, fatty liver, metformin use, GLP-1RA use between two groups

diabetes duration (AUC=0.743), and triglycerides (AUC=0.710). The four-variable combined model achieved an AUC of 0.900, representing a 0.054 improvement over the best single variable (HOMA-β), with comparable performance to the seven-variable model (AUC=0.919) (Figure 3).

3.5. Methodological Comparison

Both multivariate logistic regression and random forest models consistently

identified diabetes duration, HOMA-β, GLP-1 use, and metformin use as key predictors. The random forest model demonstrated superior prediction accuracy (99% vs. 87%), while logistic regression provided OR values that offered more intuitive clinical risk assessment.

4. Discussion

This retrospective cohort study successfully developed and validated a com-

Table 4. results of multivariate logistic regressions analysis

Variables	β	P	OR	95%CI
Duration	0.315	<0.001	1.371	1.211-1.552
Fatty liver	0.198	0.582	1.219	0.603-2.464
Metformin	3.712	0.001	4.927	4.512-7.123
GLP-1RA	-5.291	<0.001	0.005	0.001-0.042
Fasting insulin	-0.022	0.492	0.978	0.919-1.041
HOMA- β	-0.050	<0.001	0.951	0.936-0.966
TG	0.071	0.405	1.074	0.909-1.269

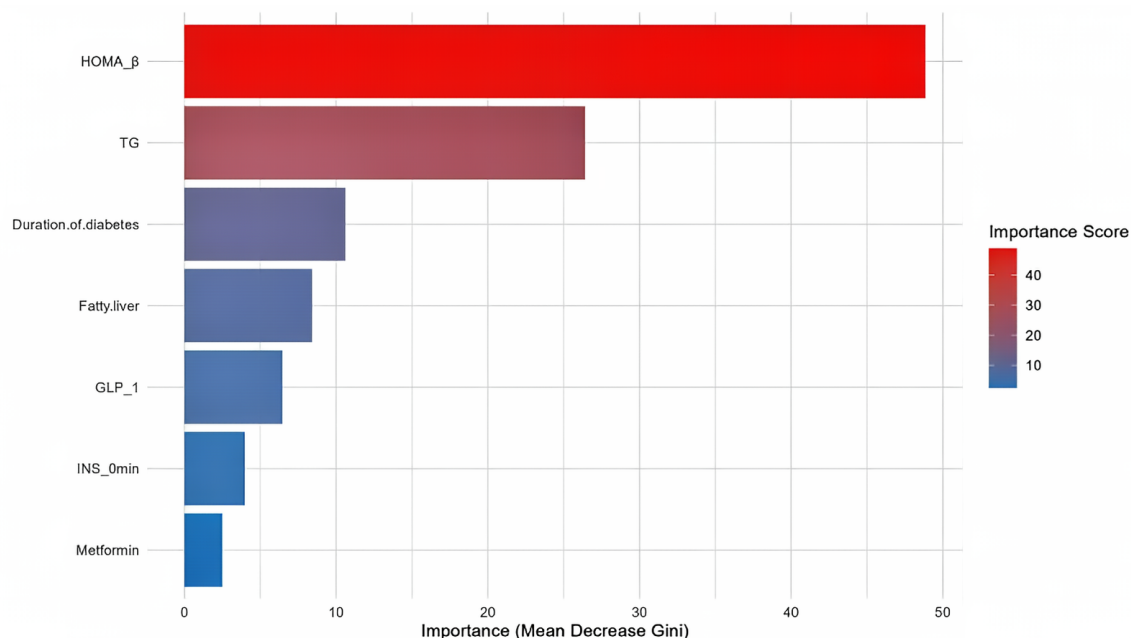


Figure 2. Key factors predicting Dorzagliatin efficacy

prehensive predictive model for glycemic response to dorzagliatin in patients with type 2 diabetes mellitus (T2DM) using real-world clinical data. By synergistically integrating multivariable logistic regression with a random forest machine learning algorithm, we identified and quantified several key predictors, including HOMA- β , diabetes duration, and concomitant use of GLP-1 receptor agonists (GLP-1 RAs). The final composite model demonstrated exceptional discriminative performance (AUC=0.900), positioning it as a potentially powerful tool for optimizing patient selection and personalizing treatment strategies in clinical practice, thereby addressing a significant unmet need in the era of precision diabetology.

Our analysis robustly identified HOMA- β as the single most powerful predictor of dorzagliatin response, exhibiting the strongest negative correlation in logistic regression (OR=0.951) and achieving the highest rank in random forest variable importance, with an individual ROC-AUC of 0.847. This finding provides compelling mechanistic validation for the proposed mode of action of glucokinase activators (GKAs) like dorzagliatin, which function primarily by enhancing glucokinase activity in pancreatic β -cells and hepatocytes^[1]. The inverse relationship suggests that patients with higher preserved β -cell function, as quantified by HOMA- β , possess a greater intrinsic capacity for glucose-sensing and insulin secretory response, which can be effectively potentiated by GKAs^[15]. This aligns with the fundamental pathophysiology of T2DM, where progressive β -cell failure ultimately limits the efficacy of most insulin secretagogues. Our data thus strongly advocate for the stratification of T2DM patients based on residual β -cell function when considering dorzagliatin therapy. It is noteworthy that HOMA- β and diabetes duration, as two core predictors, are physiologically interrelated. However, they capture different dimensions of T2DM heterogeneity. HOMA- β reflects the immediate insulin secretory capacity, while disease duration likely represents the cumulative metabolic burden and potential irreversible damage to β -cells over time. Therefore, combining both may more accurately define the disease continuum—from early-stage with preserved β -cell function to advanced-stage with β -cell failure^[2]. This approach could optimize the selection of target populations for GKAs.

Complementing this finding, our study revealed that shorter diabetes duration served as a significant positive predictor of treatment response (OR=1.371, AUC=0.743). This observation powerfully underscores the critical importance of early pharmacological intervention in T2DM management. A shorter

disease duration is intrinsically linked to a greater preserved functional β -cell mass and reduced severity of insulin resistance, creating a more favorable pathophysiological substrate for a drug that augments the body's own glucose-responsive insulin secretion. This finding resonates with the natural history of T2DM progression and echoes the therapeutic principle observed with other agents, where early intensive control yields superior and more durable outcomes^[16]. It suggests the existence of a potential “therapeutic window” for GKAs early in the disease course, a concept that warrants further prospective investigation. Our results thereby add to the growing body of evidence advocating for a proactive, personalized approach to T2DM treatment selection from diagnosis, rather than a uniform, stepwise protocol^[17].

A particularly intriguing and somewhat unexpected finding was the potent negative association between concomitant GLP-1 RA use and dorzagliatin response (OR=0.005, AUC=0.746). This robust statistical signal necessitates careful and nuanced interpretation. On one hand, it may reflect a significant confounding by indication, wherein patients prescribed GLP-1 RAs inherently represent a distinct clinical phenotype, typically characterized by more advanced disease, higher cardiovascular risk, greater obesity, or prior failure of multiple antidiabetic agents^{[18],[19]}. In such a cohort with potentially diminished β -cell reserve, the efficacy of any insulin secretagogue, including dorzagliatin, would understandably be blunted. The observed negative correlation is more likely an indicator of underlying disease severity rather than a direct drug-drug antagonism. This association was particularly pronounced in the subgroup of patients with lower HOMA- β values. These findings support the following biological interpretation: when intrinsic β -cell function is severely compromised, even potent insulin secretion stimulation by GLP-1 RAs through the cAMP signaling pathway reaches a “ceiling effect” due to substantially reduced β -cell mass and functional exhaustion^[4]. Under such conditions, adding a glucose kinase activator that primarily enhances glucose sensing provides minimal additional stimulation to insulin secretion. On the other hand, a pharmacodynamic or pharmacokinetic interaction cannot be ruled out. Both drug classes ultimately augment glucose-dependent insulin secretion, albeit via distinct molecular targets—dorzagliatin via glucokinase and GLP-1 RAs via the incretin receptor. Theoretical concerns regarding receptor downregulation, β -cell exhaustion, or competition in downstream signaling pathways in the presence of potent dual secretagogue stimulation merit further basic and clinical investi-

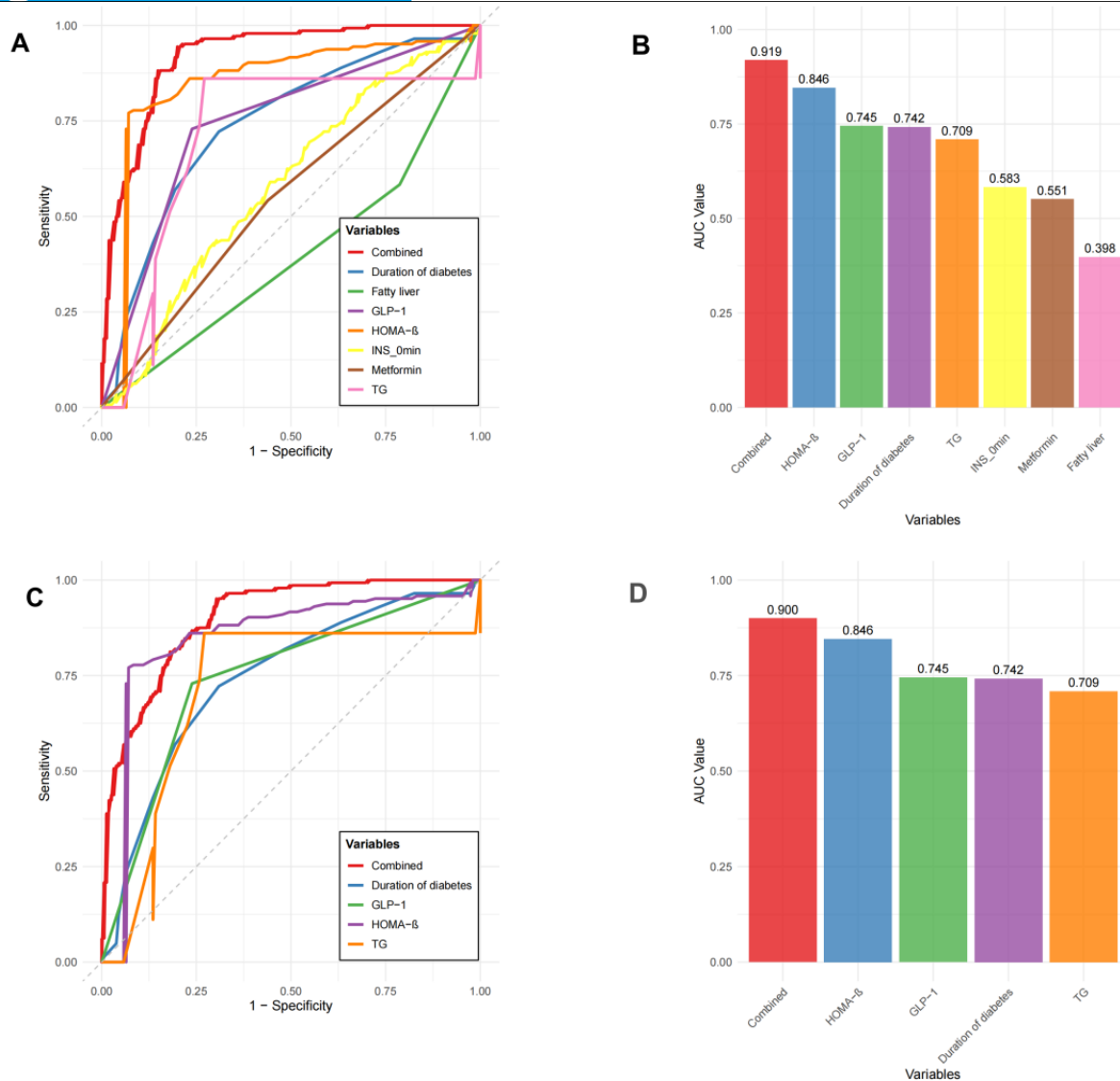


Figure 3. Univariate ROC analysis results

gation^[20]. Future studies designed to explore this interaction are crucial, as the combination of a GKA and a GLP-1 RA could, paradoxically, be either beneficial or antagonistic depending on the patient's specific metabolic context. The predictive model itself, which integrated seven significant clinical and biochemical variables, represents the cornerstone of this study's contribution. Its outstanding performance (AUC=0.900) significantly surpassed that of any single predictor, validating the superiority of a multimodal, integrative approach over reliance on isolated parameters. The complementary use of traditional logistic regression, which provides transparent, interpretable odds ratios, with the robust, non-parametric random forest algorithm, which captures complex non-linear relationships and interactions, represents a sophisticated methodology for clinical prediction rule development. The fact that this high level of accuracy was achieved using routinely available clinical data enhances its potential for real-world implementation and integration into electronic health record systems to support clinical decision-making. This approach aligns with the broader movement towards data-driven precision medicine in diabetes care, aiming to move beyond the "one-size-fits-all" paradigm and instead match the right patient with the right drug at the right time^[21] mean body mass index $35.4 \pm 6.4 \text{ kg/m}^2$.

Notwithstanding its strengths, our study has several limitations that must be acknowledged. The single-center, retrospective design inevitably introduces the potential for selection bias and may limit the generalizability of our findings to broader, more diverse T2DM populations. While internal validation was performed, the near-perfect accuracy (99%) reported by the random forest model raises a legitimate concern for overfitting, a common pitfall in machine learning applied to finite datasets. External validation in independent, multi-ethnic cohorts is an essential next step to confirm the model's robustness

and transportability. Furthermore, the "black box" nature of complex machine learning models, including random forests, somewhat compromises clinical interpretability, making it challenging to fully deconstruct the decision-making process for individual predictions. The enigmatic negative association with GLP-1 RAs, while statistically strong, is susceptible to residual confounding, as data on precise indications for GLP-1 RA therapy, duration of use, and detailed metrics of disease severity were not fully available. Finally, the relatively short observation period precludes any assessment of the model's ability to predict long-term durability of glycemic response or its impact on hard clinical endpoints, such as microvascular and macrovascular complications. Despite these limitations, our work holds significant clinical implications. We provide a validated, practical tool to help clinicians identify patients most likely to benefit from the novel agent dorzagliatin. The four-factor predictive framework demonstrates considerable potential for clinical translation. This model could be integrated into clinical decision support systems, such as EMR-embedded tools or mobile health applications. Using routine clinical data, these tools can generate individualized response probabilities for dorzagliatin therapy. Further validation in large-scale prospective cohorts is required before implementing this model in clinical practice. Our findings also reinforce fundamental pathophysiological principles in T2DM management, highlighting the central role of preserved β -cell function and the advantage of early intervention. Future research should focus on prospective, multi-center validation studies. It should also explore integrating omics data to enhance predictive accuracy and investigate the mechanistic basis of the observed interaction with GLP-1 RAs. In summary, this study represents a significant step toward personalized medicine in T2DM. We provide a data-driven framework to optimize dorzagliatin use, potentially paving the way for similar approaches

5. Conclusion

This study identified HOMA- β , use of GLP-1 receptor agonists and metformin, and diabetes duration as key predictors of the glycemic response to dorzagliatin. The prediction model developed based on these factors demonstrated robust discriminatory ability, offering a valuable tool for optimizing patient selection and advancing personalized treatment in type 2 diabetes.

Funding

No funding.

Acknowledgements

We thank Head Nurse Liu Lele for her valuable assistance in this study.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Yue Liu: Conceptualization, Investigation, Formal analysis, Writing - Original Draft. Jia Liu: Data Curation, Visualization, Writing - Review & Editing. Zhongqing Yan: Formal analysis, Validation. Ying Chen: Resources, Data Curation. Liping Han (Corresponding Author): Supervision, Conceptualization, Writing - Review & Editing, Project administration.

References

- [1] Y. Jiang, L. Wang, Z. Dong, B. Xia, S. Pang. "Recent drug development of dorzagliatin, a new glucokinase activator, with the potential to treat type 2 diabetes: a review study." *J. Diabetes*. **2024**, *16*, 6, e13563.
- [2] W. Yang, D. Zhu, S. Gan, X. Dong, J. Su, W. Li, H. Jiang, W. Zhao, M. Yao, W. Song, Y. Lu, X. Zhang, H. Li, G. Wang, W. Qiu, G. Yuan, J. Ma, W. Li, Z. Li, X. Wang, J. Zeng, Z. Yang, J. Liu, Y. Liang, S. Lu, H. Zhang, H. Liu, P. Liu, K. Fan, X. Jiang, Y. Li, Q. Su, T. Ning, H. Tan, Z. An, Z. Jiang, L. Liu, Z. Zhou, Q. Zhang, X. Li, Z. Shan, Y. Xue, H. Mao, L. Shi, S. Ye, X. Zhang, J. Sun, P. Li, T. Yang, F. Li, J. Lin, Z. Zhang, Y. Zhao, R. Li, X. Guo, Q. Yao, W. Lu, S. Qu, H. Li, L. Tan, W. Wang, Y. Yao, D. Chen, Y. Li, J. Gao, W. Hu, X. Fei, T. Wu, S. Dong, W. Jin, C. Li, D. Zhao, B. Feng, Y. Zhao, Y. Zhang, X. Li, L. Chen. "Dorzagliatin add-on therapy to metformin in patients with type 2 diabetes: a randomized, double-blind, placebo-controlled phase 3 trial." *Nat. Med.* **2022**, *28*, 5, 974–981.
- [3] U. Kaur, B. K. Pathak, T. J. Meerashahib, D. V. V. Krishna, S. S. Chakrabarti. "Should glucokinase be given a chance in diabetes therapeutics? a clinical-pharmacological review of dorzagliatin and lessons learned so far." *Clin. Drug. Invest.* **2024**, *44*, 4, 223–250.
- [4] D. Zhu, X. Li, J. Ma, J. Zeng, S. Gan, X. Dong, J. Yang, X. Lin, H. Cai, W. Song, X. Li, K. Zhang, Q. Zhang, Y. Lu, R. Bu, H. Shao, G. Wang, G. Yuan, X. Ran, L. Liao, W. Zhao, P. Li, L. Sun, L. Shi, Z. Jiang, Y. Xue, H. Jiang, Q. Li, Z. Li, M. Fu, Z. Liang, L. Guo, M. Liu, C. Xu, W. Li, X. Yu, G. Qin, Z. Yang, B. Su, L. Zeng, H. Geng, Y. Shi, Y. Zhao, Y. Zhang, W. Yang, L. Chen. "Dorzagliatin in drug-naïve patients with type 2 diabetes: a randomized, double-blind, placebo-controlled phase 3 trial." *Nat. Med.* **2022**, *28*, 5, 965–973.
- [5] C. Li, J. Zhang, Y. Sun, Y. Zhao, X. Liu, Z. Fang, L. Feng, B. He, Q. Zou, G. J. Tracey. "A phase I open-label clinical trial to study drug-drug interactions of dorzagliatin and sitagliptin in patients with type 2 diabetes and obesity." *Nat. Commun.* **2023**, *14*, 1, 1405.
- [6] A. L. Lynam, J. M. Dennis, K. R. Owen, R. A. Oram, A. G. Jones, B. M. Shields, L. A. Ferrat. "Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults." *D.A.P.R.* **2020**, *4*, 6.
- [7] G. Shi, G. Liu, Q. Gao, S. Zhang, Q. Wang, L. Wu, P. He, Q. Yu. "Comparison between traditional logistic regression and machine learning for predicting mortality in adult sepsis patients." *Front. Med.* **2024**, *11*, 1496869.
- [8] G. Shi, G. Liu, Q. Gao, S. Zhang, Q. Wang, L. Wu, P. He, Q. Yu. "A random forest algorithm-based prediction model for moderate to severe acute postoperative pain after orthopedic surgery under general anesthesia." *BMC anesthesiology*. **2023**, *23*, 1, 361.
- [9] Y. Ma, Q. Lu, F. Yuan, H. Chen. "Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after pcp for osteoporotic vertebral compression fractures." *J. Orthop. Surg. Res.* **2023**, *18*,

1, 62.

- [10] Y. Jin, A. Lan, Y. Dai, L. Jiang, S. Liu. "Development and testing of a random forest-based machine learning model for predicting events among breast cancer patients with a poor response to neoadjuvant chemotherapy." *Eur. J. Med. Res.* **2023**, *28*, 1, 394.
- [11] C. Z. Wu, L. Y. Huang, F. Y. Chen, C. H. Kuo, D. F. Yeih. "Using machine learning to predict abnormal carotid intima-media thickness in type 2 diabetes." *Diagnostics (Basel, Switzerland)*. **2023**, *13*, 11, 1834.
- [12] J. Tan, E. Huang, Y. Hao, H. Wan, Q. Zhang. "Risk factors associated with severe progression of parkinson's disease: random forest and logistic regression models." *Front. Neurol.* **2025**, *16*, 1550789.
- [13] V. B. Berikov, O. A. Kutnenko, J. F. Semenova, V. V. Klimontov. "Machine learning models for nocturnal hypoglycemia prediction in hospitalized patients with type 1 diabetes." *J. Pers. Med.* **2022**, *12*, 8, 1262.
- [14] J. Huang, W. Liu. "Comparison of machine learning models for predicting stroke risk in hypertensive patients: lasso regression model, random forest model, boruta algorithm model, and boruta algorithm combined with lasso regression model." *Medicine*. **2025**, *104*, 22, e42690.
- [15] Y. Ren, L. Li, L. Wan, Y. Huang, S. Cao. "Glucokinase as an emerging anti-diabetes target and recent progress in the development of its agonists." *J. Enzyme. Inhib. Med. Chem.* **2022**, *37*, 1, 606–615.
- [16] W. Jia, J. CN. Chan, T. Y. Wong, E. B. Fisher. "Diabetes in china: epidemiology, pathophysiology and multi-omics." *Nat. Metab.* **2025**, *7*, 1, 16–34.
- [17] J. Fetzner, E. Rafi. "Glycemic, cardiorenal, and weight implications on noninsulin pharmacotherapy for the management of type 2 diabetes." *J. Clin. Endocr. Metab.* **2025**, *110*, 1496869.
- [18] N. Marx, M. Husain, M. Lehrke, S. Verma, N. Sattar. "GLP-1 receptor agonists for the reduction of atherosclerotic cardiovascular risk in patients with type 2 diabetes." *Circulation*. **2022**, *146*, 24, 1882–1894.
- [19] J. B. McGill, I. B. Hirsch, C. G. Parkin, G. Aleppo, C. J. Levy, J. R. Gavin. 3rd. "The current and future role of insulin therapy in the management of type 2 diabetes: a narrative review." *Diabetes. Ther.* **2024**, *15*, 5, 1085–1098.
- [20] M. P. Klein, H. K. Kurturk, J. K. Snell-Bergeon, V. N. Shah. "Reduced efficacy of glucagon-like peptide-1 receptor agonists therapy in people with type 1 diabetes and genetic forms of obesity." *J. Diabetes. Sci. Techn.* **2025**, *19*, 2, 297–303.
- [21] A. Kyriakidou, A. V. Kyriazou, T. Koufakis, Y. Vasilopoulos, M. Grammatiki, X. T. Sekmekidou, L. Avramidis, S. Baltagiannis, D. G. Goulis, P. Zebekakis, K. Kotsa. "Clinical and genetic predictors of glycemic control and weight loss response to liraglutide in patients with type 2 diabetes." *J. Pers. Med.* **2022**, *12*, 3, 424.