Optimizing Claims Risk Prediction for Commercial Health Insurance Users by Combining the DeepFM Algorithm

Fei Mi[1], Shuanghong Mi[2]
[1]Department of Health Economics and Medical Insurance, School of Humanities and Management, Guangdong Medical University, Dongguan 523121, Guangdong, China
[2]Department of Finance, College of Digital Economics, Dongguan City University, Dongguan 523121, Guangdong, China
Corresponding author: Fei Mi (e-mail: 15562771399@163.com)

ABSTRACT Commercial health insurance claims prediction is challenged by label delay, temporal distribution drift, and complex high-dimensional feature interactions, which reduce the reliability and stability of conventional risk assessment models. To address these issues, this study proposes a DeepFM-based dual-correction framework that integrates inverse probability weighting, symmetric cross-entropy loss, temporal adversarial learning, and maximum mean discrepancy constraints to jointly mitigate label lag and distribution drift. A shared embedding architecture is employed to capture both low-order and high-order feature interactions, while an adaptive gated fusion strategy balances the contributions of the FM and DNN branches for robust probability estimation. Experimental results demonstrate that the proposed model achieves an AUC of 0.853 under standard conditions and maintains competitive performance under severe label delay and temporal drift scenarios, with limited degradation in calibration metrics. The framework significantly enhances prediction accuracy, robustness, and probability consistency for commercial health insurance claims risk assessment. Furthermore, its capability for modeling dynamic feature interactions and time-varying data distributions provides methodological insights for intelligent signal interpretation and adaptive prediction tasks in electromagnetic wave propagation environments and related antenna data analysis applications.

INDEX TERMS commercial health insurance, deep factorization machine, claims risk prediction, distribution drift, electromagnetic signal processing

I. INTRODUCTION

AGAINST the backdrop of the rapid expansion of the insurance market, accurately predicting a customer's claims risk has become an important support for underwriting, pricing, and risk management for companies [1,2]. Similarly,, accurately forecasting demand and material quality is critical for inventory and pricing strategies. However, in both actual business scenarios, claims data (or demand data) often suffer from problems such as labeling time lags and a large number of unsettled policies (or long-cycle production batches), resulting in inconsistencies between observed labels and actual risk status, thus generating serious noise [3,4]. At the same time, under the influence of factors such as policy adjustments, product iterations, and user behavior evolution, the joint distribution of features and labels experiences systematic drift over time, i.e., temporal distribution drift [5]. These two types of dynamic disturbance factors seriously affect the generalization ability and stability of the prediction model, resulting in a significant reduction in its performance when deployed across time periods. At present, deep learning models have been extensively studied in multiple fields, but the joint modeling problem of label lag and distribution drift is still widely ignored, limiting the credible application of models in real insurance environments. In the field of health insurance claims risk prediction, existing research generally focuses on improving the accuracy and efficiency of predictions through machine learning (ML) methods [6–8]. To improve the accuracy of insurance claims processing, Alam Ashrafe used six ML algorithms to predict health insurance claims risks. The results showed that the eXtreme Gradient Boosting (XGBoost) and Random Forest (RF) models performed better than other algorithms, had the lowest prediction error, and could effectively identify key risk variables [9]. Thakre Vaishnavi P used precision, recall, and F1 score metrics to comprehensively compare the application of logistic regression and other ML models in claims risk prediction. The results showed that the accuracy of logistic regression was 82%, which could provide practical guidance for insurance companies to enhance customer service and reduce financial losses by exploring the trade-off between accuracy and interpretability [10]. Devaguptam Sreegeethi introduced an ML-driven automation

framework that used patient medical history to calculate relevant risk factors and premiums, aiming to reduce human intervention, protect insurance operations, identify high-risk customers, detect fraudulent claims, and reduce financial losses in the insurance industry. Finally, the effectiveness of the proposed framework was verified through insurance processing tasks [11]. To automate claims risk prediction, Nalluri Venkateswarlu compared the applications of Support Vector Machine (SVM), Decision Tree (DT), RF, and Multi-Layer Perceptron (MLP) in medical fraud detection. By extracting 19 key features of medical insurance fraud, he provided valuable suggestions for insurance management departments to establish an automatic audit mechanism to eliminate fraud [12]. Anam Syaiful proposed using SVM and Particle Swarm Optimization (PSO) for health claim insurance prediction. Experimental results showed that support vector machines using the PSO algorithm performed well in health claim insurance prediction, and it was proven that the performance of support vector machines using the PSO algorithm was better than that of standard support vector machines [13]. These studies have achieved positive results in model selection and practical applications, but they still focus on the comparison and application of traditional ML models, lacking the systematic use of complex feature interactions and deep learning structures.

As research deepens, scholars have gradually tried to combine deep learning with factorization machines to capture complex feature relationships. DeepFM can automatically learn complex feature interactions and performs well on high-dimensional sparse data [14,15]. To identify predictive features from complex raw data, Chen Ling proposed an ensemble diversity-enhanced Extreme DeepFM model and designed an ensemble diversity metric in each hidden layer. He considered ensemble diversity and prediction accuracy in the objective function and proved the effectiveness of the proposed model through datasets [16]. To better explore the potential relationship between data, Yu Huaidong proposed a Gate Attention Factorization Machine (GAFM) model based on the dual factors of accuracy and speed, and used the gate structure to control speed and accuracy. The results showed that the GAFM model was superior to the existing factorization machine in terms of speed and accuracy [17]. To process complex industrial data with different features, nonlinear relationships, and a large number of unlabeled samples, Ren Lei proposed a data-driven self-supervised Long Short-Term Memory (LSTM)-DeepFM model for industrial soft measurement, exploring the interdependence between features and the dynamic fluctuations in time series. Experiments showed that the proposed method reached the state-of-the-art level compared with existing methods [18]. These works have significantly enhanced the model's ability to process high-dimensional sparse features and have been verified to outperform traditional methods on multiple datasets. However, the label lag noise and time distribution drift problems in the data are still generally ignored, resulting in insufficient stability of the model in inter-period prediction.

To address the above challenges—which are prevalent in dynamic risk scenarios ranging from insurance claims to supply chain management—this paper proposes a DeepFM framework with dual correction for lag and drift specifically tailored for commercial health insurance claims risk prediction. In terms of the label lag problem, the Inverse Probability Weighting (IPW) mechanism is applied to reweight the missing labels of unresolved samples to alleviate the model offset caused by pseudo-negative samples. To address the issue of temporal distribution drift, timestamp information is embedded in a shared embedding space, and adversarial training constraints are combined to minimize distribution differences across different time windows, enabling the model to achieve stronger temporal invariance in representation learning. Compared to traditional FM, DeepFM not only explicitly captures second-order feature interactions through the FM branch but also learns complex high-order nonlinear patterns through the DNN branch, enabling the model to better fit high-dimensional, sparse feature environments. By improving the model's stability and predictive accuracy in dynamic business environments, this paper provides a more practical reference not only for intelligent insurance risk management but also for sectors facing similar data challenges, such as optimizing risk control and decision-making in the complex, long-cycle operations of the industry.

II. COMMERCIAL HEALTH INSURANCE CLAIMS RISK MODELING USING DEEPFM ALGORITHM OPTIMIZATION

A. MULTI-SOURCE FEATURE PROCESSING AND EMBEDDING REPRESENTATION

In commercial health insurance claims risk modeling, this paper integrates four core data sources, whose feature information is shown in Table 1:

TABLE 1. Data sources and feature information

Classification	Sequence	Feature
Demographic information	1	Age
	2	Gender
	3	Occupation Code
	4	Region identification (ID)
Health examination data	1	Body Mass Index (BMI)
	2	Blood pressure
	3	Blood sugar level
	4	Medical history code
Policy attributes	1	Coverage level
	2	Payment period
	3	Insurance type
	4	Deductible setting
Historical claims behavior	1	Annual frequency of medical visits
	2	Drug consumption categories
	3	Previous claims status

In Table 1, demographic information includes age, gender, occupation code, and region identification; health examination data covers BMI, blood pressure, blood sugar level, and medical history code; policy attributes include

insurance coverage level, payment period, insurance type, and deductible setting; historical claim behavior records annual medical frequency, drug consumption category, and previous claim status. First, the demographic information is classified and sorted, and the categorical variables are initially represented by one-hot encoding. On this basis, the high-cardinality categories are mapped to a low-dimensional continuous vector space through the embedding layer to alleviate the curse of dimensionality and sparsity problems. For each high-cardinality categorical feature, its embedding dimension d_i is dynamically determined based on the number of unique values v_i for that feature, following the empirical formula $d_i = \min(64, \lceil \sqrt{|v_i|} \rceil)$, with an upper limit of 64 to control model complexity. Categories appearing in the training set with a frequency below the threshold (set to 10) are uniformly mapped to a special label [RARE], and a learnable shared vector is assigned to them in the embedding layer. Categories appearing during the testing phase but not seen in training are also mapped to [UNK] unknown labels, and the [RARE] corresponding embedding vectors are reused. The embedding vectors of all categorical features are optimized through end-to-end joint training, sharing the same set of embedding parameters between the FM and DNN branches of DeepFM. This sharing mechanism not only significantly reduces the number of parameters but also enhances the generalization ability of sparse feature combinations through explicit modeling of second-order interactions in the FM branch. The embedding operation can be expressed as:

$$v_i = E_i \cdot x_i \tag{1}$$

In Eq. (1), x_i is the one-hot encoding vector, and $E_i \in \mathbb{R}^{|C_i| \times d_i}$ is the learnable embedding matrix. The embedding matrix is updated by backpropagation during the model training process so that each category can capture the potential semantic features related to claim risk in the low-dimensional space. For the continuous variables in the health examination data, missing value processing and outlier detection are performed. Missing values are filled using the mean imputation method:

$$\hat{x}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ij} \tag{2}$$

Outliers are identified and truncated using the Z-score method to prevent extreme values from having an adverse effect on model training. Continuous features are then standardized to have zero mean and unit variance, that is [19]:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{3}$$

where, μ_j and σ_j are the mean and standard deviation of the feature in the training set, respectively. The policy attribute is a categorical feature with high cardinality. To fully express the impact of different policy combinations on claim risk, this paper uses embedded coding to map the categorical feature c to a continuous vector in a d-dimensional dense vector space, which is mathematically expressed as:

$$e_c = W_c \cdot v_c \tag{4}$$

where, $W_c \in \mathbb{R}^{d \times k}$ is a trainable embedding matrix. Through end-to-end optimization, the model automatically learns semantic similarity. The embedding dimension d is dynamically set based on the feature information entropy: $d = 64$ is used for high-information features, and $d = 32$ is used for low-information features. The embedded policy attribute vector, demographic features, and health examination features together constitute a preliminary low-dimensional feature vector set X_{init}.

Historical claims behavior, as a time-accumulated feature, is sparse and non-uniformly distributed. Directly inputting it into a deep network may lead to unstable gradients. This paper first logarithmically transforms the numerical claims feature $y' = \log(y + 1)$ to compress the magnitude difference, and then standardizes it to make it have a similar scale to other continuous features. At the same time, the categorical historical event features are mapped to vector representations through embedding coding, and then concatenated with the numerical features to form a unified feature vector X_{hist}. The process is expressed as:

$$X_{\text{hist}} = [\tilde{x}_{\text{num}}, v_{\text{cat}}] \tag{5}$$

where, $\tilde{x}_{\text{num}}$ is the standardized continuous claims feature, and v_{cat} is the embedded categorical claims event feature vector. Finally, demographic information, health examination data, policy attributes, and historical claims behavior features are integrated into a unified low-dimensional dense vector $X \in \mathbb{R}^{d_{\text{total}}}$ through vector concatenation operations, that is:

$$X = [v_{\text{demo}}, \tilde{x}_{\text{health}}, v_{\text{policy}}, X_{\text{hist}}] \tag{6}$$

where, v_{demo} represents the demographic feature embedding vector; $\tilde{x}_{\text{health}}$ represents the standardized health examination feature; v_{policy} represents the embedded policy feature; X_{hist} represents the historical claims feature vector.

B. DEEPFM DUAL-CHANNEL MODELING

Commercial health insurance claims data typically exhibit high-dimensional sparsity and strong feature interactions. While traditional deep neural networks can fit high-order nonlinear relationships for this data structure, their generalization ability to sparse categorical features is weak, and they struggle to explicitly model meaningful low-order interactions. Pure sequence models, while suitable for time series, are problematic because health insurance claims predictions are usually based on static snapshot features, lacking strong temporal dependencies; forcibly introducing

sequence modeling would only increase redundancy and complexity. In contrast, DeepFM uses a shared embedding layer to uniformly represent sparse features and integrates FM and DNN channels in parallel. This avoids the overfitting risk of DNNs on sparse data and overcomes the limitation of traditional FM in modeling high-order interactions, balancing expressive power, generalization performance, and business interpretability. After completing the structured processing and embedding of multi-source features, this paper constructs a DeepFM dual-channel architecture based on shared embedding vectors to simultaneously capture the expressive power of second-order explicit interactions and higher-order nonlinear interactions, as shown in Figure 1. The FM branch analytically models second-order interactions between domains, identifying statistically significant pairwise risk associations. The DNN branch then performs multi-layer nonlinear combinations on the embedded features to capture complex risk patterns under high-order, multivariate coupling. Both channels share the same embedding representation, and joint training achieves coordinated optimization of feature representation and interaction modeling. Finally, the fusion layer generates a claim probability prediction.

It is assumed that the field set of the entire sample is $F = \{1, \ldots, m\}$, and the corresponding input of each field i after preprocessing is x_i. For categorical fields, they have been mapped to low-dimensional vectors $v_i \in \mathbb{R}^d$ through the embedding matrix E_i; for scalar continuous fields, they are denoted as $\tilde{x}_{ji} \in \mathbb{R}$ after normalization. The original values are retained in the model for linear terms, and vectorized representations are obtained through linear mapping to participate in interactions. For unified representation, the field vector is defined as:

$$e_i = \begin{cases} v_i, & i \in C, \\ \tilde{x}_{ji} \cdot u_i, & i \in N, \end{cases} \tag{7}$$

where, $u_i \in \mathbb{R}^d$ is the learnable basis vector of the continuous field, and C and N represent the categorical field set and the continuous field set, respectively. The shared embedding set is denoted as $E = \{e_1, e_2, \ldots, e_m\}$.

1) FM Branch: Second-Order Explicit Interaction Modeling

The FM branch adopts the classic second-order factorization machine form. The model output consists of global bias, linear terms, and second-order interaction terms, which are formalized as [20,21]:

$$y_{\mathrm{FM}}(x) = b_0 + \sum_{i=1}^{m} \omega_i x_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \langle e_i, e_j \rangle x_i x_j \tag{8}$$

where, b_0 is the global bias; ω_i is the linear weight; $\langle e_i, e_j \rangle = e_i^\top e_j$ represents the vector inner product to characterize the second-order interaction strength. To improve computational efficiency and avoid explicit double summation, an equivalent optimization expression is adopted [22]:

$$\sum_{i<j} \langle e_i, e_j \rangle x_i x_j = \frac{1}{2} \left(\left\| \sum_{i=1}^{m} x_i e_i \right\|_2^2 - \sum_{i=1}^{m} x_i^2 \|e_i\|_2^2 \right) \tag{9}$$

Based on this branch, a direct and interpretable contribution measure is provided for the pairwise association between high-cardinality categories and sparse fields, identifying second-order risk factors that are strongly correlated with claims.

2) DNN Branch: High-Order Nonlinear Interactive Learning

The DNN branch accepts the concatenated field vector as input and implements high-order feature combinations in the form of MLP. All field vectors are concatenated into the input vector in field order [23]:

$$h^{(0)} = [e_1; e_2; \ldots; e_m] \in \mathbb{R}^{m \cdot d} \tag{10}$$

After LL layers of fully connected transformation and nonlinear activation, the calculation is performed layer by layer:

$$h^{(l)} = \zeta \left(\mathrm{BN} \left(W^{(l)} h^{(l-1)} + b^{(l)} \right) \right), \quad l = 1, \ldots, L \tag{11}$$

where, $W^{(l)}$ and $b^{(l)}$ are the weights and biases of the l-th layer; $\mathrm{BN}(\cdot)$ is batch normalization; $\zeta(\cdot)$ is the activation function. This paper adopts ReLU (Rectified Linear Unit) and applies Dropout and L2 regularization after each layer to suppress overfitting. The final DNN output is a one-dimensional score [24]:

$$y_{\mathrm{DNN}}(x) = w_o^\top h^{(L)} + b_o \tag{12}$$

The DNN branch learns arbitrary order interactions by layer-by-layer combination, fitting the complex nonlinear effects of demographics, health check indicators, policy attributes, and claim behavior sequences on claim risk when they co-occur, and capturing the implicit multivariate coupling effect.

3) Channel Fusion Strategy

To take into account the interpretable second-order contribution of FM and the high-order expression of DNN, this paper adopts an adaptive gated fusion mechanism instead of a simple weighted sum, and dynamically adjusts the contribution ratio of the two channels according to the sample characteristics. The gate weight is defined as:

$$\alpha(x) = \delta \left(w_g^\top \phi(h^{(0)}) + b_g \right) \tag{13}$$

where, $\phi(\cdot)$ is a shared compression mapping of one layer, so that $\alpha(x) \in (0, 1)$. The final combined score is:

$$\hat{s}(x) = \alpha(x) \cdot y_{\mathrm{FM}}(x) + (1 - \alpha(x)) \cdot y_{\mathrm{DNN}}(x) \tag{14}$$

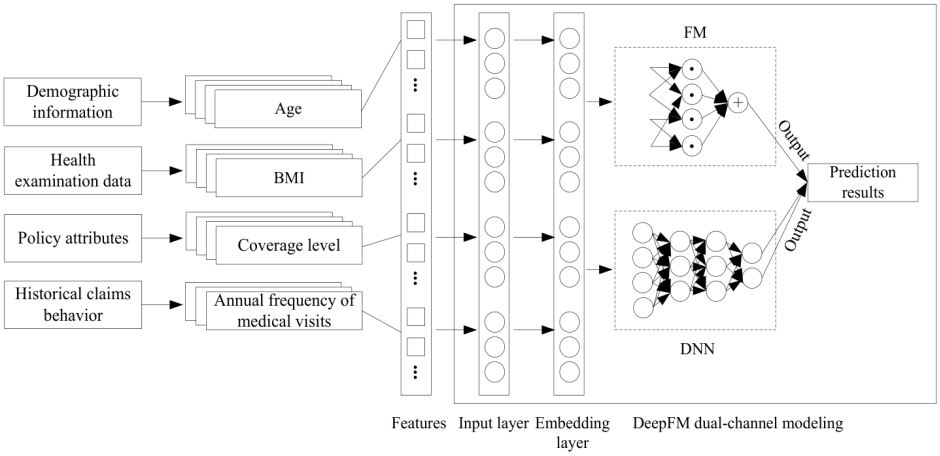

FIGURE 1. DeepFM dual-channel modeling

The claim probability prediction is obtained by Sigmoid mapping:

$$\hat{p}(x) = \text{Sigmoid}(\hat{s}(x)) \tag{15}$$

When the second-order interpretable interaction in the sample is dominant, the model can be biased towards FM output; when complex medical history and long-term medical behavior act together, the high-order coupling effect is more important, and the model can be biased towards DNN output, taking into account both interpretability and expressiveness.

C. TIME LAG AND DRIFT CORRECTION MECHANISM

To address the impact of label time lag and distribution drift, this paper constructs a dual correction mechanism that effectively corrects for label time lag noise and temporal distribution drift in the training data, improving the model's stability and generalization capabilities. This correction mechanism consists of two parts: 1) label time lag correction: by applying inverse probability weighting and a robust loss function, it corrects for label bias caused by unresolved cases, ensuring that the model learns accurate claims risk patterns; 2) time drift correction: by applying time-domain adversarial training and distribution consistency constraints in the embedding layer, the model maintains stable prediction performance for data from different time periods.

1) Label Time Lag Correction Mechanism

The label lag correction mechanism proposed in this paper is a joint correction strategy of probability weighting and robust loss embedded in the model training process. In claims data, some policies are incorrectly labeled as unclaimed (i.e., pseudo-negative samples) due to long settlement cycles [25,26]. However, the actual claims status of these samples is uncertain. To alleviate the label noise introduced by this, this paper uses IPW to dynamically reweight the loss of each sample and combines it with symmetric cross-entropy loss to improve the robustness of the model to noisy labels, as shown in Figure 2.

First, by assigning a weight to each sample, which is inversely proportional to the probability that the sample belongs to the unresolved case category, the model pays more attention to those more reliable labels during training, reducing the influence of unresolved case samples. Assuming that the weighting factor of label lag noise is W_i, and the label of sample i is y_i, then the weighted loss function of the sample is [27,28]:

$$L_{\text{IPW}} = -\sum_{i=1}^{N} W_i \cdot \log p(y_i|x_i) \tag{16}$$

where, $p(y_i|x_i)$ is the model's predicted probability for sample i, and W_i is calculated by the inverse probability, which represents the label reliability of sample i. The specific calculation formula is:

$$W_i = \frac{1}{P(\text{sample is final}|x_i)} \tag{17}$$

where, $P(\text{sample is final}|x_i)$ is the probability that sample i belongs to the closed case category under the given feature x_i. The design of the robust loss function aims to further enhance the robustness of the model to noisy labels. Symmetric Cross-Entropy Loss (SCEL) is applied to adjust the importance of difficult and easy samples by weighting. Its form is:

$$L_{\text{SCEL}} = -\sum_{i=1}^{N} \left[\alpha_x \cdot y_i \log p(y_i|x_i) + (1-\alpha_x) \cdot (1-y_i) \log(1-p(y_i|x_i))\right] \tag{18}$$

When dealing with lag noise, the adjustment of α_x enables the model to maintain efficient prediction capabilities in the case of unbalanced samples.

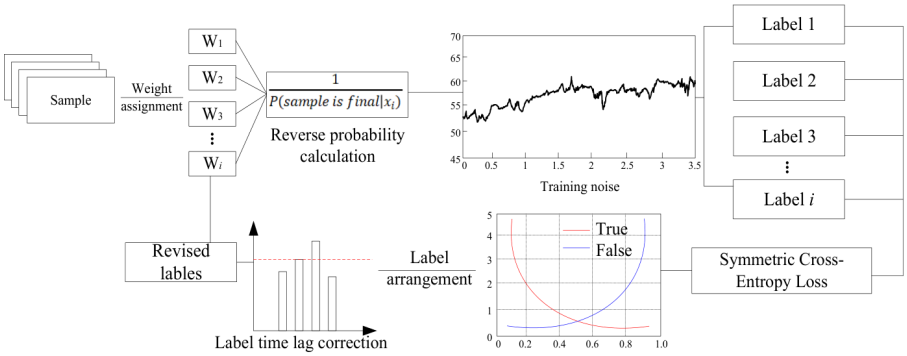

FIGURE 2. Label lag correction

2) Temporal Drift Correction Mechanism

Distribution drift arises from factors such as time, policy, and product changes, causing the distribution of data in different time periods to change, resulting in a decrease in the model's predictive ability across time periods [29–31]. This paper applies temporal adversarial training and distribution consistency constraints in the embedding layer of the DeepFM model to maintain stable performance in cross-time scenarios. As shown in Figure 3, this approach first eliminates the effects of temporal drift through an adversarial process. The time drift adversarial mechanism adopted in this paper is essentially a feature decoupling method based on domain adversarial training. Its goal is to force the shared embedding layer of DeepFM to learn a risk representation that is independent of time and has cross-time period generalization ability. The observation time window is used as the domain variable, and a discrete time domain label is assigned to each sample. Under this setting, the model contains two core components: (1) Feature encoder: the shared embedding layer of DeepFM, which outputs a low-dimensional dense representation; (2) Time domain discriminator: a lightweight MLP that takes the low-dimensional dense representation as input and predicts the time domain to which it belongs. The training process adopts an adversarial optimization strategy: the feature encoder tries to minimize the loss of the main task (claims prediction) while maximizing the classification error of the time domain discriminator (i.e., making the discriminator unable to accurately determine which time period the sample comes from); while the time domain discriminator tries to accurately predict the time domain label. This adversarial goal is achieved through the gradient reversal layer—during backpropagation, the discriminator gradient is reversed after passing through the GRL, thereby driving the encoder to generate a feature representation that is indistinguishable from the time domain.

It is assumed that t_i is the time label corresponding to sample i; $f(x_i)$ is the output representation of input feature x_i in the embedding layer; $D_{\text{adv}}(f(x_i))$ is the predicted probability of the adversarial network for the time label of sample i. The loss function of time domain adversarial training is [32]:

$$L_{\text{adv}} = -\sum_{i=1}^{N} \log\left(1 - D_{\text{adv}}(f(x_i))\right) \tag{19}$$

By minimizing the prediction ability of the adversarial network for time labels, the DeepFM model learns a stable and time-independent feature representation. Then, the distribution consistency constraint is applied to learn the same feature distribution on the training data of different time periods. The temporal stability of the model is guaranteed by minimizing the difference in feature distribution of the model in different time periods. This paper adopts Maximum Mean Discrepancy (MMD) as the distribution consistency measure to calculate the difference in data distribution from different time periods:

$$L_{\text{MMD}} = \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} f(x_i) - \frac{1}{N_2} \sum_{i=1}^{N_2} f(x_i') \right\|_2^2 \tag{20}$$

where, $f(x_i)$ and $f(x_i')$ are the embedding representations of sample features from time period 1 and time period 2, respectively, and N_1 and N_2 are the numbers of samples in the two time periods. By minimizing this loss function, the distribution difference between time periods is minimized to the greatest extent. Combining the mechanisms of label lag correction and time drift correction, the final comprehensive correction loss function is obtained:

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{IPW}} + \mathcal{L}_{\text{SCEL}} + \lambda_1 \mathcal{L}_{\text{adv}} + \lambda_2 \mathcal{L}_{\text{MMD}} \tag{21}$$

where, λ_1 and λ_2 are the weight hyperparameters that control time domain adversarial training and distribution consistency constraints. The final loss function combines corrections for label lag noise and temporal distribution drift to ensure model stability and generalization in dynamic environments.

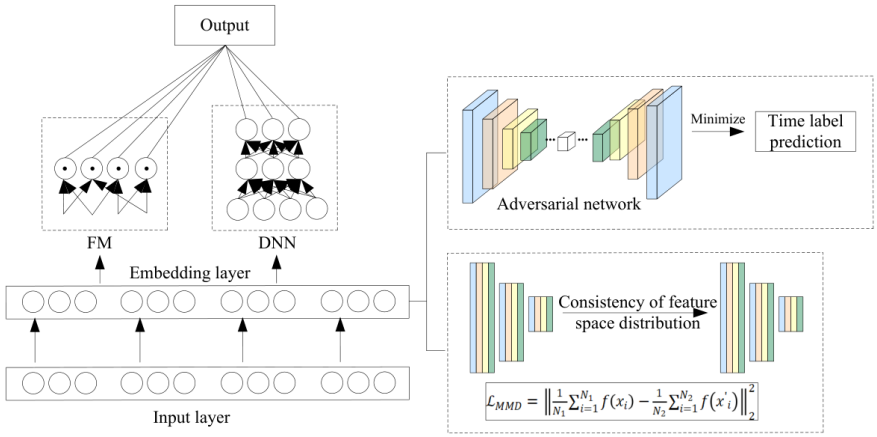

FIGURE 3. Time drift correction

D. RISK PROBABILITY OUTPUT AND OPTIMIZATION

Based on the temporal drift correction mechanism, the risk probability output and optimization process address the significant class imbalance problem in claims risk prediction. Various user feature data are input and learned through the FM branch and the DNN branch, respectively, to obtain two predicted probability outputs. A gating mechanism then weights and fuses the outputs of the FM and DNN branches to obtain the final risk prediction probability. Focal Loss is then applied during the training process, allowing the model to focus more on high-risk users who are difficult to classify and enhancing its predictive power for minority class samples. Finally, the Brier Score is minimized to optimize the calibration of the output probabilities. First, the learning results of the FM and DNN branches are fused using a gated fusion mechanism to construct a prediction framework. The final risk prediction probability is obtained through the weighted fusion of the FM and DNN outputs. Assuming that the output of the FM branch is p_{FM}, and the output of the DNN branch is p_{DNN}, the output of the gated fusion layer is:

$$p_{\mathrm{final}} = g \cdot p_{\mathrm{FM}} + (1 - g) \cdot p_{\mathrm{DNN}} \tag{22}$$

where, g is the learned gating parameter, which represents the weight of the FM and DNN outputs in the final probability prediction. Through the gating mechanism, the model can adaptively select the contribution of the FM or DNN branch for different samples, thereby improving the prediction performance. Focal Loss is applied to alleviate the problem of class imbalance. In the commercial health insurance claim prediction task, the imbalance of positive and negative samples causes the model to pay too much attention to negative samples (i.e., low-risk users) and ignore positive samples (i.e., high-risk users). By assigning lower weights to easy-to-classify samples, focusing on those difficult-to-classify samples, especially positive samples, the model's learning ability for minority classes is enhanced [33,34]:

$$L_{\mathrm{Focal}} = -\sum_{i=1}^{N} \phi (1 - p_i)^{\gamma} \log(p_i) \tag{23}$$

where, p_i is the predicted probability of sample i; ϕ is the balance factor used to control the ratio of positive and negative samples; γ is a hyperparameter for adjusting the attention to easy and difficult samples. When $\gamma > 0$, Focal Loss reduces the attention to easy-to-classify samples and strengthens the training of difficult-to-classify samples [35]. In this way, the prediction ability for high-risk users is improved. Brier Score fine-tuning is an effective means to optimize the calibration of predicted probabilities. Brier Score measures the mean square error between the predicted probability and the true label:

$$B = \frac{1}{N} \sum_{i=1}^{N} (p_i - y_i)^2 \tag{24}$$

By minimizing the Brier Score, the model can improve the calibration of predicted probabilities, making the predicted probabilities closer to the true probability of occurrence, thereby improving the credibility and usability of the prediction results.

III. COMMERCIAL HEALTH INSURANCE USER CLAIMS RISK ASSESSMENT

A. EXPERIMENTAL SETUP

The experimental data used in this paper come from claims data from a commercial health insurance company, covering the period from January 2021 to June 2023. The dataset includes users who purchased health insurance products from the insurance company and covers a sample of users from different regions. The dataset labels each user as to whether they have a claim during the prediction period, with a value of 0 indicating no claim and 1 indicating a claim. To further validate the model's generalization capabilities, this paper also uses the publicly available insurance.csv dataset as a supplementary dataset. This dataset contains

customer characteristics and insurance-related data, suitable for exploring the relationship between health insurance claims and risk. Details of the dataset are shown in Table 2:

TABLE 2. Dataset details

Classification	Commercial health insurance data	insurance.csv
Sample size	4,721	1,338
Number of features	15	6
Main features	Features shown in Table 1	Age, Gender, BMI, Smoking status, Number of children, Health insurance premiums
Data type	Numerical type, categorical type	Numerical type, categorical type

Table 2 shows a dataset containing 4,721 policy records, corresponding to 3,892 unique policyholders (some users hold multiple policies). To simulate the time distribution drift in real-world business, the data were divided into five consecutive semi-annual windows (2021H1, 2021H2, 2022H1, 2022H2, 2023H1) to ensure natural time evolution during the training and testing phases, reflecting covariate drift and concept drift caused by policy adjustments, product iterations, and changes in user behavior.

During data preprocessing, one-hot encoding is used to initially represent categorical variables in demographic information. To mitigate high-dimensional sparsity, an embedding layer is used to map these categorical features into a low-dimensional continuous vector space. For continuous variables in the health examination data, mean imputation is used to address missing values, and Z-score normalization is used for outliers. Embedding encoding is also used to map categorical features in policy attributes and historical claims. Data are desensitized using k-anonymization and differential privacy techniques to ensure compliance with insurance industry data security standards. Finally, all features are standardized and merged using vector concatenation to form a unified low-dimensional dense feature vector for further model training. Table 3 shows the hyperparameter settings for this model:

TABLE 3. Hyperparameter settings

Module	Parameter	Specifications
DeepFM	FM embedding dimension	64
	DNN layers	3
	Number of nodes per layer in DNN	256
	Activation function	ReLU
	Dropout rate	0.2
Loss function	Focal Loss γ	2
	Brier Score adjustment factor	0.1
Training optimization	Learning rate	0.001
	Batch size	128
	Training rounds	20

To determine the optimal weight coefficients in the combined loss function, a grid search was performed on the validation set with candidate values ranging from {0.01, 0.05, 0.1, 0.2, 0.5, 1.0}. The weighted score combining ECE and discriminative ability was used as the selection criterion. A weight coefficient of 0.1 was ultimately chosen. To further evaluate the model's sensitivity to the weight coefficients, its performance variation within the range of [0.05, 0.2] was tested. When the weight coefficients varied within the range of 0.05–0.2, the AUC fluctuation was less than ± 0.005, and the ECE variation did not exceed ± 0.003, indicating that the overall model performance is not sensitive to small changes in the weight coefficients. This demonstrates that the proposed dual-loss framework has good stability, and that 0.1 is a robust and effective trade-off. After model training, this paper further conducts feature importance analysis. By comprehensively measuring the embedding vector weights, the FM branch interaction coefficient, and the DNN network gradient contribution, the impact of various features on claims risk prediction is quantified, as shown in Table 4:

TABLE 4. Feature importance analysis

Feature	Significance score (%)	Feature	Significance score (%)
Previous claims status	18.8	Deductible setting	4.7
Medical history code	15.5	Smoking status	3.9
Age	12.9	Drug consumption categories	3.2
BMI	10.6	Occupation Code	2.6
Annual frequency of medical visits	9.7	Region ID	1.8
Blood pressure	7.9	Health insurance premiums	0.9
Blood sugar level	6.8	Gender	0.7

As can be seen from Table 4, health and behavior-related features such as historical claim status, past medical history, age, and body mass index contribute the most to claim risk prediction, indicating that the user's past health status and claim record are the core determinants of claim risk. Health monitoring indicators such as annual visit frequency, blood pressure, and blood sugar also show high importance, reflecting the model's sensitivity to dynamic health status.

To fully verify the performance of the DeepFM model in this paper, it is evaluated from the perspectives of risk prediction, model calibration, and stability, and compared with mainstream baseline models:

(1) DNN-Embedding Through the neural network, the embedding layer is connected to the target loss for joint training to learn its feature expression. This is an end-to-end embedding training method.

(2) WDL WDL is a structure that combines a wide model (linear model) and a deep model (neural network). The wide part uses a linear model to capture the low-order interactions of sparse features, while the deep part uses a deep neural network to capture the complex interactions of high-order features.

(3) Gradient Boosting Decision Tree (GBDT) GBDT, as an ensemble learning method widely used in classification tasks, has strong model expression capabilities, especially in high-dimensional sparse data.

B. EXPERIMENTAL RESULTS

1) Risk Prediction Results

In risk prediction assessment, this paper uses the AUC metric to evaluate the classification performance of the model. The classification ability of the model at different decision thresholds is comprehensively considered. In the experiments, all models are evaluated using the same

training set and test set, with the training set accounting for 80% and the test set accounting for 20%. The output probability value of each model is input into the ROC (Receiver Operating Characteristic) curve, and the overall classification ability of the model is finally evaluated by calculating the AUC value. The results are shown in Figure 4:

The results in Figure 4 show that the proposed models perform better in the classification task of commercial health insurance claim risk prediction, demonstrating strong discriminative power in distinguishing high-risk from low-risk users. In Figure 4A, the DeepFM model achieves the highest AUC of 0.853, demonstrating its ability to maintain a low false positive rate while maintaining a high recall rate. The DNN-Embedding and WDL models in Figure 4B and 4C perform worse, with AUCs of 0.820 and 0.824, respectively. Both models outperform traditional methods in modeling feature interactions, but fall short of DeepFM in capturing high-order nonlinear feature combinations. In contrast, the GBDT model in Figure 4D performs the worst, with an AUC of 0.818, showing some limitations in handling complex, high-dimensional features and adapting to distribution shifts. The proposed model achieves a better balance between feature interactions and generalization. It unifies multi-source features through shared embeddings, explicitly models second-order interactions with FM, captures high-order nonlinearities with DNN, and dynamically balances the contributions of both through gated fusion. This ensures interpretable recognition of high-cardinality sparse feature pairs while also enabling the learning of implicit risk patterns from complex coupling factors, resulting in strong classification performance in claims risk prediction.

To further comprehensively evaluate model performance, this paper applies the Precision and F1-Score metrics to comprehensively measure the model's balanced ability in identifying the positive class (high-risk users). The experiments use 5-fold cross-validation, and the results are shown in Table 5:

TABLE 5. Comparison of F1-Score and Precision

Model	Precision	F1-Score
DeepFM	0.842 ± 0.011	0.801 ± 0.009
DNN-Embedding	0.815 ± 0.014	0.763 ± 0.015
WDL	0.821 ± 0.013	0.776 ± 0.012
GBDT	0.798 ± 0.016	0.751 ± 0.017

In Table 5, the mean Precision values for the DNN-Embedding, WDL, and GBDT models are 0.815, 0.821, and 0.798, respectively; the mean F1-Score values are 0.763, 0.776, and 0.751, respectively. The DeepFM model proposed in this paper significantly outperforms the other comparison models in both Precision and F1-Score, achieving a Precision of 0.842. This indicates that the highest proportion of actual claims are made among users predicted as high-risk by the model. High Precision is crucial for insurance companies' practical applications, as it directly impacts the cost-effectiveness of risk intervention actions and avoids wasting excessive resources on misclassified low-risk users. DeepFM also achieves the optimal F1-Score of 0.801, demonstrating that the model strikes a good balance between precisely identifying high-risk users and minimizing underidentification of high-risk users, resulting in stronger overall classification performance.

2) Model Calibration Results

To analyze the calibration performance of the DeepFM model proposed in this paper in real-world applications, Brier Score (mean squared error) and ECE are used for evaluation. Under a stratified five-fold cross-validation framework, the samples are binned into deciles based on the predicted probabilities output by the model on the test set, with risk groups plotted on the horizontal axis. The Brier Score and ECE, $\sum_k (n_k/N)|acc_k - conf_k|$, are calculated for each group, comparing the calibration performance of each model across different risk segments, as shown in Figures 5 and 6.

Figure 5 shows significant differences in the probability prediction accuracy of the four models across different risk groups. The DeepFM model performs best overall, achieving a Brier Score of only 0.079 across all risk groups. This indicates that its predicted probabilities best fit the true distribution and exhibit minimal fluctuation across different risk groups, demonstrating strong stability and generalization capabilities. The WDL and DNN-Embedding models perform second best, with Brier Scores slightly higher than DeepFM overall, suggesting that DeepFM still has some shortcomings in probability calibration for complex samples. The GBDT model performs the worst overall, with Brier Scores exceeding 0.1 in multiple risk groups, indicating significant deviations between its probability predictions and actual risk. Overall, DeepFM, leveraging its deep feature interactions and nonlinear fitting capabilities, achieves superior probability calibration. However, WDL, DNN-Embedding, and GBDT struggle to account for all risk groups and perform relatively poorly. The results in Figure 6 show that the proposed DeepFM model performs better in terms of calibration. In Figure 6A, DeepFM has the lowest average ECE value of 0.030, indicating the best agreement between predicted probabilities and actual risk. The WDL model in Figure 6C achieves an average ECE of 0.053, while the DNN-Embedding model in Figure 6B achieves an average ECE of 0.031. The GBDT model in Figure 6D exhibits the largest calibration error, with an average ECE of 0.065. Compared to other models, the DeepFM model, thanks to its dual-channel architecture, is able to simultaneously capture low-order feature interactions and high-order nonlinear relationships, maintaining relatively stable calibration performance across all risk segments. This excellent calibration performance is crucial for the practical application of commercial health insurance, as it ensures that the risk probabilities output by the model are more

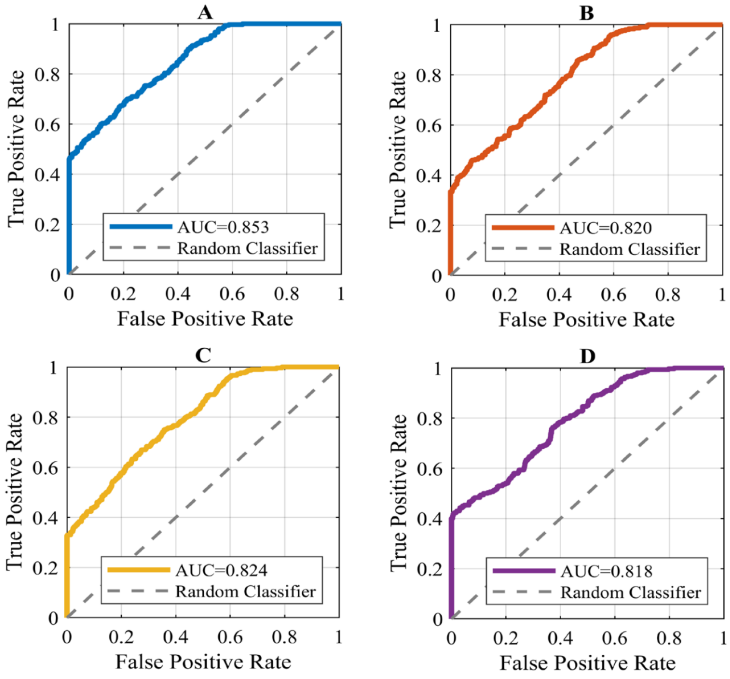

FIGURE 4. AUC comparison. (A) AUC of DeepFM; (B) AUC of DNN-Embedding; (C) AUC of WDL; (D) AUC of GBDT

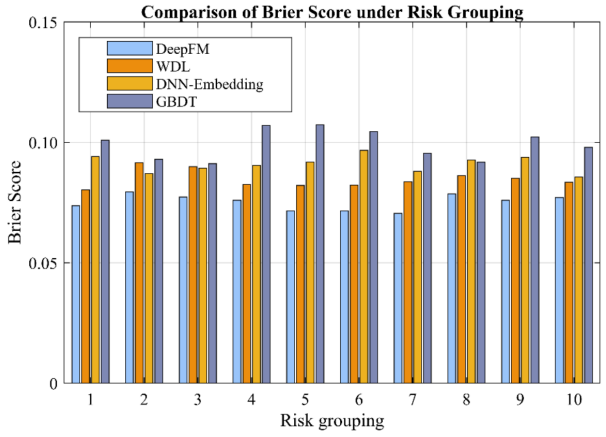

FIGURE 5. Brier Score comparison

reliable and can serve as a basis for premium pricing and risk management. The results demonstrate that the DeepFM model not only excels in discriminative ability but also exhibits significant advantages in probability calibration, making it more suitable for commercial insurance applications requiring precise probability estimates.

3) Stability Results

To test the stability of claims risk prediction, the model's ability to handle time-delay noise and adaptability to time-varying distribution drift are evaluated. In the time-delay noise processing evaluation, different levels of outstanding case ratios are constructed by simulating the injection of time-delay noise, set to 0%, 10%, 20%, 30%, and 40%, respectively, to simulate label lag in the claims process. Four models—DeepFM, WDL, DNN-Embedding, and GBDT—are trained and predicted at various noise levels. A comprehensive comparison of the three metrics, AUC, Brier Score, and ECE, is conducted from the perspectives of classification ability, probability calibration, and stability, as shown in Figure 7.

The results in Figure 7 show that as the pending case ratio increases, the predictive performance of all models is affected to varying degrees. In Figure 7A, the AUC shows an overall downward trend, but DeepFM maintains a score of 0.789 at a 40% pending case ratio, representing a

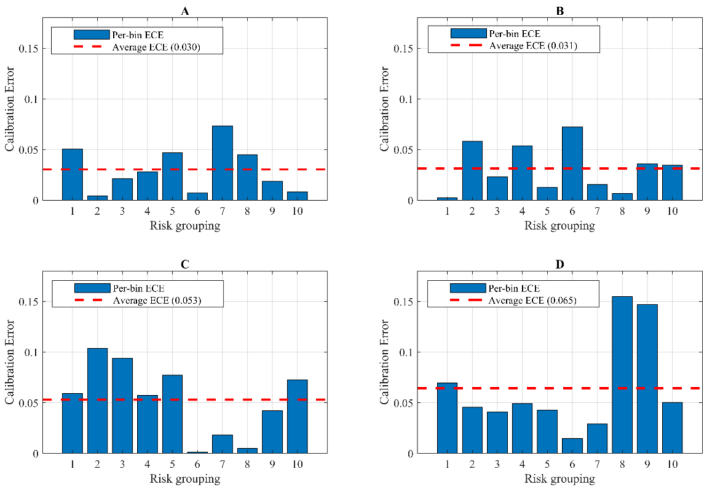

FIGURE 6. ECE Comparison. (A) ECE of DeepFM; (B) ECE of DNN-Embedding; (C) ECE of WDL; (D) ECE of GBDT

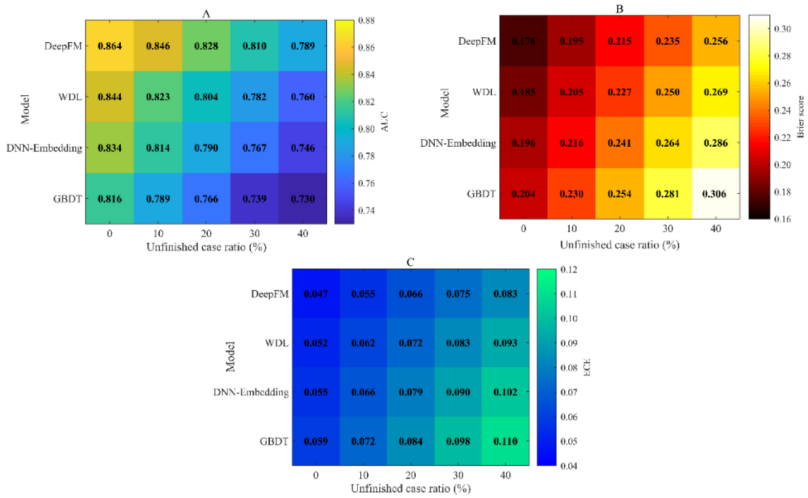

FIGURE 7. Comparison of time-delay noise processing capabilities. (A) AUC under time-delay noise; (B) Brier Score under time-delay noise; (C) ECE under time-delay noise

3.8% and 5.8% improvement over the next-best WDL and DNN-Embedding scores of 0.760 and 0.746, respectively. This demonstrates its superior ability to distinguish positive and negative examples under high noise conditions. The Brier Score in Figure 7B and the ECE in Figure 7C show an upward trend, indicating a gradual weakening of the calibration and confidence of the predicted probabilities. However, DeepFM's Brier Score and ECE at a 40% pending case ratio remain relatively small, consistently below 0.085, demonstrating its ability to maintain good consistency in probabilistic predictions despite noise. In contrast, other models perform poorly in high-noise environments. Overall, the DeepFM method proposed in this paper, with its targeted loss imbalance design and rigorous regularization and training strategies, demonstrates superior stability and robustness against label lag noise in the commercial health insurance claims risk prediction task.

In the evaluation of adaptability to time-varying distribution drift, time-varying distribution drift is categorized as no drift, mild drift, moderate drift, and severe drift. A comprehensive comparison is also conducted using the AUC, Brier Score, and ECE metrics, as shown in Figure 8.

The results in Figure 8 show significant differences in the stability of different models under time-varying distribution drift. In Figure 8A, DeepFM experiences the smallest decrease in AUC, dropping only from 0.853 to 0.830, maintaining a relatively high overall performance. In contrast, GBDT experiences the largest drop, from 0.818 to 0.759, indicating its greatest sensitivity to drift. WDL and DNN-Embedding experience intermediate decreases, demonstrating some adaptability, but not as stable as DeepFM. The Brier Score in Figure 8B increases overall with drift. However, the increase in DeepFM's Brier Score is only 0.024, while GBDT's Brier Score increases by 0.063, showing a

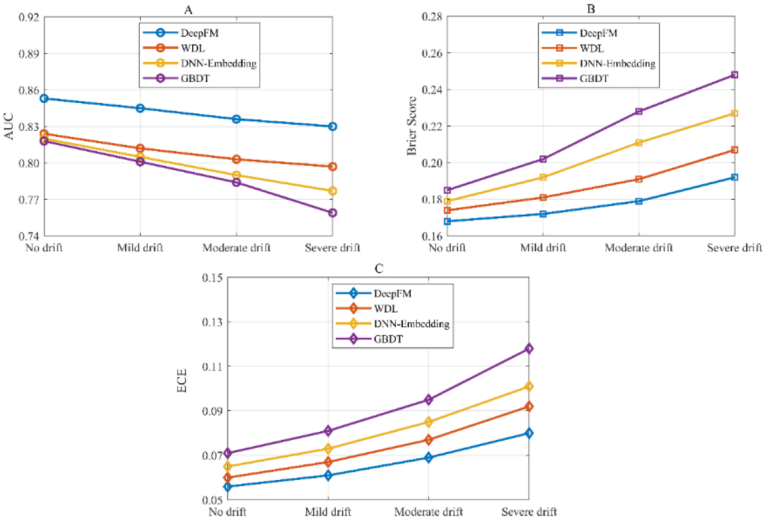

FIGURE 8. Comparison of adaptability to time-varying distribution drift. (A) AUC values under time-varying distribution drift; (B) Brier Score values under time-varying distribution drift; (C) ECE values under time-varying distribution drift

significant difference. Finally, in the ECE metric in Figure 8C, DeepFM increases from 0.056 to 0.080, also by only 0.024, remaining relatively low; GBDT, on the other hand, increases from 0.071 to 0.118, indicating a rapid decline in probability calibration performance under drift. Overall, DeepFM exhibits minimal fluctuations across all three metrics, demonstrating its ability to effectively address the decline in prediction accuracy caused by distribution drift while maintaining a more stable probability calibration.

IV. CONCLUSION

To improve the stability of commercial health insurance claims risk prediction—a critical goal shared by firms seeking to stabilize risk management in high-variability sectors —this paper combines DeepFM with a dual-correction framework for lag and drift. This methodology's focus on handling time-lagged and shifting data patterns is highly relevant for optimizing operational decision-making in both insurance risk control. Label lag noise is corrected using inverse probability weighting and a symmetric cross-entropy loss. Distribution drift is addressed using temporal adversarial training and a maximum mean difference constraint. Furthermore, an adaptive gated fusion mechanism is used to balance the FM and DNN branches. On the commercial health insurance dataset presented in this paper, the model achieves an AUC of 0.853, which is 4.0%, 3.5%, and 4.3% higher than DNN-Embedding, WDL, and GBDT, respectively. Regarding model calibration, the highest Brier Score across all risk groups is 0.079, with an average ECE of 0.030. At the methodological level, the approach not only significantly enhances the robustness of the model in a dynamic business environment, but also provides an effective reference for intelligent risk control—applicable to insurance —that takes into account both label dynamics and distribution time-varying properties; at the practical level, it provides insurance companies with a more robust, explainable, and efficient risk identification tool, offering reliable support for claims risk assessment in actual business decisions and serving as a blueprint for improving operational and quality control risk forecasting.

AUTHOR CONTRIBUTIONS

Conceptualization – Fei Mi; methodology – Fei Mi and Shuanghong Mi; investigation – Fei Mi and Shuanghong Mi; writing-original draft preparation – Fei Mi and Shuanghong Mi. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was funded by, 1. Social Science of Dongguan City, titled "Research on the Development Path Design and Practice of Commercial Medical Insurance Participation in chronic Disease Management" (No. 2023CG10)

2. Basic and Applied Basic Research Foundation of Guangdong Province - Surface Project, titled "Chemotherapy-immune synergistic anti-tuberculosis function and mechanism of Host cell targeted Nanoscale graphene oxide system" (No.: 2022A1515011223)

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This survey was conducted in compliance with Ethics Committee of Guangdong Medical University. Participants were informed of the study's purpose and data usage prior to participation. All data used in this study has undergone rigorous anonymization to fully protect user privacy. All analytical processes adhered to relevant laws, regulations, and industry data security standards. As this is a retrospective data analysis and does not interfere with actual business processes or user rights, no direct contact or intervention

was made with individuals, and no additional sensitive personal information was obtained. The relevant analyses strictly followed current laws, regulations, and research ethics to ensure the compliance, transparency, and social responsibility of the research.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Wu, J. Qiao, S. Nicholas, Y. Liu, and E. Maitland, "The challenge of healthcare big data to China's commercial health insurance industry: Evaluation and recommendations," BMC Health Services Research, pp. 22(1), 2022, doi: 10.1186/s12913-022-08574-2.

[2] L. Ajijola, O. Akindipe, A. Lawal, O. Afuwape, and P. Olaleye, "Evaluation of the effectiveness and accuracy of predictive modeling techniques on healthcare claims in Nigeria," Nigerian Journal of Management Studies, vol. 27, no. 1, pp. 175-199, 2025, [Online]. Available: https://njms.unilag.edu.ng/article/view/2568.

[3] A. Pnevmatikakis, S. Kanavos, G. Matikas, K. Kostopoulou, A. Cesario, and S. Kyriazakos, "Risk assessment for personalized health insurance based on real-world data," Risks, vol. 9, no. 3, pp. 46, 2021, doi: 10.3390/risks9030046.

[4] X. Liu, Y. Xiao, J. Chen, P. Ma, and W. Zhu, "Analysis of data exchange technology between medical and health insurance institutions from the perspective of patent measurement," Journal of Medical Informatics, vol. 45, no. 5, pp. 59-64, 2024, doi: 10.3969/j.issn.1673-6036.2024.05.010.

[5] G. Patra, C. Kuraku, S. Konkimalla, V. Boddapati, M. Sarisa, and M. Reddy, "An analysis and prediction of health insurance costs using machine learning-based regressor techniques," Journal of Data Analysis and Information Processing, vol. 12, no. 4, pp. 581-596, 2024, doi: 10.4236/jdaip.2024.124031.

[6] A. Abuzaid and I. Alkrunz, "Forecasting next year's health insurance claims using machine learning models," JITSI: Jurnal Ilmiah Teknologi Sistem Informasi, vol. 6, no. 2, pp. 120-131, 2025, doi: 10.62527/jitsi.6.2.443.

[7] S. Moon, D. Jagadeesh, S. Sultana, and M. Prasanna, "AI-driven machine learning model for health insurance claim prediction," International Journal of Communication Networks and Information Security, vol. 17, no. 3, pp. 900-906, 2025, [Online]. Available: https://www.proquest.com/scholarly-journals/ai-driven-machine-learning-model-health-insurance/docview/3232790510/se-2.

[8] K. Kaushik, A. Bhardwaj, A. Dwivedi, and R. Singh, "Machine learning-based regression framework to predict health insurance premiums," International Journal of Environmental Research and Public Health, vol. 19, no. 13, pp. 7898, 2022, doi: 10.3390/ijerph19137898.

[9] A. Alam and V. Prybutok, "Use of responsible artificial intelligence to predict health insurance claims in the USA using machine learning algorithms," Exploration of Digital Health Technologies, vol. 2, no. 1, pp. 30-45, 2024, doi: 10.37349/edht.2024.00009.

[10] V. Thakre, R. Poul, and A. Sawarkar, "Predictive precision: Unraveling health insurance claim patterns with logistic regression and decision trees," Cureus Journal of Computer Science, vol. 3, no. 1, pp. 1-10, 2025, doi: 10.7759/s44389-025-03010-y.

[11] S. Devaguptam, S. Gorti, T. Akshaya, and S. Kamath, "Automated health insurance processing framework with intelligent fraud detection, risk classification and premium prediction," SN Computer Science, pp. 5(5), 2024, doi: 10.1007/s42979-024-02801-9.

[12] V. Nalluri, J. Chang, L. Chen, and J. Chen, "Building prediction models and discovering important factors of health insurance fraud using machine learning methods," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 7, pp. 9607-9619, 2023, doi: 10.1007/s12652-023-04633-6.

[13] S. Anam, M. Putra, Z. Fitriah, I. Yanti, N. Hidayat, and D. Mahanani, "Health claim insurance prediction using support vector machine with particle swarm optimization," BAREKENG: Jurnal Ilmu Matematika dan Terapan, vol. 17, no. 2, pp. 797-806, 2023, doi: 10.30598/barekengvol17iss2pp0797-0806.

[14] Q. Wang, M. Zhang, Y. Zhang, J. Zhong, and V. Sheng, "Location-based deep factorization machine model for service recommendation," Applied Intelligence, vol. 52, no. 9, pp. 9899-9918, 2022, doi: 10.1007/s10489-021-02998-9.

[15] R. Madani, A. Ez-Zahout, F. Omary, and A. Chedmi, "Advancing context-aware recommender systems: A deep context-based factorization machines approach," International Journal of Computing and Digital Systems, vol. 15, no. 1, pp. 353-363, 2024, doi: 10.12785/ijcds/160128.

[16] L. Chen and H. Shi, "DexDeepFM: ensemble diversity enhanced extreme deep factorization machine model," ACM Transactions on Knowledge Discovery from Data, vol. 16, no. 5, pp. 1-17, 2022, doi: 10.1145/3505272.

[17] H. Yu and J. Yin, "Deep reinforcement factorization machines: A deep reinforcement learning model with random exploration strategy and high deployment efficiency," Applied Sciences, vol. 12, no. 11, pp. 5314, 2022, doi: 10.3390/app12115314.

[18] L. Ren, T. Wang, Y. Laili, and L. Zhang, "A data-driven self-supervised LSTM-DeepFM model for industrial soft sensor," IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 5859-5869, 2022, doi: 10.1109/TII.2021.3131471.

[19] E. Merza and N. Mohammed, "Fast ways to detect outliers," Journal of Techniques, vol. 3, no. 1, pp. 66-73, 2021, doi: 10.51173/jt.v3i1.287.

[20] Q. Zhou and N. Zhou, "Neural network-enhanced pairwise bilinear factorization machines," Computer Engineering & Science, vol. 46, no. 9, pp. 1648-1659, 2024, [Online]. Available: http://joces.nudt.edu.cn/CN/.

[21] J. Quan and X. Sun, "Credit risk assessment using the factorization machine model with feature interactions," Humanities and Social Sciences Communications, pp. 11(1), 2024, doi: 10.1057/s41599-024-02700-7.

[22] T. Liu and N. Zhou, "Deep input-aware factorization machine based on setwise sorting," Computer Engineering & Science, vol. 45, no. 10, pp. 1891-1900, 2023, doi: 10.3969/j.issn.1007-130X.2023.10.020.

[23] I. Cohen Sabban, O. Lopez, and Y. Mercuzot, "Automatic analysis of insurance reports through deep neural networks to identify severe claims," Annals of Actuarial Science, vol. 16, no. 1, pp. 42-67, 2021, doi: 10.1017/S174849952100004X.

[24] M. Samiuddin, G. Rajender, K. Varma, A. Kumar, and S. Shaik, "Health insurance cost prediction using deep neural network," Asian Journal of Research in Computer Science, vol. 16, no. 2, pp. 46-53, 2023, doi: 10.9734/ajrcos/2023/v16i2338.

[25] F. Vandervorst, W. Verbeke, and T. Verdonck, "Claims fraud detection with uncertain labels," Advances in Data Analysis and Classification, vol. 18, no. 1, pp. 219-243, 2023, doi: 10.1007/s11634-023-00568-0.

[26] X. Zhang, "Optimal strategies for multi-insurance company competition under time-delay effects," Advances in Applied Mathematics, vol. 14, no. 5, pp. 476-489, 2025, doi: 10.12677/aam.2025.145276.

[27] X. Ma and J. Wang, "Robust inference using inverse probability weighting," Journal of the American Statistical Association, vol. 115, no. 532, pp. 1851-1860, 2019, doi: 10.1080/01621459.2019.1660173.

[28] V. Avagyan and S. Vansteelandt, "Stable inverse probability weighting estimation for longitudinal studies," Scandinavian Journal of Statistics, vol. 48, no. 3, pp. 1046-1067, 2021, doi: 10.1111/sjos.12542.

[29] Y. Xie and Z. Li, "Health insurance actuarial theory and methodology," Journal of Shandong University (Health Sciences), vol. 57, no. 8, pp. 20–38, 2022, doi: 10.6040/j.issn.1671-7554.0.2018.1016.

[30] Y. Gao, "Optimal reinsurance and investment problems with delay effects in a thinning dependent risk model," Advances in Applied Mathematics, vol. 13, no. 4, pp. 1523-1535, 2024, doi: 10.12677/aam.2024.134143.

[31] L. Settipalli, G. Gangadharan, and U. Fiore, "Predictive and adaptive drift analysis on decomposed healthcare claims using ART-based topological clustering," Information Processing & Management, vol. 59, no. 3, Art. no. 102887, 2022, doi: 10.1016/j.ipm.2022.102887.

[32] A. Gallego, J. Calvo-Zaragoza, and R. Fisher, "Incremental unsupervised domain-adversarial training of neural networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4864-4878, 2021, doi: 10.1109/TNNLS.2020.3025954.

[33] A. Ghosh, T. Schaaf, and M. Gormley, "Adafocal: Calibration-aware adaptive focal loss," Advances in Neural Information Processing Systems, vol. 35, no. 116, pp. 1583-1595, 2022, doi: 10.48550/arXiv.2211.11838.

[34] J. Singh, C. Beeche, Z. Shi, O. Beale, B. Rosin, J. Leader, and J. Pu, "Batch-balanced focal loss: A hybrid solution to class imbalance in deep learning," Journal of Medical Imaging, pp. 10(5), 2023, doi: 10.1117/1.JMI.10.5.051809.

[35] L. Yang and B. Shi, "Credit risk assessment model based on modified cross-entropy loss function with focal loss and empirical evidence," Chinese Journal of Management Science, vol. 30, no. 5, pp. 65-75, 2022, doi: 10.16381/j.cnki.issn1003-207x.2020.2188.