A Hybrid Neural Collaborative Filtering and Word Embedding Approach for Personalised Diabetes Drug Recommendation

Original Scientific Paper

Chen Xiaoying
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, Malaysia
p139588@siswa.ukm.edu.my

Shahrul Azman Mohd Noah*
Centre for Artificial Intelligence Technology,
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia
shahrul@ukm.edu.my

Li Huiting
Faculty of Foreign Languages, Guangzhou University of Chinese Medicine,
Guangzhou, China

*Corresponding author

Abstract *– Personalised diabetes drug recommender systems have the potential to enhance treatment efficacy and advance precision medicine. However, their real-world implementation remains challenging due to issues such as data sparsity and the cold-start problem, which affect both new patients and newly introduced drugs. In this study, we propose a hybrid recommendation framework that integrates Neural Collaborative Filtering (NCF) with biomedical word-embedding techniques to capture complex patient–drug interactions while alleviating cold-start limitations. The NCF component learns nonlinear relationships from structured clinical variables and historical medication records, whereas the word-embedding component leverages BioBERT to extract semantic representations of drugs from textual descriptions, enabling effective retrieval of appropriate treatments even when interaction data is unavailable. Our approach was evaluated using the UCI Diabetes 130-US Hospitals dataset and a curated DrugBank corpus comprising 71 anti-diabetic drugs. The optimised NCF model achieved strong predictive performance, with an accuracy of 0.9004, precision of 0.7658, recall of 0.4209, F1-score of 0.5433, specificity of 0.9789, and AUC of 0.8717 on held-out data. Furthermore, the BioBERT-based semantic module generated clinically coherent drug similarity rankings, suitable for recommending alternatives in cold-start scenarios. By fusing patient-centric probability estimates with semantic similarity scores, the proposed hybrid strategy delivers ranked drug recommendations, supporting personalised and data-efficient diabetes treatment.*

Keywords: *recommender system, neural collaborative filtering, word embedding, BioBERT, cold-start problem, diabetes management*

Received: Received: June 26, 2025; Received in revised form: October 31, 2025; Accepted: November 3, 2025

1. INTRODUCTION

Diabetes mellitus encompasses a set of chronic metabolic disorders characterised by sustained hyperglycemia due to defects in insulin secretion or action. Common types include type 1, type 2, and gestational diabetes. Diabetes affects multiple organs and systems and is associated with increased mortality. Globally, diabetes was among the top ten causes of death, accounting for 4.2 million deaths in 2019 [1]. The disease is therefore a significant and growing public health concern.

Managing diabetes is a complex and highly personal process. The treatment plan depends on factors such as a patient's age, existing health conditions, kidney function, blood sugar levels, and their response to previous therapies. With numerous available drug options, doctors often employ combination therapies to help patients achieve their glucose goals. This diversity makes personalisation essential, and precision medicine holds great promise for improving diabetes care [2].

Recommender systems are algorithmic frameworks that predict items likely to be relevant to users by lever-

aging historical interactions, preferences, and item features [3]. These systems are widely applied in domains such as e-commerce and social media [4, 5], and they have recently been adapted for healthcare tasks, including disease prediction, diagnosis support, and treatment recommendation [6]. Traditional collaborative filtering (CF) methods have the potential to improve treatment selection because they identify patterns in shared patient behaviour. However, CF often struggles when data are limited or when new patients or drugs enter the system; a challenge known as the cold-start problem.

In this paper, we focus on two types of cold-start issues:

a. New-patient cold start, where there's no prior medication history.

b. New-drug cold start, where a new or rarely used medication lacks enough prescription data.

While some existing methods attempt to overcome these challenges by utilising drug information or clustering patients, few effectively handle both simultaneously in a clinically meaningful manner [7-8].

To solve this, we propose a hybrid model that combines Neural Collaborative Filtering (NCF) with BioBERT-based semantic retrieval. The NCF module learns patterns from structured clinical data to generate personalised recommendations for patients with prior records. Meanwhile, BioBERT transforms drug descriptions into rich semantic vectors, helping the system recognise and suggest new but relevant medications. Together, they produce ranked drug recommendations that balance personalisation with pharmacological reasoning.

The main contributions of this paper are as follows:

- We propose a hybrid recommendation framework that integrates patient-level Neural Collaborative Filtering (NCF) with BioBERT-based drug semantic retrieval to address both patient and drug cold-start challenges effectively.
- We conduct a comprehensive evaluation, including a quantitative assessment of the NCF model's predictive performance and a qualitative analysis of BioBERT-based drug retrievals against authoritative clinical guidelines.

2. LITERATURE REVIEW

Recommender systems have increasingly been explored to support diabetes management by synthesising patient records, prescription histories, and drug metadata to generate tailored pharmacological and lifestyle suggestions. Early efforts relied primarily on CF techniques that exploit patterns of co-occurrence between patients and drugs. Such approaches are straightforward and effective when sufficient interaction data exist. However, they degrade markedly under data sparsity and in cold-start scenarios. To mitigate sparsity, researchers have proposed hybrid and preprocessing methods that enrich interaction signals. For example, clustering combined with cosine-similarity CF can group similar patients or encounters to increase effective density, thereby improving recommendations for frequently occurring patterns while remaining sensitive to the granularity of clustering and input quality [7]. Explainable hybrid CF models aim to make recommendations more transparent by combining memory-based and model-based components with probabilistic or Bayesian overlays. These enhancements improve interpretability but typically require relatively dense interaction matrices and comprehensive user-drug records [8].

Building on the idea that semantics can complement co-occurrence signals, Wang and Li [9] introduced a CF algorithm that fuses user behaviour with drug semantic similarity derived from Word2Vec embeddings, demonstrating improved recommendation accuracy when drug semantics are informative. Complementary work has extended hybrid designs to incorporate behavioural and genetic data for lifestyle or non-pharmacological recommendations, showing promise but also revealing scalability constraints when specialised data are required [10].

Beyond static and hybrid CF methods, approaches that explicitly model sequential or decision-making processes, such as reinforcement learning (RL), have emerged for adaptive treatment planning. RL models, including Deep Q-Networks trained with clinical guidance, can learn dynamic policies that adapt to evolving patient states and have shown encouraging results for glycemic control in retrospective simulations. However, their dependence on richly labelled longitudinal data and the difficulty of validating learned policies in novel clinical scenarios remain essential limitations [11]. These methodological strands highlight a central tension for diabetes recommender systems. Models that rely on interaction histories can achieve strong personalisation when data are ample. In contrast, methods that exploit content or semantics can recommend plausible options for rare or new items but may lack patient-specific nuance.

Neural Collaborative Filtering (NCF) represents an evolution of CF paradigms by leveraging embedding layers and deep architectures to capture high-order, nonlinear patient-drug interactions. NCF architectures enable the integration of side information (clinical features, temporal signals, item attributes), allowing for richer representations than linear matrix factorisation while preserving the core of item-user interactions [12]. In healthcare, NCF and its variants have been adapted for use in treatment recommendations, safety-aware combination selection, and temporally sensitive personalisation. For instance, SafeDrug integrates molecular graph encoders into an NCF pipeline to account for the molecular structures of drugs and their interaction risks [13]. In contrast, dynamic network and multi-view drug representation approaches extend NCF to model patients' health trajectories and comorbidity contexts, thereby improving temporally aware recommendations

[14]. Privacy-preserving variants, such as federated NCF, enable multi-institutional learning without sharing raw data, addressing a practical deployment barrier in cross-site healthcare applications [15]. Applications of NCF beyond medication, such as personalised rehabilitation and skincare, illustrate its adaptability to diverse clinical and consumer health tasks when appropriate contextual features are included [16, 17].

Parallel to these interaction-based advances, text-based embedding methods have matured into powerful tools for mitigating item cold-start. Static embeddings (Word2Vec) and contextual transformers (BERT and its biomedical adaptations, such as BioBERT) produce dense semantic vectors from drug labels, mechanisms, and indications, enabling similarity-based retrieval even when prescription histories are absent.

Several lines of work have demonstrated the utility of such embeddings for cold-start and metadata-sparse settings. Incremental or meta-embedding schemes can learn useful item representations from limited interaction data [18, 19], while content-link networks and BERT-enhanced memory recommenders have improved cold-start resilience and contextual relevance in medical domains [20, 21]. Empirically, domain-adapted transformers (e.g., BioBERT) tend to outperform general-domain embeddings on biomedical similarity tasks because they capture domain-specific relations and terminology nuances, making them particularly suitable for recommending pharmacologically coherent alternatives.

The literature suggests complementary strengths and weaknesses across three families of approaches: interaction-centric CF/NCF methods excel at patient-level personalisation when sufficient data exist but struggle with new items [22]; content-centric embedding methods provide semantically plausible options for new drugs but may not fully capture patient heterogeneity; and sequential decision frameworks like RL can optimise dynamic policies but demand longitudinal, labeled data and careful validation. These observations motivate the development of hybrid architectures that integrate patient-level neural collaborative models with biomedical semantic retrieval to address the drug cold-start problem. By combining the expressive power of high-capacity interaction models with the contextual understanding provided by domain-specific text embeddings, such hybrid pipelines aim to retain the personalisation strengths of NCF while expanding coverage to novel or sparsely represented drugs. This design effectively tackles the dual cold-start challenges encountered in diabetes drug recommendation tasks, ensuring both adaptability and clinical relevance.

3. METHODOLOGY

The overall architecture of the proposed hybrid recommender is illustrated in Fig. 1. It consists of three modules: (1) an NCF-based recommender that learns from patient-drug interactions and clinical features, (2) a word-embedding-based semantic recommender that encodes drug descriptions for cold-start drugs, and (3) a fusion module that consolidates NCF predictions with embedding-based alternatives to produce the final ranked suggestions.

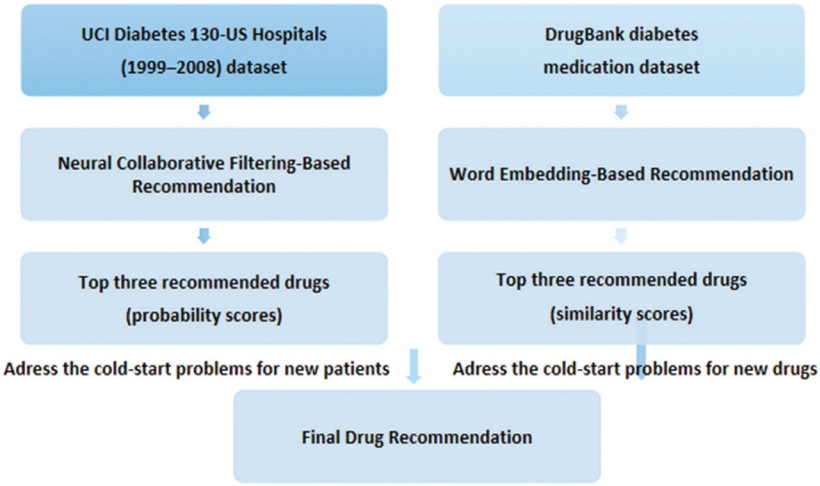

Fig. 1. Hybrid drug recommendation framework

3.1. NEURAL COLLABORATIVE FILTERING-BASED RECOMMENDATION

a) Data acquisition and preprocessing

We utilised the UCI Diabetes 130-US Hospitals dataset (1999-2008), which comprises 101,766 encounter records from 71,518 unique patients and over 50 attributes.

We selected key patient features and the 11 most commonly used diabetes drugs. Data preprocessing included deduplication, numerical encoding of lab tests and ICD-9 codes, normalisation of drug use, discretisation of continuous features, and binary encoding of categorical variables(details shown in Table 10). These steps ensured that the input for the NCF module was structured and clean.

Table 1. Select features and data processing methods

Feature Name	Meaning	Processing Method
patient_nbr	Unique patient identifier	Used directly as a patient ID for the interaction matrix
A1Cresult	Glycosylated haemoglobin result	Encoded: None=0, Norm=1, >7=2, >8=3
max_glu_serum	Maximum blood glucose	Encoded: None=0, Norm=1, >200=2, >300=3
diag_1, diag_2, diag_3	ICD-9 diagnostic codes	Discretized: low (1-3)=0, medium (4-6)=1, high (>6)=2
number_diagnoses	Total diagnoses	Mapped to disease categories and numerically encoded
Age	Age interval	Converted to a median numeric value and standardised
Change	Whether medication changed	Binary: No=0, Ch=1
diabetesMed	Whether the diabetes medication used	Binary: No=0, Yes=1
num_medications	Total Medications	Discretized: low (1-5)=0, medium (6-15)=1, high (>15)=2
time_in_hospital	Days in hospital	Discretized: short (1-4)=0, medium (5-8)=1, long (>8)=2
Use the status of the 11 most commonly used drugs	Drug-specific use status	Binary: No=0, Steady/Up/Down=1

b) Patient-drug interaction matrix construction

After preprocessing, the patient-drug interaction matrix contained 56,103 patients and 11 drugs.

The matrix was sparse: only 14.07% of all potential patient-drug pairs were positive, indicating a strong class imbalance.

c) Model design and training

The NCF model embeds patient and drug IDs into low-dimensional vectors, concatenates them with clinical features, and passes them through multi-layer perceptrons (MLP) with ReLU activations and dropout regularisation. A sigmoid output predicts the probability of drug use (Fig. 2).

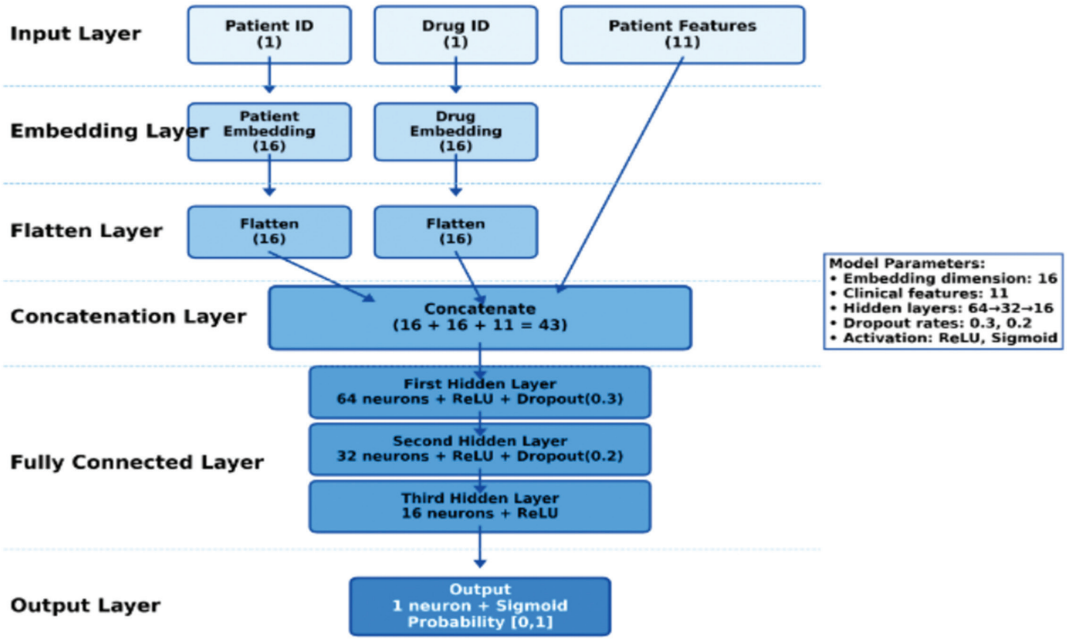

Fig. 2. Structure of the NCF model

Training used an 80/20 train-test split and 3-fold cross-validation. The binary cross-entropy loss and Adam optimiser were employed, with a batch size of 128 and early stopping to prevent overfitting. To further improve model performance, three optimisation strategies were explored as shown in Table 2.

Table 2. Model optimisation methods

Optimisation Aspect	Baseline Model	Class weight optimisation Model	Learning Rate optimisation Model
Optimizer	Adam	Adam	Adam
Initial Learning Rate	0.001	0.001	0.0005
Learning Rate Adjustment	None	None	Dynamic decay
Class Weights	Not used	Used	Not used
Overfitting Control	Early stop	Early stop + weights	Early stop + learning rate decay
Sample Imbalance Treatment	None	Class weight	None
Main Optimisation Objective	Baseline	Imbalance solution	Training stability

- Baseline: no class weighting, fixed learning rate.
- Class weight: apply class weights to the loss to penalise misclassification of the minority class.
- Learning rate optimisation: use a dynamic learning rate decay schedule to improve stability and generalisation.

d) Model evaluation

Model performance was evaluated using multiple metrics, including accuracy, precision, recall, and F1 score.

3.2. WORD EMBEDDING-BASED RECOMMENDATION

a) Data and preprocessing

Descriptions of 71 anti-diabetic drugs were retrieved from DrugBank[1], encompassing major pharmacological classes such as insulin analogues, sulfonylureas, biguanides, α-glucosidase inhibitors, GLP-1 receptor agonists, DPP-4 inhibitors, and SGLT-2 inhibitors. Eleven drugs used in the NCF module served as baseline items, while the remaining drugs were treated as unseen new drugs. Descriptions were preprocessed through lowercasing, punctuation removal, tokenisation, stop-word elimination, and lemmatisation.

b) Semantic representation

Three embedding models with progressively higher levels of linguistic and biomedical sophistication were employed: Word2Vec, BERT, and BioBERT. Each model encoded the preprocessed drug-related text into dense semantic vectors, capturing latent pharmacological semantics. These embeddings capture latent pharmacological semantics, enabling computation of inter-drug similarity without prior interaction data.

c) Similarity computation and evaluation

The cosine similarity was used to calculate the similarity between the baseline drugs and all the new drugs. Based on the similarity scores, the top three drugs with the highest similarity were selected as the recommended results. For each baseline drug d_i, the cosine similarity between its embedding v_i and those of all other drugs v_j is as follows:

$$sim(d_i, d_j) = \frac{v_i \cdot v_j}{\| v_i \| \| v_j \|}$$

The top three drugs with the highest similarity were selected as recommendations for d_i.

Clinical interpretability was evaluated using an Extended Semantic Coherence Score (SCS^+) that considers both identical and mechanistically related pharmacological classes. For each baseline drug d_i,

$$SCS^+ = \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j\in TopK(i)} R\left(C_i, C_j\right)}{K}$$

where $R(C_i, C_j)$=1 if classes are identical, 0.5 if therapeutically or mechanistically related, and 0 otherwise.

Higher SCS^+ values indicate better alignment between embedding similarity and therapeutic mechanism.

3.3. HYBRID RECOMMENDATION STRATEGY

This module integrates outputs from the NCF and word embedding models to generate a ranked list of drug recommendations. For existing or new patients, the NCF model predicts the probability of each drug by inputting and preprocessing patient IDs or clinical data, then selects the top three highest-probability drugs as recommendations. For each drug recommended by the NCF model, the three most similar drugs are generated using a word embedding model. The primary drug and its similar alternatives are integrated into a recommendation list, providing both primary and extended drug options for each patient.

4. RESULTS AND DISCUSSION

4.1. NEURAL COLLABORATIVE FILTERING-BASED RECOMMENDATION

The UCI Diabetes 130-US Hospitals dataset provides a realistic representation of clinical heterogeneity among diabetic patients. After data filtering, the final dataset comprises 56,103 patients and 11 target drugs. The age distribution indicates that most patients fall within the 60–80 age range (Fig. 3).

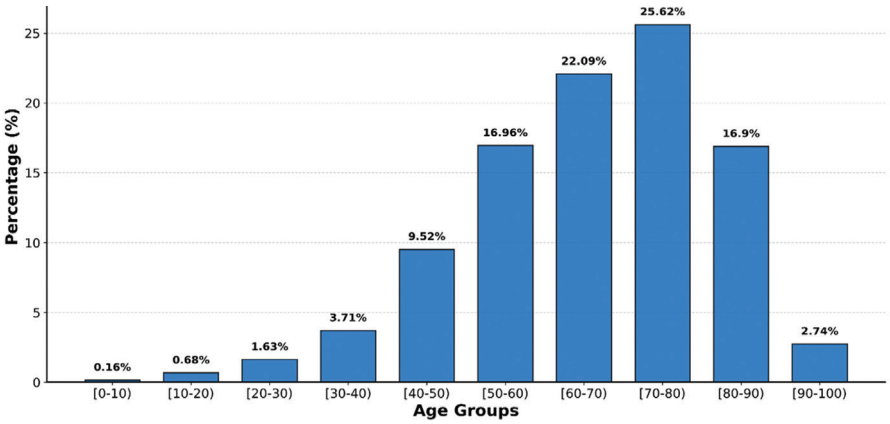

Fig. 3. Age distribution of patients

[1] https://go.drugbank.com/categories/DBCAT002117

Polypharmacy is highly prevalent, with 96.18% of patients prescribed more than five medications. Similarly, multimorbidity is common, as 66.64% of patients present with more than six diagnoses, with cardiovascular disease being a frequent comorbidity (~30%) (Figs. 4–5). Drug usage patterns are highly imbalanced: insulin is the most frequently prescribed medication (67.44%), followed by metformin (27.40%), glipizide (16.63%), and glyburide (14.35%) (Fig. 6).

The resulting patient–drug interaction matrix contains 617,133 potential interactions, of which only 14.07% are positive, highlighting the inherent sparsity of the dataset.

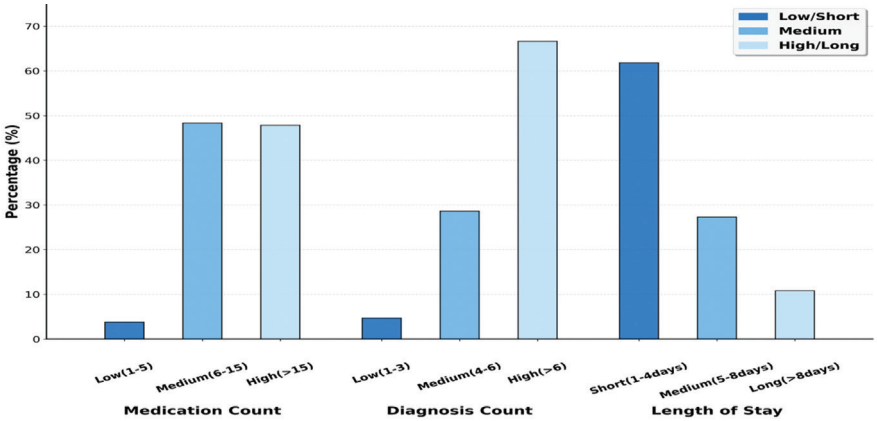

Fig. 4. Patient clinical characteristics distribution

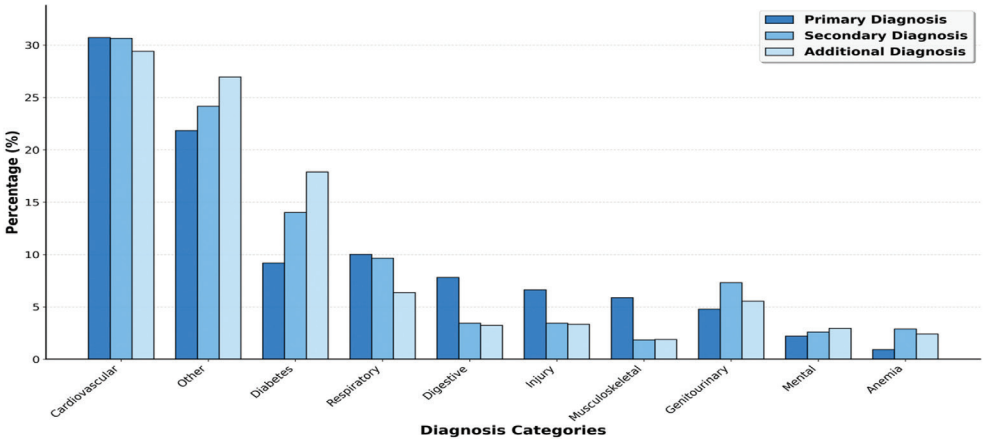

Fig. 5. Diagnosis categories distribution

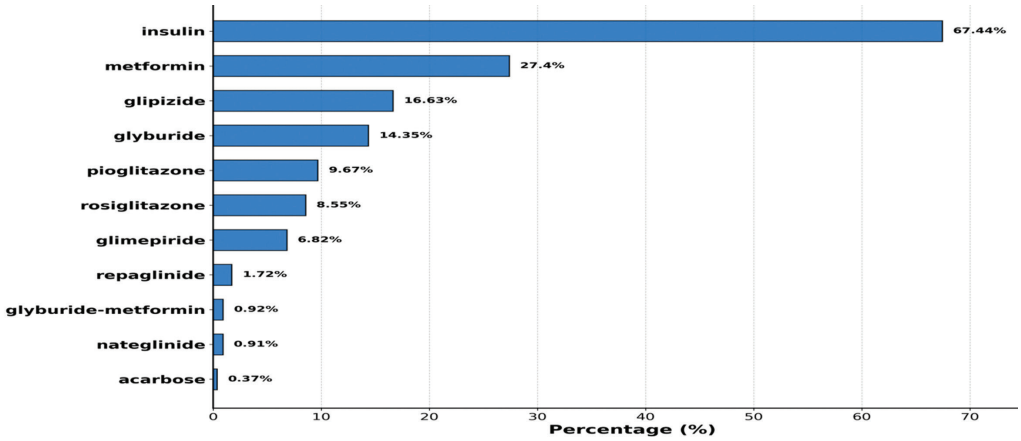

Fig. 6. Medication Usage Distribution

We evaluated three training strategies (baseline, class-weighted, and learning-rate-optimised) under identical data splits and evaluation protocols. As illustrated in Fig. 7, the learning-rate-optimised configuration achieved the best or performance on all key metrics for the training set, except the recall metric. At the same time, the baseline model ranked second, and the class-weighted model traded precision for recall (high

recall but substantially lower precision and overall accuracy). On the held-out test set, the learning-rate-optimised model achieved strong discrimination and clinically relevant performance: accuracy = 0.900, precision = 0.763, recall = 0.421, F1 score = 0.543, specificity = 0.979, and AUC = 0.873 (Table 3).

Considering the balance between precision and recall, the stability of validation performance, and the modest computational overhead of the scheduling scheme, we selected the learning-rate-optimised model as our final configuration for downstream analyses.

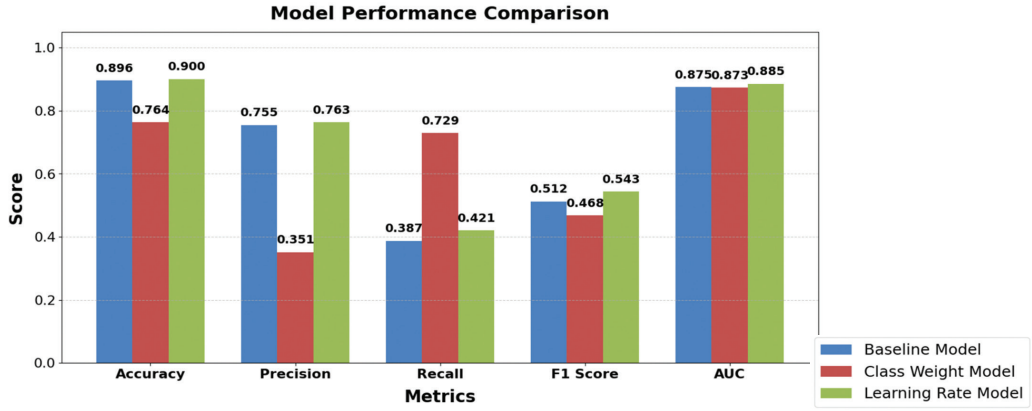

Fig. 7. Model Performance Comparison

Table 3. Test-set performance (learning-rate-optimised NCF)

Evaluation Metric	Value
Accuracy	0.900
Precision	0.766
Recall	0.421
F1 Score	0.543
Specificity	0.979
AUC	0.873

4.2. WORD EMBEDDING-BASED RECOMMENDATION

The three embedding models were evaluated across two complementary dimensions: semantic precision, measured by the mean Top-3 cosine similarity, and clinical interpretability, quantified using the Extended Semantic Coherence Score (SCS^{+}). As shown in Fig. 8, BioBERT achieved the best overall balance between linguistic and pharmacological understanding. Its embeddings attained the best Mean Cosine Similarity (0.936) and SCS^{+}(0.76), indicating the highest internal semantic cohesion and good clinical interpretability.

BERT achieved sufficient semantic precision, 0.911, while it showed low domain comprehension; the pre-training corpus for BERT lacks biomedical specificity. Word2Vec was able to capture basic lexical proximity at 0.842, which, due to its shallow co-occurrence semantics, resulted in recommendations lacking pharmacological coherence.

This, in turn, reflects the clear performance hierarchy of the tested models in terms of their ability to encode domain-specific biomedical knowledge into dense vector representations. BioBERT provided the most clinically coherent recommendations. For example, it associated metformin with liraglutide, dulaglutide, and insulin degludec, consistent with ADA-endorsed treatment intensification pathways that involve GLP-1 receptor agonists and insulin analogues. Word2Vec, on the other hand, demonstrated a tendency to generate lexically similar but pharmacologically irrelevant associations. For example, linking metformin with insulin detemir. This underlines how static word embeddings fail to capture the mechanistic and therapeutic context within medical texts. These results confirm that domain-specific contextual embeddings enhance both semantic precision and therapeutic interpretability, thereby providing robust support for explainable cold-start recommendations in diabetes pharmacotherapy.

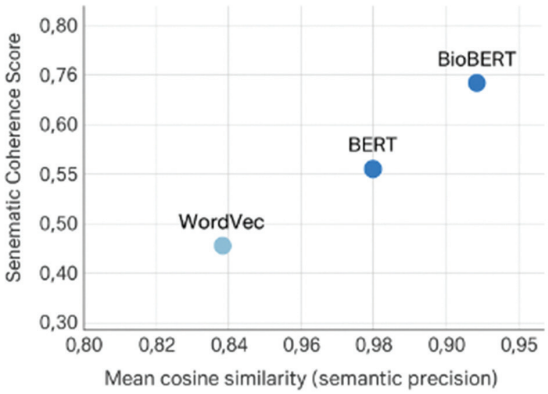

Fig. 8. Comparative performance of word embedding models

4.3. HYBRID RECOMMENDATION OUTCOMES

The hybrid recommender system developed shows impressive technical strength and real-world medical relevance. By combining NCF with BioBERT-based semantic embeddings, it effectively tackles the common cold-start challenges, where systems struggle with new patients or newly approved drugs.

The NCF module captures complex relationships between patients' health profiles and their past prescriptions, allowing for highly personalised suggestions. Meanwhile, BioBERT interprets biomedical language to include new or rarely prescribed medications in the recommendations. Together, they create a balanced, clinically consistent system that aligns with key guidelines, including the ADA Standards of Care in Diabetes 2024 [23] and the KDIGO 2023 Clinical Practice Guideline[24].

During testing, the hybrid model delivered accurate, guideline-based recommendations. For returning patients, NCF's top-ranked drug suggestions closely matched the actual prescriptions, outperforming traditional CF technique. For new patients, the system adapted well, suggesting drug combinations tailored to factors such as blood sugar levels, existing conditions, and treatment complexity.

For instance, in one case, as shown in Table 4 (Patient 90341199, age 75, with multiple conditions and a stable A1C), NCF correctly prioritised insulin and metformin, both of which were found in the real prescription record. BioBERT added further depth by suggesting related alternatives such as insulin degludec and liraglutide, consistent with ADA recommendations for patients at higher cardiovascular risk.

Table 4 also shows that, in another case involving a new patient with poor glycemic control (A1C>7), the system recommended insulin, metformin, and glipizide as top options. BioBERT's expansion identified GLP-1 receptor agonists and SGLT2 inhibitors, such as liraglutide, dulaglutide, and canagliflozin, as clinically sound alternatives that support heart and kidney health.

Overall, the model achieved strong clinical interpretability. NCF offered patient-specific probability scores, while BioBERT added context through semantic ranking. This hybrid design not only delivers accurate and personalised results but also ensures that every recommendation is grounded in sound medical reasoning.

Table 4. Hybrid Recommendation Outcomes

Patient ID	Patient information	TOP 3 NCF-recommended drugs	Hybrid recommendations (NCF + BioBERT)
Existing Patient (ID: 90341199)	• Age Value: 75 • Number of Diagnoses: High (>6) • Time in Hospital: Short (1-4) • Number of Medications: Low (1-5) • Diabetes Medication: Yes • A1C Result: Norm • Glucose Level: None • Primary Diagnosis: other • Secondary Diagnosis: other • Additional Diagnosis: diabetes	1. insulin - Probability: 0.6353 2. metformin - Probability: 0.2859 3. glipizide - Probability: 0.2284 Actually used drugs: metformin, insulin	1. NCF recommended drug: insulin Similar drugs: • Canagliflozin (Similarity: 0.8678) • Insulin degludec (Similarity: 0.8415) • Insulin detemir (Similarity: 0.8353) 2. NCF recommended drug: metformin Similar drugs: • Liraglutide (Similarity: 0.9375) • Insulin degludec (Similarity: 0.9316) • Dulaglutide (Similarity: 0.9258)
New Patient	• Age Value: 50 • Number of Diagnoses: High (>6) • Time in Hospital: Medium (5-8) • Number of Medications: High (>15) • Diabetes Medication: Yes • A1C Result: >7 • Glucose Level: >200 • Primary Diagnosis: diabetes • Secondary Diagnosis: respiratory • Additional Diagnosis: cardiovascular	1. insulin - Probability: 0.8902 2. metformin - Probability: 0.2354 3. glipizide - Probability: 0.1620	3. NCF recommended drug: glipizide Similar drugs: • Bexagliflozin (Similarity: 0.9332) • Lixisenatide (Similarity: 0.9307) • Albiglutide (Similarity: 0.9298)

4.4. COMPARISON WITH REPRESENTATIVE PRIOR METHODS

Two representative studies, Granda Morales *et* al. [6] and Valdiviezo-Díaz [7], used the same UCI Diabetes 130-US Hospitals dataset as this research, providing a consistent baseline for evaluation. Table 5 summarises their methodologies, key performance results, and the extent to which they address the cold-start problem.

Both prior studies relied primarily on CF. Granda Morales *et* al. [6] applied unsupervised clustering to reduce sparsity before CF, achieving moderate accuracy but limited personalisation due to fixed patient grouping. Valdiviezo-Díaz [7] introduced a Bayesian hybrid CF to enhance explainability, improving precision but remaining dependent on historical interactions. Neither approach explicitly addressed the cold-start problem, particularly for newly introduced drugs. In contrast, our method integrates NCF to model nonlinear patient-drug interactions and BioBERT-based embeddings to infer semantic similarity among drugs. This dual-module design enhances prediction accuracy and facilitates recommendations for previously unseen drugs and patients.

On the same dataset, our model outperforms previous methods while providing alternatives that are consistent with pharmacological logic. The improvement demonstrates the effectiveness of combining deep interaction modelling with biomedical text semantics for precision drug recommendation.

Table 5. Comparative summary of previous and proposed methods

Study	Dataset	Method	Key metrics	Cold-start addressed
Granda Morales *et* al.	UCI Diabetes 130-US	PCA and K-means clustering combined with user-based collaborative filtering	Accuracy = 0.61; MSE = 0.51	No
Valdiviezo-Díaz	UCI Diabetes 130-US	Naive Bayes Collaborative Filtering	Precision = 0.744; Recall = 0.616; MAE = 0.332	No
This work	UCI Diabetes 130-US and DrugBank	Hybrid Neural Collaborative Filtering with BioBERT	Accuracy = 0.9004, Precision = 0.7658, AUC = 0.8717; Recall = 0.4209	Yes, for both new patients and new drugs

5. CONCLUSION

This study proposes a hybrid drug recommendation framework that combines NCF with BioBERT-based semantic embeddings to enhance personalisation and address data sparsity and cold-start issues in diabetes treatment. Using the UCI Diabetes 130-US Hospitals dataset, the model achieves notably better accuracy and robustness than previous CF and clustering approaches. By jointly modelling patient-drug interactions and semantic drug relationships, the method effectively handles both new-patient and new-drug scenarios, enhancing generalizability and clinical relevance. Despite these advantages, this study has several limitations. The dataset used is dated and imbalanced, which may affect the generalisability of the results. The NCF model also suffers from limited interpretability, and the embedding performance is influenced by the quality and completeness of textual data, potentially reducing accuracy for rare or sparsely described drugs.

Future work will aim to address these limitations by employing more recent and diverse clinical datasets, integrating temporal and safety-related information, and enhancing model interpretability to facilitate the practical and trustworthy deployment of clinical applications.

6. REFERENCES

[1] D. Tomic, J. E. Shaw, D. J. Magliano, "The burden and risks of emerging complications of diabetes mellitus", Nature Reviews Endocrinology, Vol. 18, No. 9, 2022, pp. 525-539.

[2] C. M. Mutter, T. Smith, O. Menze, M. Zakharia, H. Nguyen, "Diabetes: Pathogenesis, diagnosis, and clinical management", Cureus, Vol. 13, No. 2, 2021, p. e13523.

[3] N. Tiwary, S. A. Mohd Noah, F. Fauzi, T. S. Yee, "A review of explainable recommender systems utilising knowledge graphs and reinforcement learning", IEEE Access, Vol. 12, 2024, pp. 91999-92019.

[4] N. I. Y. Saat, S. A. M. Noah, M. Mohd, "Enhanced content-based recommendation using topic modelling and knowledge graph", Elektronika Ir Elektrotechnika, Vol. 30, No. 2, 2024, pp. 73-79.

[5] N. Jamil, S. A. Mohd Noah, M. Mohd, "Collaborative item recommendations based on friendship strength in social networks", International Journal of Machine Learning and Computing, Vol. 10, No. 3, 2020, pp. 437-443.

[6] N. A. Ahmad, A. F. Mat Ludin, D. Vanoh, *et* al., "Mobile health application (WeFit) among community-dwelling older adults: Development, validation, acceptance, and usability study", Digital Health, Vol. 10, 2024.

[7] L. F. Granda Morales, P. Valdiviezo-Diaz, R. Reátegui, L. Barba-Guaman, "Drug recommender system for diabetes using a collaborative filtering and clustering approach: Development and performance evaluation", Journal of Medical Internet Research, Vol. 24, No. 7, 2022, p. e37233.

[8] P. Valdiviezo-Diaz, "Explainable recommendations of drugs for diabetic patients", Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies - Healthinf, Lisbon, Portugal, 2023, pp. 73-80.

[9] R. Wang, T. Li, "Collaborative filtering recommendation algorithm based on user behavior and drug semantics", Proceedings of the Journal of Physics: Conference Series, Vol. 1802, No. 3, 2021, p. 032005.

[10] F. R. Cavallo, C. Toumazou, "Personalised lifestyle recommendations for type 2 diabetes: Design and simulation of a recommender system on UK Biobank data", PLOS Digital Health, Vol. 2, No. 8, 2023.

[11] X. Sun, Y. Bee, S. Lam, Z. Liu, W. Zhao, S. Chia, H. Abdul Kadir, J. Wu, B. Ang, N. Liu, Z. Lei, Z. Xu, T. Zhao, G. Hu, G. Xie, "Effective treatment recommendations for type 2 diabetes management using rein-

forcement learning: Treatment recommendation model development and validation", Journal of Medical Internet Research, Vol. 23, No. 7, 2021, p. e27858.

[12] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, "Neural collaborative filtering", Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3-7 April 2017, pp. 173-182.

[13] C. Yang, C. Xiao, F. Ma, L. Glass, J. Sun, "SafeDrug: Dual molecular graph encoders for recommending effective and safe drug combinations", arXiv:2105.02711, 2021.

[14] G. Liu, X. Yu, Z. Liu, X. Li, X. Fan, X. Zheng, "DNMDR: Dynamic networks and multi-view drug representations for safe medication recommendation", arXiv:2501.08572, 2025.

[15] V. Perifanis, P. S. Efraimidis, "Federated neural collaborative filtering", arXiv:2106.04405, 2021.

[16] S. Chang, J. Kim, Y. Lee, "Personalised rehabilitation training using neural collaborative filtering in musculoskeletal health", Proceedings of the Clinical AI Rehabilitation Conference, Seoul, South Korea, 2023.

[17] T. A. Manurung, Z. K. A. Baizal, "The usage of neural collaborative filtering in enhancing personalised skincare recommendation", Proceedings of the International Conference on Data Science and Its Applications, Jakarta, Indonesia, 2024.

[18] A. Nguyen, L. Dinh, H. Nguyen, "Improving cold-start recommendations with item content and Word2Vec embeddings", Proceedings of the 12th ACM Conference on Recommender Systems, Vancouver, BC, Canada, 2-7 October 2018, pp. 136-144.

[19] J. Misztal-Radecka, M. Indyk, "Meta-User2Vec: Learning user and item embeddings jointly for cold-start recommendation", Information Processing & Management, Vol. 58, No. 3, 2021, p. 102512.

[20] T. Hoang, M. Dang, N. Nguyen, "Context-aware recommendation with BERT for user-item textual interactions", Neurocomputing, Vol. 452, 2021, pp. 672-681.

[21] W. Channarong, P. Wongthongtham, S. P. Muniandy, "HybridBERT4Rec: A hybrid recommender system using BERT for cold-start problems", Journal of Intelligent & Fuzzy Systems, Vol. 43, No. 4, 2022, pp. 5325-5335.

[22] P. Li, S. A. Mohd Noah, H. Mohd Sarim, "A survey on deep neural networks in collaborative filtering recommender systems", arXiv:2412.01378, 2024.

[23] American Diabetes Association, "Standards of Medical Care in Diabetes—2024", Diabetes Care, Vol. 47, Supplement 1, 2024, pp. S1-S350.

[24] Kidney Disease: Improving Global Outcomes (KDIGO) Diabetes Work Group, "KDIGO 2023 Clinical Practice Guideline for Diabetes Management in Chronic Kidney Disease", Kidney International, Vol. 103, No. 6, 2023, pp. 1197-1250.