Expert Systems With Applications 298 (2026) 129583

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A multi-objective framework for predicting public opinion trends on infectious diseases using NSGA-II and interval predictions

Futian Weng [a,b], Meng Su [c,*], Petr Hajek [d], Mohammad Zoynul Abedin [e,*]

[a] School of Economics and Management, Fuzhou University, Fuzhou, 360108, China
[b] Data Mining Research Center, Xiamen University, Xiamen, 361005, China
[c] School of mathematics and statistics, Central south university, Changsha, 410083, China
[d] Faculty of Economics and Administration, University of Pardubice, Studentska 95, Pardubice, 53210, Czech Republic
[e] School of Management, Swansea University, Bay Campus, Fabian Way, SA1 8EN, Swansea, Wales, UK

ARTICLE INFO

Keywords:
Public opinion prediction
NSGA-II algorithm
Feature selection
Infectious diseases
Interval prediction

ABSTRACT

Predicting public opinion trends during major infectious disease outbreaks is critical for guiding effective public health responses. However, predicting public opinion remains challenging because it is influenced by socio-economic, psychological, and media factors. This paper presents a novel framework for predicting public opinion trends related to significant infectious diseases, with a focus on COVID-19 as a case study. The proposed framework identifies the key factors influencing public opinion development and enables both point and interval predictions. The framework uses information ecology theory and applies the NSGA-II algorithm to select the features that best drive public opinion trends. By incorporating this framework, accurate point forecasts are produced alongside prediction intervals, effectively quantifying the uncertainty inherent in public opinion dynamics. This approach minimizes the quality-driven loss function to generate precise prediction intervals, providing decision-makers with critical insights into public opinion fluctuations during epidemics. The results offer valuable, real-time public sentiment warnings, supporting timely and effective interventions in epidemic prevention and control efforts.

1. Introduction

In recent years, the world has witnessed an increase in public health emergencies, particularly large-scale outbreaks of infectious diseases such as COVID-19 (Mikulic & Baumgartner, 2025). These outbreaks pose serious threats to public health, economic stability and social order. During such crises, individuals increasingly rely on social media platforms such as Weibo to receive real-time information about the pandemic (Li et al., 2024; Liu et al., 2024). The widespread use of these platforms increases the reach and speed of information dissemination, potentially exacerbating extreme sentiments and crisis events (Kellner et al., 2023). To mitigate the negative impact of online public opinion during epidemics, it is essential to accurately predict trends in public sentiment and manage public emotions to promote constructive discourse. Real-time forecasting enables governments to respond quickly, implement policies effectively, and maintain public trust during crises (Gizelis & Karim, 2024; Yao et al., 2024).

Understanding the factors that shape online public opinion is key to predicting future sentiment trends (Kumar & Taylor, 2024; Subramanian et al., 2024). Several researchers have analyzed the variables that influence the evolution of public opinion during infectious disease outbreaks. For example, Liu and Fu (2022) found a correlation between the severity of an epidemic and public opinion risk, while Su et al. (2023) examined how public discussion and spatial factors influence opinion dynamics. Other studies have investigated the mechanisms underlying the spread of public discourse. Wang et al. (2022) applied information ecology theory to model public opinion as a dynamic system consisting of three elements: information, the informant, and the information environment. However, much of this prior research focuses on content analysis and evolutionary trends without systematically examining the various factors that influence the underlying mechanisms of public opinion dissemination during epidemics.

Extracting key features from various information sources is essential for building predictive models of public opinion. Feature selection is a

* Corresponding authors.
E-mail addresses: wengfutian@stu.xmu.edu.cn (F. Weng), sumeng@stu.xmu.edu.cn (M. Su), petr.hajek@upce.cz (P. Hajek), m.z.abedin@swansea.ac.uk (M.Z. Abedin).

https://doi.org/10.1016/j.eswa.2025.129583
Received 25 January 2025; Received in revised form 17 August 2025; Accepted 30 August 2025
Available online 16 September 2025

multi-objective optimization problem that aims to maximize model performance while selecting the optimal combination of features (Eshkiti et al., 2023). Among the available approaches, multi-objective genetic algorithms, such as NSGA-II, have proven to be effective in solving complex optimization problems, especially in time series feature selection (Espinosa et al., 2023; Jiménez et al., 2017).

Predicting trends in public opinion is inherently a time series forecasting problem. Scholars have made significant progress in this area by developing predictive models for public opinion related to infectious diseases. Mahdikhani (2022) predicted public interest on Twitter during the COVID-19 pandemic, and Xu et al. (2023) improved model accuracy by identifying factors that drive fluctuations in public opinion. However, most existing models focus on point predictions, which provide a single, deterministic value but fail to account for uncertainty in future opinion trends Du et al. (2022), Yan et al. (2022). This oversight is problematic because the uncertainty associated with public opinion can affect the precision of strategic responses, particularly in volatile epidemic situations. It is therefore critical to address this gap by incorporating uncertainty into public opinion forecasting through interval forecasting (Chen et al., 2024; Yan et al., 2024).

Interval forecasting provides a comprehensive approach to quantifying the uncertainty in public opinion trends, providing both upper and lower bounds on predictions (Nix & Weigend, 1994). This can be achieved through indirect methods, which assume specific error distributions, or direct methods, which model prediction intervals from the data itself without distributional assumptions (Simhayev et al., 2022; Xue et al., 2024). The latter has become increasingly popular due to its flexibility and efficiency. Recently, scholars have proposed models that simultaneously generate point estimates and prediction intervals without increasing computational complexity (Salem et al., 2020). Applying these methods to predict public opinion trends during infectious disease outbreaks can help decision-makers better anticipate shifts in public sentiment and take pre-emptive action.

In this study, we aim to advance the understanding of the spread of public opinion during infectious disease outbreaks by introducing a novel predictive framework. Our approach uses information ecology theory to identify key factors influencing public opinion, applies the NSGA-II algorithm for optimal feature selection, and uses a Quality Driven Plus (QD +) loss function for both point and interval predictions. The main contributions of this study are:

1. We establish a comprehensive framework, based on information ecology theory, to identify the key factors influencing the evolution of public opinion during major infectious disease outbreaks.
2. We implement the NSGA-II multi-objective evolutionary algorithm to select the optimal feature subsets, thereby improving the accuracy of predictions.
3. We develop a prediction framework that simultaneously produces point and interval predictions of public opinion trends using the QD + loss function, allowing for a more nfine-grained understanding of uncertainty.
4. We propose a versatile prediction framework suitable for different infectious disease scenarios, providing operational evidence for public opinion management during health crises.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature on public opinion prediction and feature selection methods. Section 3 introduces the proposed framework, detailing the feature selection process and the prediction model. Section 4 presents the data, and Section 5 describes the experimental setup and presents the results of the empirical evaluation. Section 6 discusses the results and limitations of the proposed model. Finally, Section 7 concludes the paper.

2. Literature review

2.1. Feature construction and selection

The spread of online discourse during major infectious disease outbreaks operates like a control system, influenced by multiple input and output variables, such as social network structures, public attention, and external factors (Zhang et al., 2018). Several studies have investigated the drivers of public opinion trends during infectious disease outbreaks. Liu and Fu (2022) developed a theoretical model linking the severity of an epidemic to increased public attention, revealing a strong correlation between event intensity and public interest. Chen and Zhang (2022) found that surges in event-related tweets during the COVID-19 pandemic were largely fueled by personal narratives and heightened media coverage, highlighting the media's critical role in shaping online discourse. Zhao et al. (2022) emphasized the influence of social network structures by incorporating user follower counts into their model, while Su et al. (2023) pointed to spatial factors, noting that regions hit harder by the epidemic exhibited more negative public sentiment. Complementing these findings, Zhu et al. (2020) showed that political and economic centers tend to draw greater social media attention during crises.

While insightful, existing studies often overlook integrating factors into a comprehensive framework for understanding public opinion spread. Information ecology theory offers a robust approach, analyzing interactions between information, informants, and environment. It shows online discourse as a dynamic community shaped by content, environment, and social interactions. Yet, its use in predicting outbreak-related opinion trends is limited. Applying this theory is crucial for understanding and forecasting sentiment (Wang et al., 2022).

In the context of time series prediction of sentiment trends, feature selection is essential for reducing data complexity, improving model performance, and enhancing interpretability. This paper focuses on feature selection methods most relevant to our study. Among them, filter-based methods offer computational efficiency by selecting features based on statistical criteria; however, they often overlook inter-feature dependencies and may underperform in capturing complex relationships. Wrapper-based methods, while more computationally intensive, evaluate feature subsets using a predictive model and tend to yield higher accuracy (Gonzalez-Vidal et al., 2019); Karasu et al., 2020; Ameur et al. (2024). Embedded methods integrate feature selection into the model training process itself, striking a balance between performance and efficiency.

Given the need for multi-objective optimization, which balances prediction accuracy with the number of features selected, multi-objective genetic algorithms (Jiménez et al., 2017); Kumar and Taylor, 2024, like NSGA-II, prove to be highly effective. Jimenez et al. (2020) successfully utilized NSGA-II to predict infection rates, combining it with performance metrics like RMSE and MAE. Similarly, Espinosa et al. (2023) introduced a hybrid approach that extended NSGA-II, enhancing both computational speed and generalization capabilities. These studies highlight the efficacy of NSGA-II in optimizing feature selection, particularly in datasets with public sentiment or epidemic-related data.

2.2. Interval prediction

Most existing research on public opinion trends focuses on point forecasts, which provide a specific future value but do not take uncertainty into account (Liu et al., 2022; Zhong et al., 2023). In contrast, interval forecasting provides a more comprehensive view by quantifying the uncertainty associated with the predicted values. Interval prediction methods generate both upper and lower bounds on predicted values, providing a clearer understanding of possible outcomes (Nardi & O'day, 2000). These methods are particularly valuable for modelling public opinion, where fluctuations and uncertainty are inherent due to the volatile nature of online discourse.

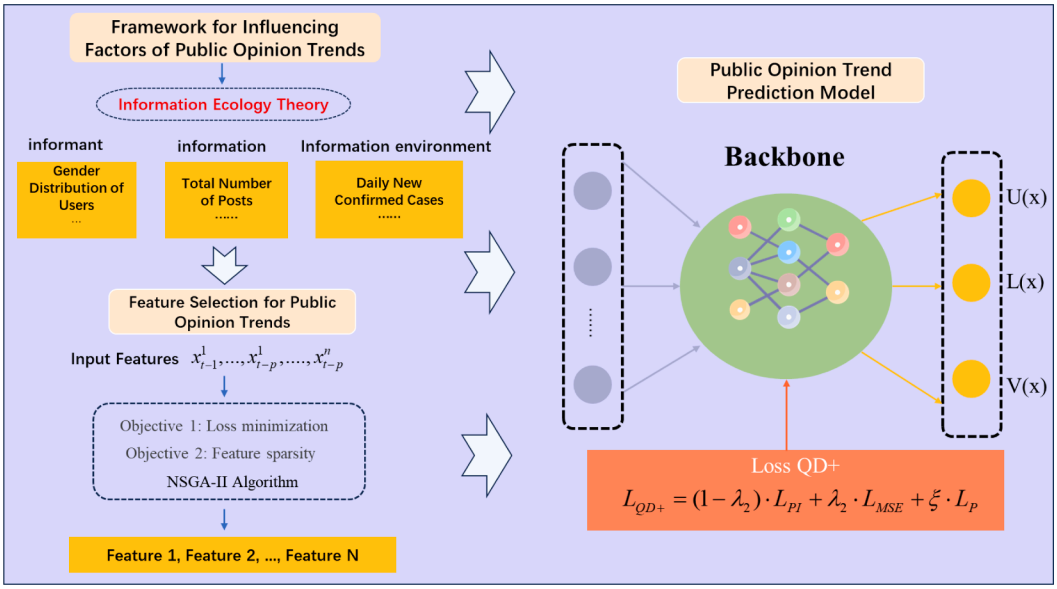


Fig. 1. Public opinion forecasting framework.

Two main sources of uncertainty in public sentiment trends are intrinsic randomness, which is due to unpredictable external factors, and subjective uncertainty, which arises from errors in data collection or model assumptions (Eldred et al., 2018). These uncertainties can be addressed using interval prediction methods, such as confidence intervals (CIs) and prediction intervals (PIs). CIs assume that data follow a specific probability distribution, allowing researchers to estimate the range within which the true value is likely to fall based on statistical analysis of the deviations between observed and true values (Sluijterman et al., 2024). However, PIs offer a more flexible alternative by predicting uncertainty directly through the model itself, making them particularly suitable for real-time public opinion forecasting (Simhayev et al., 2022).

Recent advances in artificial intelligence, particularly neural networks (NNs), have substantially improved interval prediction techniques. NNs for interval prediction are typically divided into two categories: indirect and direct methods. Indirect methods, such as those proposed by Veaux et al. (1998) and Bishop (1995), use standard regression models with constant error variance assumptions. Direct methods, such as the Minimum Volume Ellipsoid (MVE) proposed by Nix and Weigend (1994), use two NN models to estimate the upper and lower bounds, improving computational efficiency while maintaining accuracy. In addition, bootstrap-based methods (Heskes, 1996) have been used to resample data and generate robust interval predictions.

More recent approaches, such as the LUBE method (Khosravi et al., 2011), have minimized objective functions to directly estimate prediction intervals. Building on this, Pearce et al. (2018) introduced a quality-driven (QD) loss function that allows for gradient-based optimization, simplifying the training process while achieving accurate interval predictions. Salem et al. (2020)) further refined this approach by integrating point and interval predictions into a multi-objective loss function. Simhayev et al. (2022) extended these methods by selecting weights for the combination of upper and lower bounds, eliminating some of the limitations found in earlier models and improving overall prediction performance.

2.3. Research gap

While there has been extensive research on feature selection and time series forecasting, most studies have focused on optimizing models for point forecasting and have not sufficiently addressed the uncertainty inherent in public opinion trends during major infectious disease outbreaks. Furthermore, many studies do not incorporate relevant public opinion variables into their prediction frameworks. This paper seeks to fill these gaps by introducing a comprehensive approach that integrates NSGA-II for feature selection and a QD + loss function for both point and interval predictions. Our model not only captures the underlying uncertainty in public opinion trends, but also provides real-time predictions that can be applied to various infectious disease outbreaks, providing operational insights for public health decision making.

3. Design of interval prediction framework

Fig. 1 presents the overall structure of the proposed framework for public opinion forecasting, which consists of three primary components: (1) feature construction, (2) feature selection using NSGA-II, and (3) interval prediction leveraging deep learning models with the Quality Driven Plus (QD +) loss function. In the first component, we use information ecology theory to select variables related to sentiment intensity from three aspects: the informant, the information environment, and the information itself. To validate the effectiveness of the second component, we not only performed a detailed comparison with other wrapper-based feature selection approaches, but also compared different lag periods, including 1, 3, 5 and 7 days (Weng et al., 2022). Finally, we used deep learning with the QD + framework to predict both sentiment values and their intervals for infectious diseases.

3.1. Feature construction

We constructed features from sentiment metrics (e.g., Baidu Index), informant demographics, and real-time epidemic data. This feature construction process ensures that we capture the multidimensional aspects of public opinion and its development over time, which is crucial for accurate prediction.

The Baidu Index is a big data analytics tool developed by Baidu, China's largest search engine, based on the search behaviors of a vast number of internet users. By tracking the frequency and trends of specific keyword searches on the Baidu platform, it objectively reflects the public's level of attention toward particular events or topics. Since its official launch in 2006, the Baidu Index has become a key metric for measuring online public opinion and sentiment in China, known for its real-time responsiveness, broad coverage, and high representativeness. It has been widely used in various research fields, including public health and socio-economic studies (Mikulic and Baumgartner, 2025).

Table 1
Framework of influencing factors.

Feature category	Feature	Values (if categorical)
Predicted popularity of public opinion	Baidu Index	–
Informant	Distribution of various user types	[0, 1, 2, 3, 4]
	Gender distribution of users	[female, male]
	Age distribution of users	[youth, youth_adult, middle, older]
	City of residence distribution	[huadong, huazhong, huabei, huanan, dongbei, xinan, xibei, haiwai]
	Number of followers of users	–
Information	Total number of posts	–
	Number of likes	–
	Number of shares	–
	Number of comments	–
Information environment	Daily new confirmed cases	–
	Daily new death cases	–
	Daily new suspected cases	–
	Daily new recovered cases	–
	Sentiment of posts	[positive, negative, happy, neutral, angry, sad, fear, surprise]

In this study, we used the keyword 'COVID-19 pandemic' to derive sentiment intensity, which serves as the dependent variable in our analysis. The final set of constructed features is presented in Table 1.

Following the structure of information ecology theory (Luo et al., 2023; Yang et al., 2025), features were organized into three categories:

- Informant features, including demographic distributions (e.g., gender, age groups, city of residence) and user activity metrics (e.g., number of followers).
- Information features, such as the number of posts, likes, comments, and reposts.
- Information environment features, including daily new confirmed COVID-19 cases, daily new deaths, and emotion categories of user posts (e.g., positive, angry, fearful).

This structured feature design enables the model to consider the complex interactions between individuals, information flow, and external epidemic developments that drive changes in public opinion.

3.2. Feature selection

3.2.1. NSGA-II algorithm

Among evolutionary algorithms, the NSGA-II (Non-dominated Sorting Genetic Algorithm II) is a prominent approach for multi-objective optimization problems. NSGA-II was developed to address three primary issues in the original NSGA algorithm: (1) excessive computational time required for calculating dominance relationships under multiple objectives, (2) lack of an elitist strategy, and (3) reliance on a sharing mechanism for population diversity, which overly depends on subjectively determined parameters. NSGA-II overcomes these challenges by introducing a fast-sorting method to reduce algorithmic complexity, incorporating an elitist preservation strategy to prevent the loss of high-quality solutions during evolution, and employing crowding distance as a measure of diversity. In cases where individuals belong to the same non-dominated front, those with larger crowding distances are prioritized, enhancing solution diversity.

The core steps of the NSGA-II algorithm are as follows:

1. Initialization: Generate an initial population P_0 of size N and compute the fitness values for each individual based on multiple objectives. Perform fast non-dominated sorting to assign each individual to a non-dominated rank.
2. Selection, crossover, and mutation: Use tournament selection, reproduction, and mutation operators to generate an offspring population P_1.
3. Merging and sorting: Merge populations P_0 and P_1 to form a combined population of size $2N$. Perform fast non-dominated sorting on the combined population and calculate the crowding distance for individuals in the same non-dominated front. Retain individuals with higher crowding distances until the new population P_2 reaches size N.
4. Termination: Repeat the process until the stopping criteria, such as a maximum number of iterations, are met.

3.2.2. Feature selection process

Feature selection is formulated as a multi-objective optimization problem, with two objectives: minimizing the root mean square error (RMSE) of time series predictions and minimizing the number of selected features. Given a dataset with n samples, each containing m attributes, denoted as $\{x_1, x_2, ..., x_m\}$, and a target value y, we split the dataset into an 80 % training set and a 20 % testing set. The training set is used for feature selection, and to ensure robustness, we perform 10-fold cross-validation during this process. The specific optimization objective functions are defined as follows:

$$\text{Minimize } f_1(x) = \frac{1}{K}\sum_{k=1}^{K} F_k^{\phi}(x) \tag{1}$$

$$\text{Minimize } f_2(x) = \sum_{i=1}^{W} N(x_i) \tag{2}$$

where $x = \{x_1, x_2, \ldots, x_w\}$ is a boolean decision set, with $x_i \in \{\text{True}, \text{False}\}$, $i = 1, 2, \ldots, w$. In Eqs. (1) and (2), x_i = True represents the selection of the i-th feature, and x_i = False indicates that the feature is not included in the subsequent model prediction process. $F_k^{\phi}(x)$, $k = 1, 2, \ldots, K$, represents the performance of the feature selection model ϕ on the dataset x in the cross-validation fold. RMSE is used to evaluate the model's performance, while the sum of z_j represents the number of selected features. Random forest is used as the predictive model for feature selection in this study (Espinosa et al., 2022).

NSGA-II produces a set of non-dominated solutions (Pareto front) rather than a single optimal solution. Thus, further evaluation of these solutions is necessary to identify the best feature subset. The best solution is selected by evaluating the performance on the testing dataset using RMSE. Additionally, we compare NSGA-II's results with other multi-objective evolutionary algorithms, including the Multi-Objective Evolutionary Differential Algorithm (MOEDA) and Multi-Objective Particle Swarm Optimization (MOPSO). We also compare NSGA-II with non-evolutionary wrapper-based feature selection methods such as Recur-

sive Feature Elimination (RFE) (Wang et al., 2022), which iteratively removes the least important features based on model performance.

3.2.3. Evaluation metrics

After applying NSGA-II, we evaluate the resulting feature sets using RMSE, Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) on the testing dataset, which was not involved in the training process. These evaluation metrics are defined as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{4}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

To determine the best feature combination, we rank each feature subset based on RMSE, MAPE, and MAE. The feature subset with the lowest total rank across all three metrics is selected as the optimal subset for further analysis and model application Ferchichi et al. (2024). The final algorithmic process is outlined in Algorithm 1 and Algorithm 2 below.

Algorithm 1: Feature Selection.

Require:
$x = \{x_1, x_2, \ldots, x_w\}$: Individuals with a feature count of w
$D_{\text{train}}, D_{\text{test}} \subset D$; $D_k \subset D_{\text{train}}$ for $k = 1, 2, \ldots, K$. Split the data into training set D_{train} and testing set D_{test} in an 8:2 ratio. D_{train} is used for feature selection, and D_{test} for subsequent ranking of feature sets;
$O_1 \leftarrow 0$;
for $k = 1$ **to** K **do**
  Featureset $\leftarrow \{\}$;
  for $i = 1$ **to** w **do**
    if $x_i \neq 0$ **then**
      Featureset $\leftarrow$ Featureset $\cup \{x_i\}$;
  Construct D'_k based on the Featureset;
  $O_1 \leftarrow O_1 + \text{RMSE}(M_k^{\phi})$ // RMSE of the prediction model M_k^{ϕ} evaluated on dataset D'_k;
$O_1 \leftarrow \frac{O_1}{K}$;

3.3. Interval prediction model

This study builds on prior research (Salem et al., 2020; Simhayev et al., 2022) by implementing a deep learning-based framework for both point and interval predictions of public opinion trends associated with major infectious diseases. The performance of point prediction is evaluated using common error metrics such as RMSE, MAPE, and MAE. For interval predictions, the quality of the generated prediction intervals is assessed using the Probability of Interval Coverage (PICP) and the Mean Prediction Interval Width (MPIW). This comprehensive evaluation, employing a variety of metrics, provides a robust scientific basis for analyzing public opinion trends, demonstrating the effectiveness and accuracy of the proposed framework.

The PICP is calculated as follows:

$$\text{PICP} = \frac{1}{n}\sum_{i=1}^{n} k_i \tag{6}$$

where n is the sample size and k_i is a binary variable that equals 1 if the actual sample value y_i lies within the prediction interval $[L_i, U_i]$ and 0 otherwise.

Algorithm 2: Ranking Feature Set.

Require:
$D_{\text{train}}, D_{\text{test}}$
F^{ϕ}
opti_Featureset = $\{\text{Featureset}_1, \ldots, \text{Featureset}_n\}$
RMSE = [];
MAE = [];
MAPE = [];
foreach f_i ***in*** *opti_Featureset* **do**
  $D_{\text{train}} \rightarrow D_{f_i}^{\text{train}}$ based on Featureset_i;
  Obtain model $M_{f_i}^{\phi}$ trained on the training set $D_{f_i}^{\text{train}}$;
  RMSE $\leftarrow$ RMSE($M_{f_i}^{\phi}$) of the prediction model $M_{f_i}^{\phi}$ evaluated on $D_{f_i}^{\text{test}}$;
  MAE $\leftarrow$ MAE($M_{f_i}^{\phi}$) of the prediction model $M_{f_i}^{\phi}$ evaluated on $D_{f_i}^{\text{test}}$;
  MAPE $\leftarrow$ MAPE($M_{f_i}^{\phi}$) of the prediction model $M_{f_i}^{\phi}$ evaluated on $D_{f_i}^{\text{test}}$;
RMSE_{rank} = [RMSE^1_{rank}, …, RMSE^n_{rank}];
MAE_{rank} = [MAE^1_{rank}, …, MAE^n_{rank}];
MAPE_{rank} = [MAPE^1_{rank}, …, MAPE^n_{rank}];
RESULT = Mean(RMSE_{rank} + MAE_{rank} + MAPE_{rank});

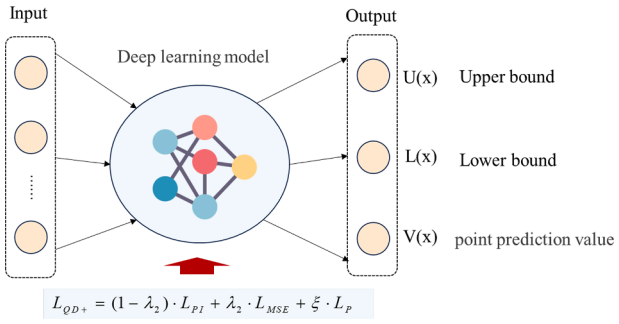


Fig. 2. Public opinion trend prediction model.

The MPIW is given by:

$$\text{MPIW} = \frac{1}{n}\sum_{i=1}^{n}(U_i - L_i) \tag{7}$$

where U_i and L_i represent the upper and lower bounds of the predicted interval for the i-th sample, respectively.

The interval prediction model based on neural networks aims to minimize the prediction interval width while maintaining an acceptable level of PICP. Most current methods for generating prediction intervals do not simultaneously provide both point estimates and prediction intervals. However, the proposed prediction framework, which utilizes the QD+ loss function, addresses this limitation by enabling both interval and point predictions without assuming any specific data distribution (Salem et al., 2020). The conceptual framework of this model is illustrated in Fig. 2. The left side represents the model inputs, while the right side shows the model's interval predictions ($L(x)$ and $U(x)$) and point prediction ($V(x)$) values. The middle section consists of a deep learning model that uses the L_{QD+} as its loss function.

3.4. Framework overview

The proposed framework consists of two main components. The first component is the selected time series prediction model, where commonly used sequence prediction algorithms such as Recurrent Neural

Networks (RNN) or Long Short-Term Memory (LSTM) networks can be employed for predicting public opinion trends. The second component is an auxiliary network responsible for generating both interval and point predictions. This network outputs three values: the upper bound $U(x)$, the lower bound $L(x)$, and the point prediction $\hat{y}$.

In the proposed model, the QD+ loss function is designed to train neural networks to produce both, a precise point prediction (e.g., sentiment index = 500), and a reliable prediction interval (e.g., between 480 and 520). The goal of QD+ is to balance two key objectives, accuracy and uncertainty estimation (the predicted interval should contain the actual value with high probability (coverage), but also be as narrow as possible (precision)). QD+ achieves this by combining several loss components: (a) an MSE term to ensure the point prediction is accurate, (b) a penalty if the true value falls outside the prediction interval, and (c) a regularization that tries to keep the width of the prediction interval small. This way, the model learns not only to predict well but also to express how confident it is in those predictions-making it especially useful for volatile situations like epidemics, where public opinion can change quickly and unpredictably.

In the interval prediction component of the proposed model, the overall objective is to minimize the MPIW while maximizing the PICP. To balance these two objectives, we define the following loss function:

$$L_{\text{PI}} = (1 - \lambda_1) \cdot L_{\text{MPIW}} + \lambda_1 \cdot L_{\text{PICP}} \tag{8}$$

where

$$L_{\text{MPIW}} = \frac{1}{c} \sum_{i=1}^{n} (U_i - L_i) \cdot k_i \tag{9}$$

$$L_{\text{PICP}} = \max\left(0, (1 - \alpha - \text{PICP})^2\right) \tag{10}$$

Here, n denotes the sample size, and λ_1 is a hyperparameter that balances the trade-off between the interval width and coverage rate. The objective is to minimize L_{MPIW} under the constraint that PICP $\geq 1 - \alpha$, where L_{MPIW} differs from MPIW as it only includes valid observations within the Prediction Interval (PI). The term L_{PICP} optimizes the interval coverage rate, with its loss increasing when PICP falls below $1 - \alpha$.

In the actual optimization process, since k_i is a discrete value, its loss function is difficult to converge. Therefore, we redefine k_i as a continuous variable:

$$k_{i-\text{soft}} = \sigma(s \cdot (y_i - L_i)) \cdot \sigma(s \cdot (U_i - y_i)) \tag{11}$$

where σ is the sigmoid function and $s > 0$ represents the softening factor. The continuous version of $k_{i-\text{soft}}$ is then substituted into Eq. (9) to complete the optimization for the interval prediction.

However, the objective function in Eq. (8) only focuses on optimizing the upper and lower bounds of the interval. To optimize point prediction values, additional loss functions are introduced. For this purpose, we define:

$$L_v = L_{\text{MSE}} + \xi \cdot L_P \tag{12}$$

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}^{(i)} - y^{(i)}\right)^2 \tag{13}$$

$$L_P = \frac{1}{n} \sum_{i=1}^{n} \left[\max\left(0, \hat{y}_L^{(i)} - \hat{y}^{(i)}\right) + \max\left(0, \hat{y}^{(i)} - \hat{y}_U^{(i)}\right)\right] \tag{14}$$

In this architecture, L_{MSE} optimizes the point prediction values, while L_P ensures that the point predictions stay within the interval bounds. The final objective function for the training process can be formalized as:

$$L_{\text{QD}+} = (1 - \lambda_2) \cdot L_{\text{PI}} + \lambda_2 \cdot L_{\text{MSE}} + \xi \cdot L_P \tag{15}$$

where λ_2 is a hyperparameter that balances the trade-off between interval predictions and point predictions. By adjusting this hyperparameter, the model can control the range and accuracy of interval predictions.

Ensemble models are employed to further reduce uncertainty in the prediction process while enhancing the performance and stability of the prediction model. Multiple models are integrated to generate the final point and interval predictions. Each ensemble model produces a PI and a point prediction, which are then aggregated to reflect the uncertainty in the results. Following the approach of Lakshminarayanan et al. (2017), the ensemble framework generates more stable outputs through the following steps.

Assume there are m models trained with the $L_{\text{QD}+}$ loss function. Each model M_j produces a prediction interval $[\hat{y}_L^{ij}, \hat{y}_U^{ij}]$ and a point prediction $\hat{y}^{ij}$ for data point y^i. The final interval and point prediction results are obtained by aggregating the predictions from all models:

$$\mu_L^i = \frac{1}{m} \sum_{j=1}^{m} \hat{y}_L^{ij} \tag{16}$$

$$\mu_U^i = \frac{1}{m} \sum_{j=1}^{m} \hat{y}_U^{ij} \tag{17}$$

$$\hat{y}_v^i = \frac{1}{m} \sum_{j=1}^{m} \hat{y}^{ij} \tag{18}$$

$$\sigma_L^i = \frac{1}{m-1} \sum_{j=1}^{m} \left(\hat{y}_L^{ij} - \mu_L^i\right)^2 \tag{19}$$

$$\sigma_U^i = \frac{1}{m-1} \sum_{j=1}^{m} \left(\hat{y}_U^{ij} - \mu_U^i\right)^2 \tag{20}$$

Finally, the aggregated prediction interval is computed as:

$$\hat{y}_U^i = \mu_U^i + z_{\alpha/2} \cdot \sigma_U^i \tag{21}$$

$$\hat{y}_L^i = \mu_L^i - z_{\alpha/2} \cdot \sigma_L^i \tag{22}$$

Here, $[\hat{y}_L^i, \hat{y}_U^i]$ and $\hat{y}_v^i$ represent the final interval prediction and point prediction for data point y^i.

4. Data description and processing

The data used in this study can be categorized into four main components. First, we utilize the Baidu Index to describe trends in public opinion on major infectious diseases. The Baidu Index[1] is a data analytics platform that leverages Baidu search data, reflecting the objective demands of the search population through user search behavior and the search volume of specific keywords (Zhang et al., 2022). The Baidu Index can be used to study and predict public opinion trends on related topics, making it an appropriate indicator for measuring the overall development of public opinion on major infectious diseases.

Second, we perform web scraping of blog posts using Python, focusing on posts related to the epidemic that were published by users between January 2020 and February 2023. The crawling frequency is set to once per hour, capturing data such as usernames, likes, comments, reposts, and basic user information. In total, 12,082,334 records were gathered.

The third component of the data concerns content related to the COVID-19 epidemic, primarily including daily reports on newly confirmed cases, daily deaths, newly suspected cases, and newly recovered cases in China.

The fourth data component pertains to public sentiment types. Following the removal of punctuation and extraneous characters from the textual corpus, we employed a pre-trained BERT model to quantify the distribution of public emotional states. we employ a pre-trained BERT model to calculate the distribution of public sentiment (Su et al., 2023). The integration of these four data sources provides multi-level and multi-perspective information, which enables a comprehensive understanding of the evolution of public opinion trends.

Based on this data, and drawing from relevant literature (Luo et al., 2023), this study constructs an influencing factor system for the popularity of public opinion on major infectious diseases. This system is

[1] http://index.baidu.com

Table 2
Results of feature selection.

Method	LAG	Feature num	RMSE	MAPE	MAE
REF	1	3/38	23738.17	**43.86**	**45161.26**
	3	12/114	23260.25	48.20	47502.94
	5	58/190	**22933.74**	46.81	46802.71
	7	93/266	24105.36	46.55	45639.10
MOEDA	1	17/38	25593.40	43.68	**43566.65**
	3	58/114	24760.34	46.65	46021.46
	5	94/190	24580.19	**43.07**	45113.37
	7	135/266	**23084.85**	43.57	43895.32
MOPSO	1	6/38	28927.56	45.60	45165.80
	3	45/114	24939.51	45.05	44703.40
	5	66/190	**23004.07**	**43.02**	44692.35
	7	121/266	23027.53	43.84	**43863.50**
NSGA2	1	7/38	23832.55	41.63	43028.72
	3	20/114	22818.64	**41.23***	**42822.68***
	5	44/190	**22545.33***	45.13	45965.33
	7	92/266	22627.74	43.75	43953.37

Table 3
DM test results of different feature selection methods.

Method	LAG	DM	P value
REF	1	–3.23	<0.001
	3	–2.69	<0.01
	5	–2.00	<0.05
	7	–3.62	<0.001
MOEDA	1	–8.20	<0.001
	3	–6.59	<0.001
	5	–4.34	<0.001
	7	–3.65	<0.001
MOPSO	1	–6.49	<0.001
	3	–5.33	<0.001
	5	–4.54	<0.001
	7	–2.13	<0.05

based on the theory of information ecology and includes three aspects: informant, information environment, and information (see Table 1).

When using the above data as input for subsequent deep learning models, data normalization to the interval $[0, 1]$ was performed to accelerate the convergence of NNs and expedite the discovery of globally optimal solutions.

5. Result analysis

5.1. Feature selection and analysis

In this section, we first compare the feature selection results of REF, NSGA2, MOEDA, and MOPSO under different lag periods, and qualitatively analyze the feature selection process under each lag period. Finally, in conjunction with the information ecology theory, we analyze the feature set under the optimal lag period from the perspectives of informant, information, and information environment.

In time series prediction problems, the choice of lag period is crucial. To determine the optimal lag period, we evaluated the feature selection and model prediction performance for lag periods of 1, 3, 5, and 7. The feature selection results of the NSGA2, MOEDA, MOPSO, and REF algorithms under different lag periods are summarized in Table 2. Furthermore, to comprehensively validate the superiority of the proposed methodology, we conducted Diebold-Mariano tests (presented in Table 3), the results of which provide additional empirical evidence substantiating its significant advantage in feature selection.

Overall, MOEDA and MOPSO show an increasing trend in scores for certain indicators as the lag period increases, while REF and NSGA2 maintain relatively stable scores across the indicators. The optimal values for each indicator under each method framework are highlighted in bold, and the best indicators across all frameworks are marked with an asterisk. From the comparative study of lag period selection, we observed that the lag period of 3 produced the most significant proportion of optimal performance indicators. Consequently, we identified the feature subset corresponding to a lag period of 3 as the optimal feature combination for subsequent analytical procedures, with detailed results presented in Table 4. The numerical suffix denotes the temporal lag structure: 0 indicates a one-day lag, with subsequent integers representing incrementally longer lag periods.

Therefore, we selected the feature subset obtained when the lag period equals 3 as the key feature combination for subsequent analysis (see Table 4).

A core focus of this study is to explore the key variables influencing the spread of public opinion during major infectious disease outbreaks. To achieve a comprehensive understanding of the intrinsic mechanisms of the NSGA2 algorithm in feature selection, we conducted detailed investigations. Initially, we examined how the multi-objective evolutionary algorithm NSGA2 performs feature selection during each iteration under different lag periods. Specifically, we calculated the frequency with which each feature was selected during the iteration process. To visually observe the evolutionary results, we plotted the frequency of feature selection and analyzed the Pareto solution sets under different lag periods to reveal the performance and variations of features within these solution sets.

Based on information ecology theory, we analyzed the final feature set from multiple dimensions. When the lag period is set to 1, variables with higher selection frequencies during the iteration process include the number of posts from northwestern provinces, the number of posts from eastern provinces, Baidu search index, and daily new confirmed cases of COVID-19 (see Fig. 3). However, the final selected variables under a lag period of 1 are baidu_index_0 and Daily_new_confirmed_cases_0, indicating a closer association between the current hot topics in public opinion and the discussion intensity of the previous day, as well as the daily new confirmed cases of COVID-19 within a relatively short timeframe.

When the lag period is set to 3, analysis of the frequency of feature selection during the iteration process reveals that features with higher selection frequencies are concentrated towards the end of the lag period (see Fig. 4). Specifically, compared to features lagged by one day, variables lagged by two and three days are more likely to be selected. Additionally, features with higher occurrence frequencies include baidu_index_2, fans_num_2, like_num_1, and huadong_1.

When the lag period is set to 5, the distribution of feature occurrences across different time periods appears to be relatively uniform, especially within the higher frequency feature sets (see Fig. 5). Among them, features with higher occurrence frequencies include baidu_index_4, netural_2, user_ver_3_0, like_num_2, and dong_bei_4, marking the first appearance of factors related to sentiment within the top-ranking features.

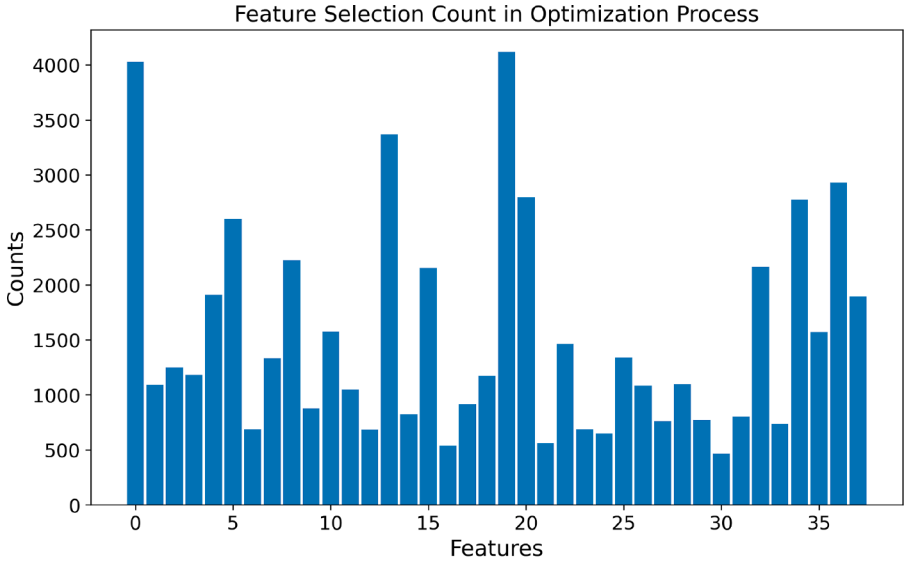
When the lag period is set to 7, the frequency distribution of features appears similar to that of the lag period 5 (see Fig. 6). However, due to the increased number of features, there is a similarity in the distribution of occurrences among the top twenty features. Unlike in previous lag periods, the top-ranking feature is not the Baidu search index itself, but rather a variable related to user factors, user_ver_3_3. Other highly ranked features include haiwai_5, repost_num_0, female_num_0, and fear_3. Notably, factors related to COVID-19 do not appear among the top-ranked features, indicating that over a longer data cycle, public opinion tends to focus more on variables related to the general public rather than COVID-related factors.

Overall, in shorter lag periods, variables closer to the current day are more likely to be selected, while as the lag period increases, the probability of feature selection becomes more evenly distributed across different time stages. Additionally, throughout the iteration process, variables related to geographical regions and COVID-19 tend to have higher occurrence frequencies compared to other variables.

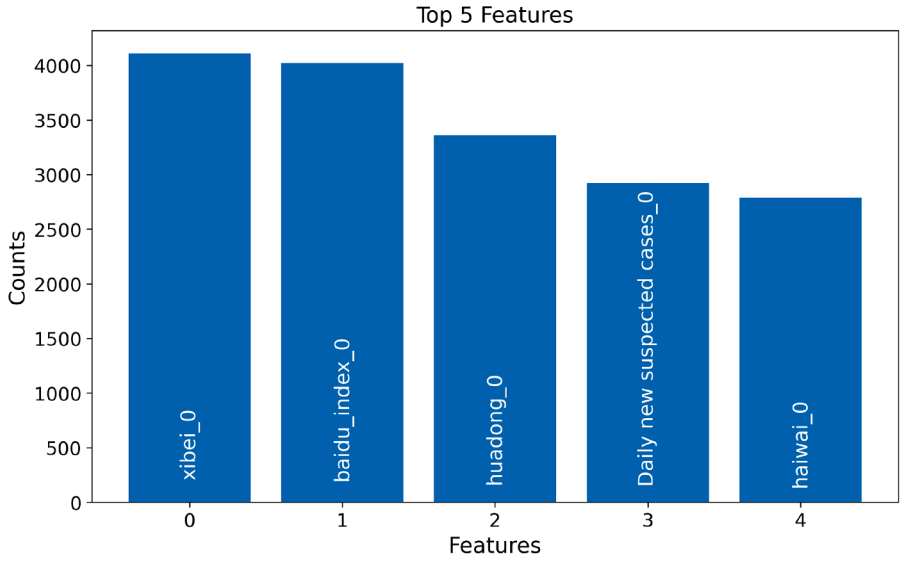
Table 4
Optimal feature combination selected by the NSGA-II algorithm (lag = 3).

Original features	Selected features
baidu_index, user_ver_0, user_ver_1, user_ver_2, user_ver_3, user_ver_4, female_num, male_num, kong_num, youth_num, youth_adult_num, middle_num, older_num, huadong, huazhong, huabei, huanan, dongbei, xinan, xibei, haiwai, positive, negative, happy, neutral, angry, sad, fear, surprise, total_post_num, like_num, repost_num, review_num, fans_num, daily new confirmed cases, daily new deaths, daily new suspected cases, daily new number of recoveries	baidu_index_2, user_ver_1_2, user_ver_4_2, older_num_1, huadong_0, huadong_1, xibei_2, happy_2, angry_2, sad_0, surprise_2, total_post_num_1, total_post_num_2, like_num_1, review_num_0, fans_num_0, daily new confirmed cases_2, daily new deaths_2, daily new suspected cases_0, daily new number of recoveries_2

Note: the numerical suffix denotes the temporal lag structure: 0 indicates a one-day lag, with subsequent integers representing incrementally longer lag periods.



(a) Feature count

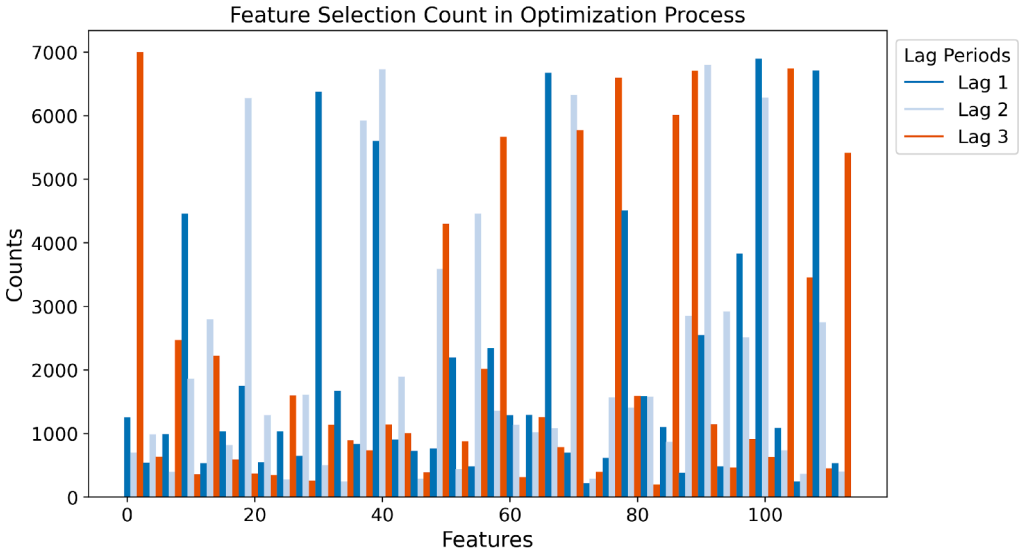


(b) Top 5 features

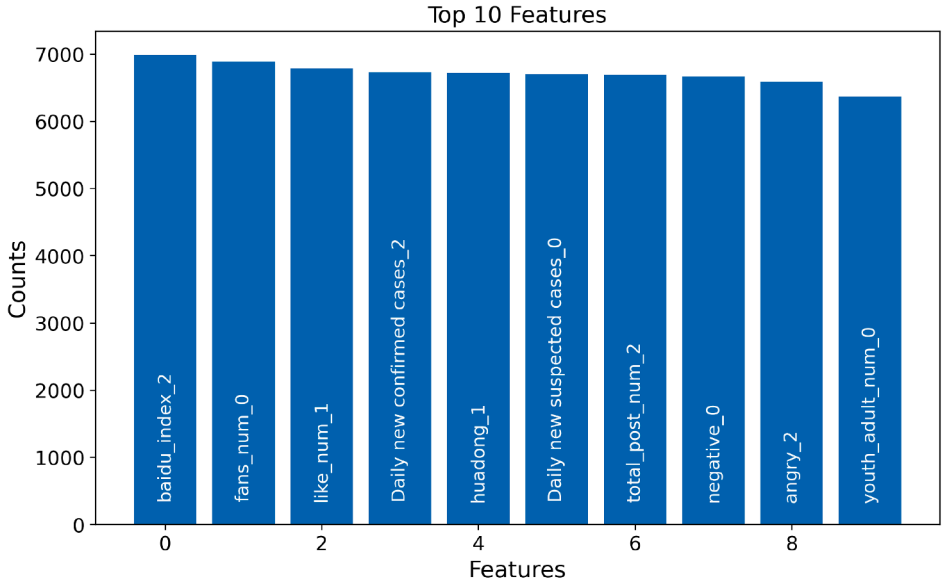
Fig. 3. Feature selection process (lag = 1). Legend: the names (and values) of top 5 features can be found in Table 1, and their numerical suffix denotes the temporal lag structure.

Based on feature evolution across different lag periods, we draw three key conclusions from an information ecology perspective:

1. Information sources: Indicators such as the total number of posts and the number of comments have relatively lower weights in the feature selection process compared to other variables. However, as the lag period increases, metrics such as repost count and number of likes begin to rank higher in the frequency of feature selection. Despite this, these metrics still have lower weights in the final selected feature set, suggesting that while variables related to information sources are somewhat associated with public opinion trends, they



(a) Feature count



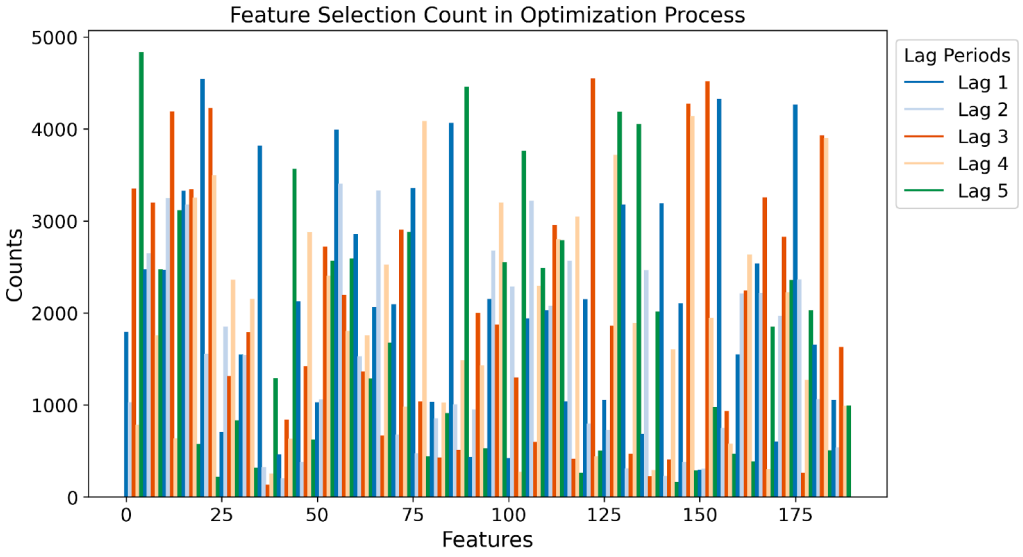
(b) Top 5 features

Fig. 4. Feature selection process (lag = 3). Legend: the names (and values) of top 5 features can be found in Table 1, and their numerical suffix denotes the temporal lag structure.

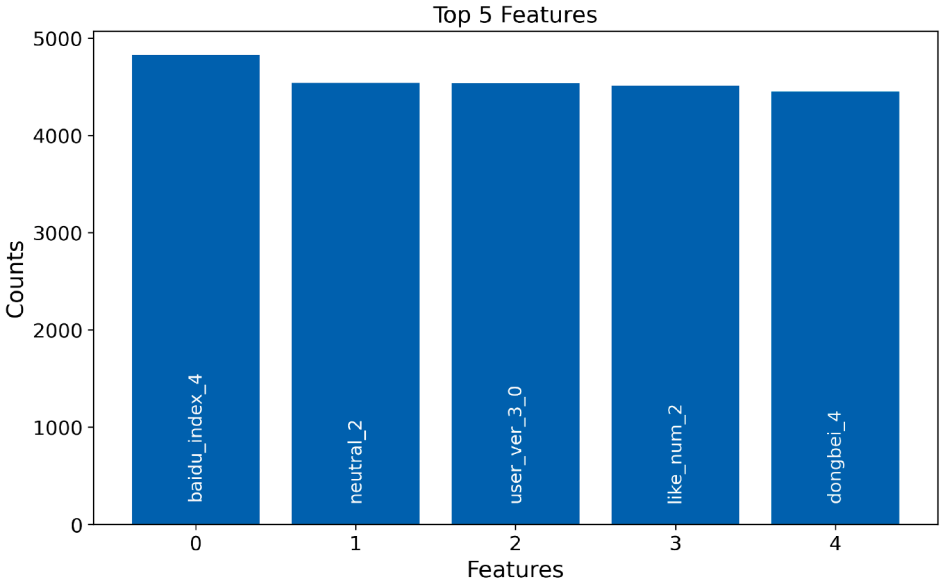
are not the primary driving factors in the final predictive models of public opinion trends.

2. Informant: As the lag period increases, the weight of variables related to the distribution of various user types continuously increases. Variables associated with age distribution and geographical distribution of users also show a similar trend. For age distribution, variables associated with middle-aged populations have a higher proportion compared to adolescents and the elderly. In terms of geographical distribution, variables representing East, Central, and North China are included in varying degrees across different lag periods. These regions share common characteristics of relatively developed economic levels and high population densities, indicating a close association between the development of public opinion trends and the discussion level in such regions. Other variables related to informants, such as gender distribution and number of followers, appear less frequently, suggesting that during the epidemic period, age and geographical characteristics of users play a more prominent role in shaping public opinion trends.
3. Information environment: The results indicate that with an increase in the lag period, the proportion of factors associated with negative sentiments also increases. Factors related to the development of the COVID-19 pandemic are present in the final feature sets across various lag periods. The Appendix A shows that the period close to the prediction day of new confirmed COVID-19 cases consistently appears in the final feature sets across different lag periods, indicating the significant impact of recent COVID-19 news on public opinion trends.

In summary, variables related to informants and the information environment significantly influence the prediction of public opinion trends. Among the informant variables, features associated with the public exert varying degrees of influence on the development and changes



(a) Feature count



(b) Top 5 features

Fig. 5. Feature selection process (lag = 5). Legend: the names (and values) of top 5 features can be found in Table 1, and their numerical suffix denotes the temporal lag structure.

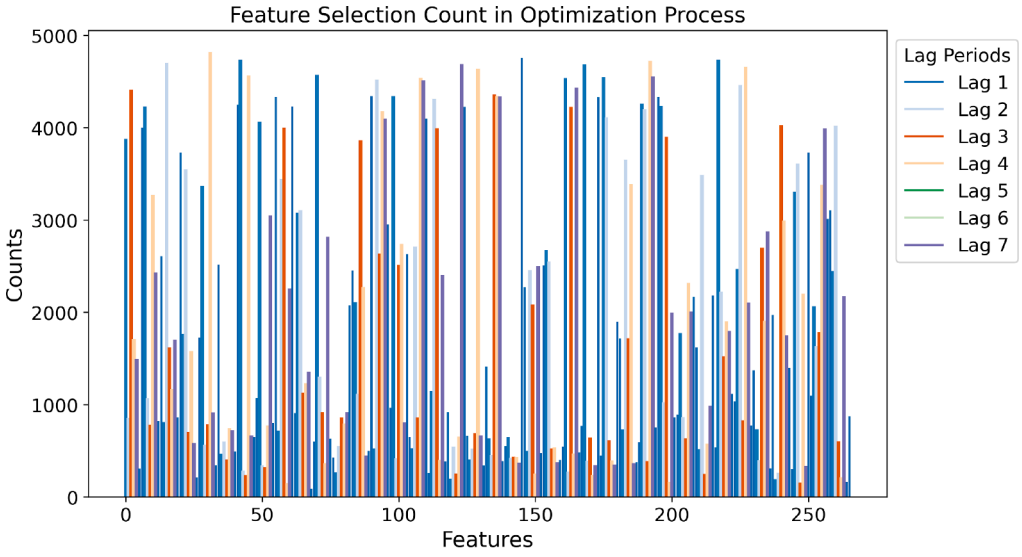
in public opinion. For instance, users from the East China region frequently appear in feature selection processes. Regarding information environment variables, the level of attention to pandemic dynamics plays a crucial role in predicting public opinion trends. This has practical implications for public opinion management and forecasting during the pandemic, emphasizing the need to focus on regional factors and public sentiment in public opinion alert systems. Conversely, variables related to information, such as the total number of posts and the number of likes, have a lower proportion in the feature selection process and final feature set. This suggests that during the pandemic, public opinion is more directly influenced by factors related to the public itself and the development of the pandemic, rather than the dissemination of information. Therefore, the influence of information itself appears to be relatively limited.

5.2. Performance of interval prediction

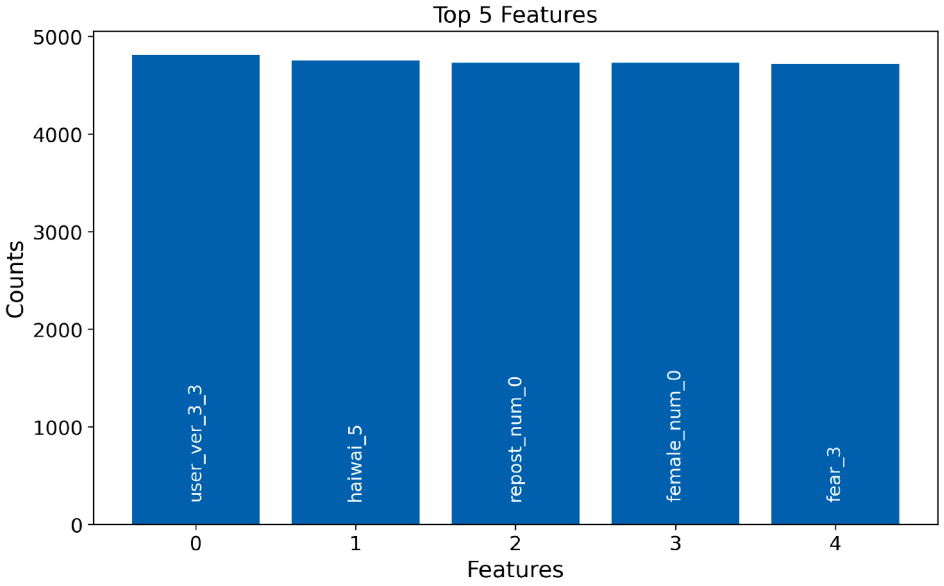
5.2.1. Experiment setting

This section proposes a time series forecasting framework that simultaneously predicts the trend intervals and points of significant infectious disease public opinion using the QD+ loss function. Based on this framework, uncertainties in public opinion trends can be quantified more accurately, facilitating the timely detection of abnormal fluctuations during early-stage public opinion monitoring. When the lower bound of public opinion intensity is high, it is important to monitor public discussions and emotional states to provide timely psychological support and prepare future epidemic policies.

To validate the performance of this framework, we compared it with other integrated deep learning models such as QD+ and IPIV. Both



(a) Feature count



(b) Top 5 features

Fig. 6. Feature selection process (lag = 7). Legend: the names (and values) of top 5 features can be found in Table 1, and their numerical suffix denotes the temporal lag structure.

the QD+ loss function and the IPIV-based public opinion trend prediction framework use NNs for interval and point prediction in regression problems. However, unlike the QD+ method, the IPIV framework does not directly output point predictions in its loss function construction. Instead, it outputs a weight value for a combined upper and lower bound interval without applying additional penalty functions. This weight value is then mathematically manipulated with the upper and lower bound values to represent the final point prediction (see Eqs. 23-25):

$$L_{\text{IPIVPI}} = \sqrt{n} \cdot \lambda \cdot \max(0, 1 - \alpha - \text{PICP})^2 \tag{23}$$

$$L_v = \frac{1}{n} \sum_{i=1}^{n} l\left(v_i \cdot U_i + (1 - v_i) \cdot L_i, y_i\right) \tag{24}$$

$$L_{\text{IPIV}} = \beta \cdot L_{\text{IPIVPI}} + (1 - \beta) \cdot L_v \tag{25}$$

Although this approach simplifies the construction of the loss function, it fundamentally smooths the prediction of public opinion trends using upper and lower bound values. Therefore, to demonstrate the superior performance of the QD+ algorithm, experiments were conducted to validate its efficacy.

Table 5
Interval prediction model parameters.

Model architecture	Parameter name	Parameter value
IPIV	λ	15
	β	0.5
	α	0.05
QD+	λ_1	0.99
	λ_2	0.3
	ξ	10
	α	0.05

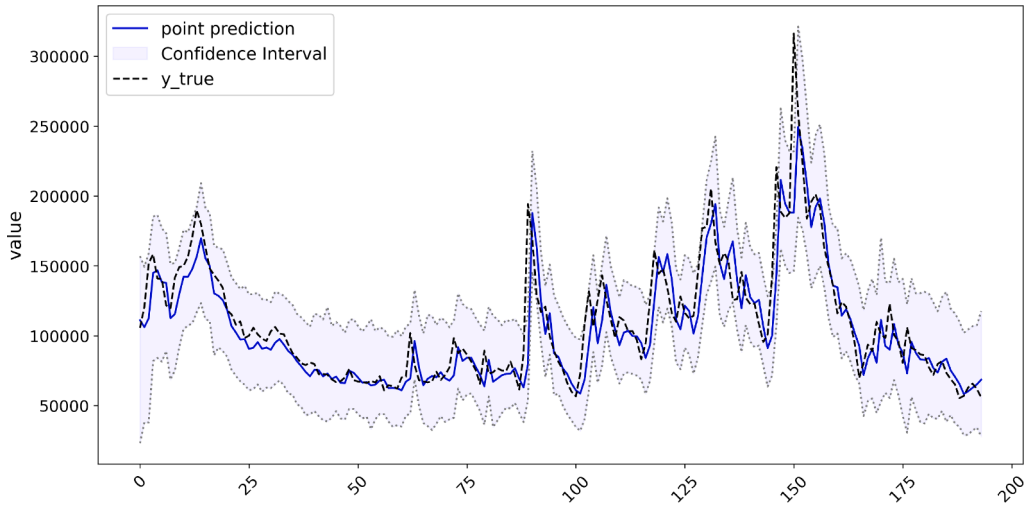


Fig. 7. Predictive results for QD+.

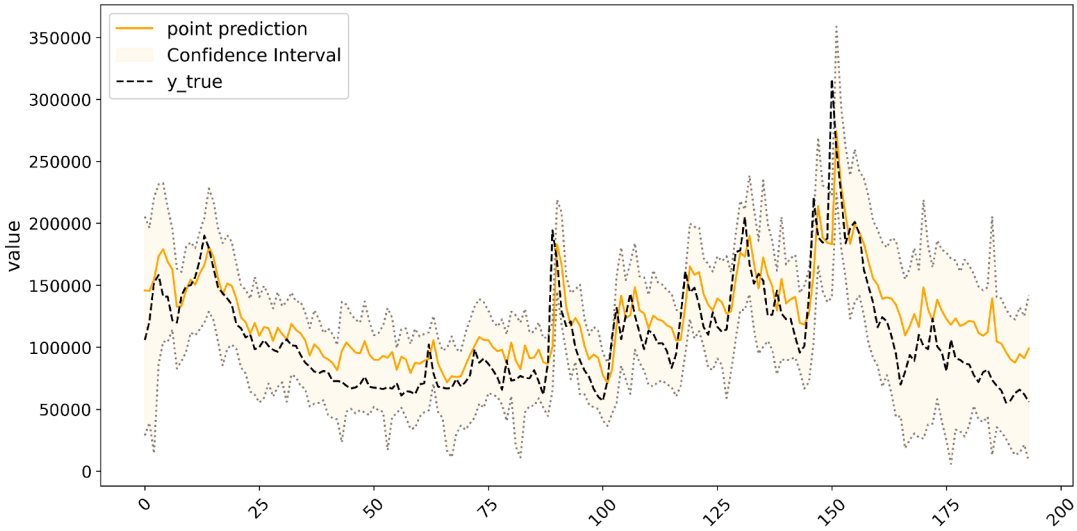


Fig. 8. Predictive results for IPIV.

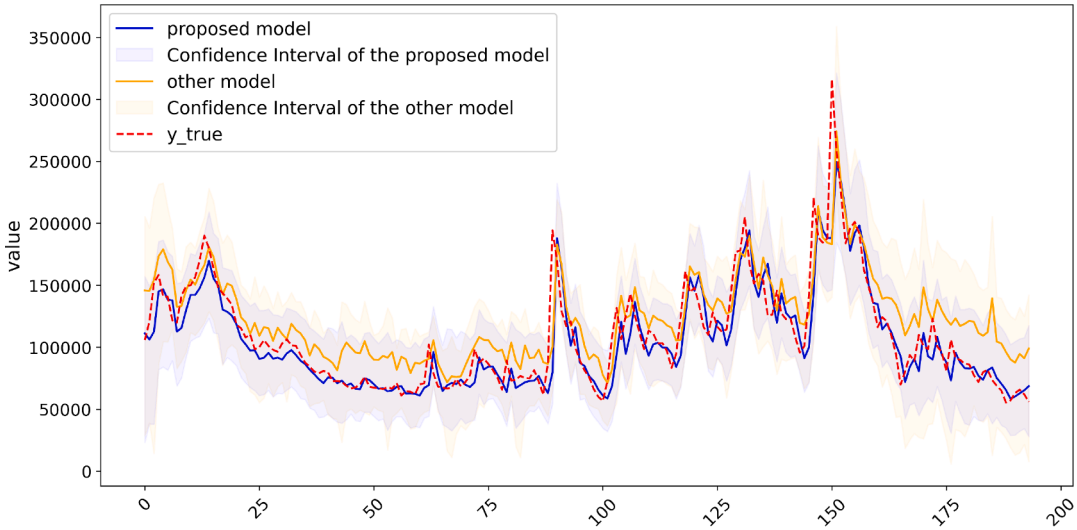


Fig. 9. Comparison of predictive results between QD+ and IPIV.

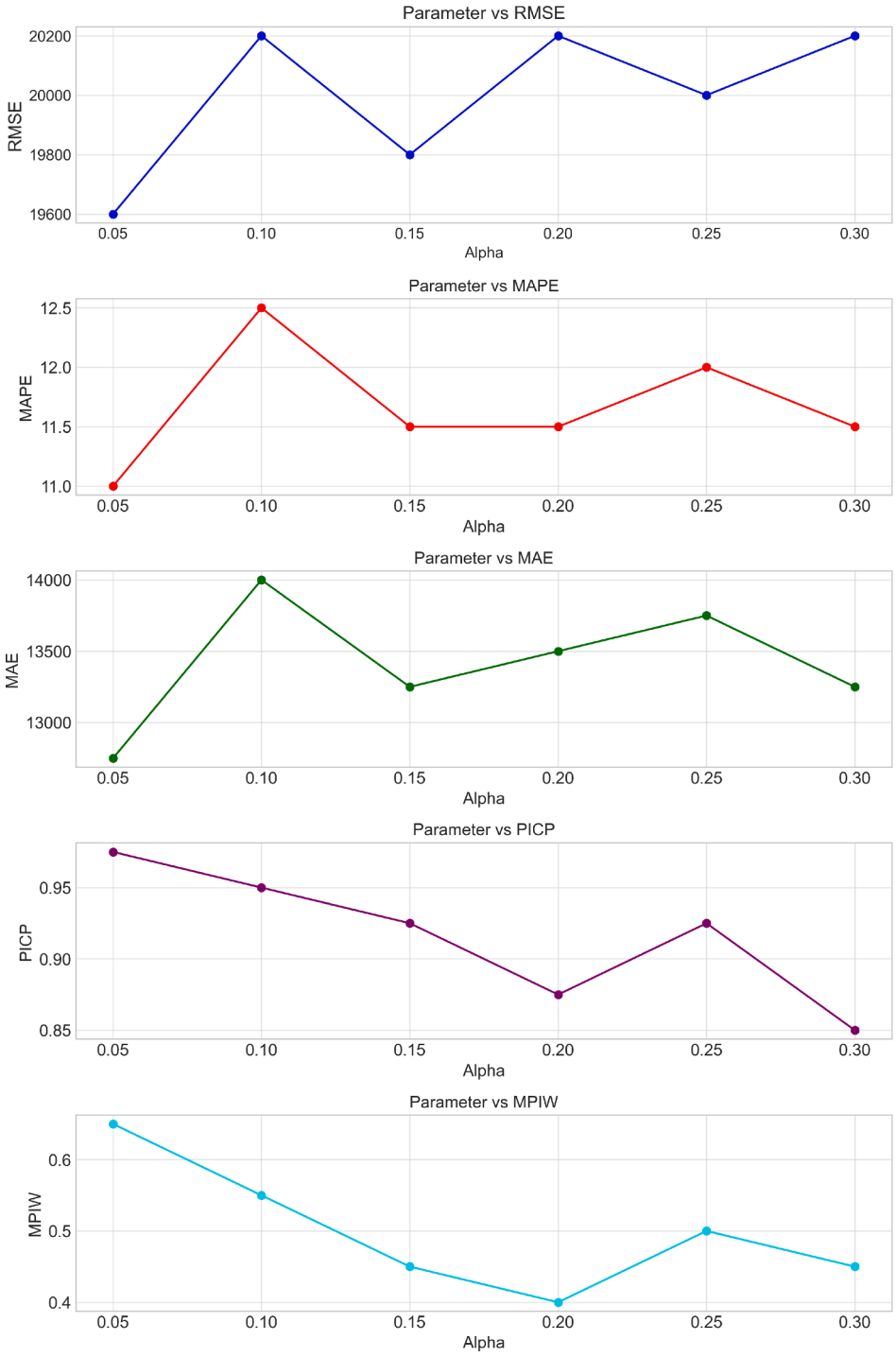


Fig. 10. Comparing metrics over different values of α.

Table 6
Model parameters.

Model	RNN	LSTM	GRU	Transformer	DNN
Hidden size	128	128	128	128	256
Num layers	3	3	3	3	5
Batch size	32	32	32	32	32
N_head	NA	NA	NA	3	NA
Max_len	NA	NA	NA	512	NA

Next, in the prediction framework, we selected a suitable deep learning model as the baseline model for interval and point prediction. As mentioned earlier, models such as DNN, RNN, LSTM, GRU, and Transformer have been widely applied in point prediction for multiple time series. However, the efficacy of these models for both interval and point prediction is less known. Therefore, we simultaneously compared multiple deep learning time series prediction models to determine the optimal model.

Furthermore, to validate the stability of the proposed framework, we conducted robustness analysis. Specifically, we used all feature sets as inputs to validate the effectiveness of this method and performed Diebold-Mariano (DM) tests on the results.

The hyperparameter settings for the QD+ and IPIV frameworks are listed in Table 5 (Salem et al., 2020; Simhayev et al., 2022), while the

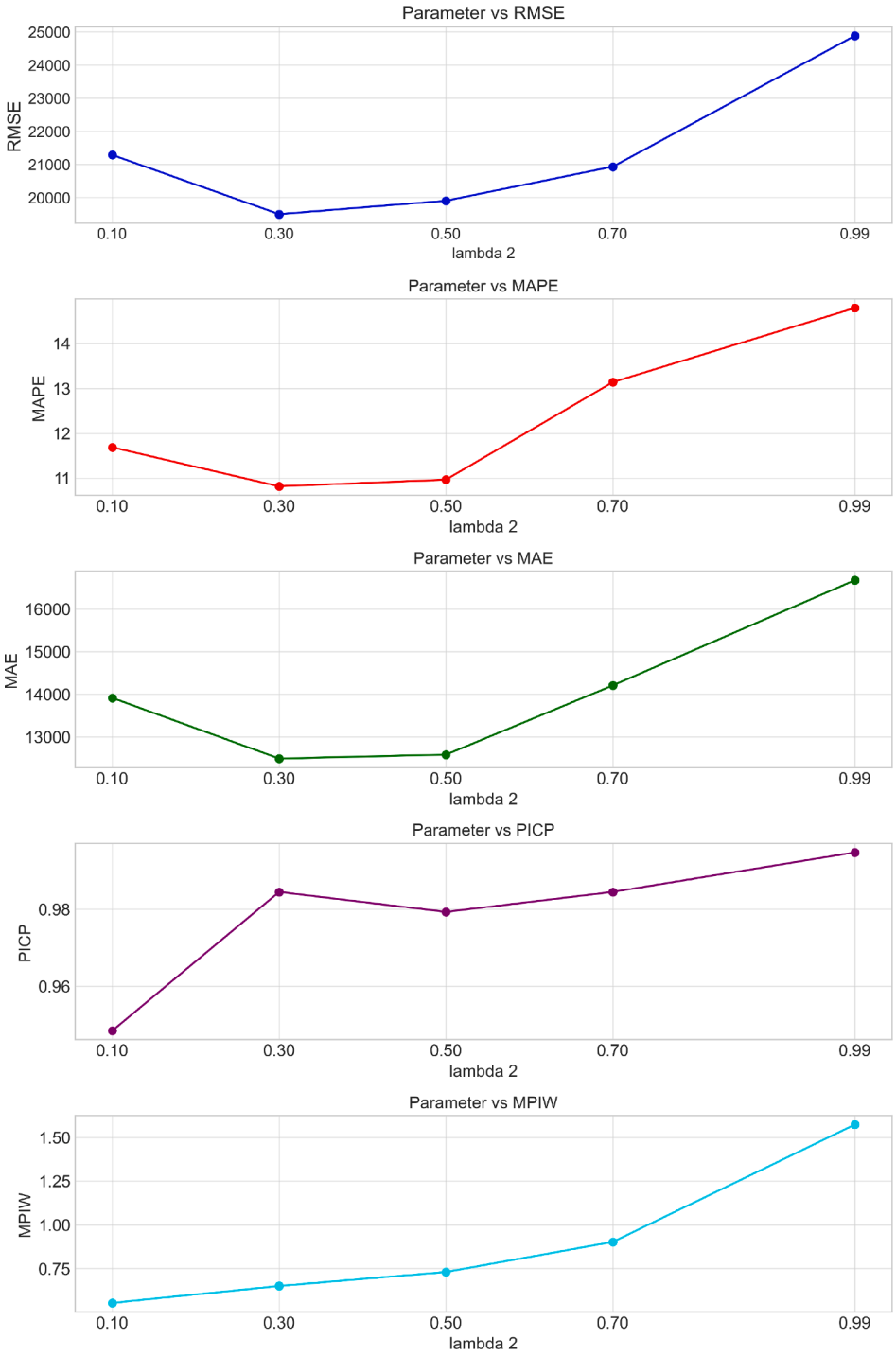


Fig. 11. Comparing metrics over different values of λ_2.

parameter settings for DNN, LSTM, RNN, GRU, and Transformer are provided in Table 6.

5.2.2. *Experiment results*

Table 7 presents the specific results of deterministic and interval prediction for each baseline model under different prediction frameworks. RMSE, MAPE, and MAE are used to measure the performance of point prediction models, while the quality of interval prediction is assessed using PICP and MPIW. From the comparison of deterministic predictions, it is evident that the results of each baseline model in the QD+ framework are generally superior to those in the IPIV framework. Among them, the best-performing model is the LSTM model used in the QD+ framework, with respective values of 19496.88, 10.82, and 12491.39 for RMSE, MAPE, and MAE.

Regarding interval prediction experimental results, considering both PICP and MPIW metrics, it can be observed that compared to the IPIV framework, the baseline models under the QD+ framework maintain a higher coverage rate while also sustaining a smaller interval width. The best-performing model in terms of interval prediction performance

Table 7
Results of model comparison.

Framework	Model	Features	RMSE	MAPE	MAE	PICP	MPIW
IPIV	**RNN**	**SUB**	**20710.88**	**12.73**	**13888.32**	0.9742	0.77
	LSTM	SUB	26563.01	16.66	16728.13	0.9226	**0.62**
	GRU	SUB	28983.70	20.85	20832.90	**0.9793**	0.77
	Transformer	SUB	27938.29	17.79	19998.22	0.9536	0.83
	DNN	SUB	29341.43	20.17	20328.04	0.9793	0.92
QD+	RNN	SUB	20104.98	**10.60**	12646.64	0.9845	0.68
	LSTM	**SUB**	**19496.88**	10.82	**12491.39**	**0.9845**	**0.65**
	GRU	SUB	21795.63	12.36	14500.21	0.9793	0.76
	Transformer	SUB	22286.00	12.07	14302.50	0.9742	0.83
	DNN	SUB	24397.00	15.76	17002.01	0.9845	1.28
QD+	RNN	ALL	32769.86	23.33	25299.92	0.9381	0.99
	LSTM	**ALL**	**32697.95**	22.95	**24230.69**	0.9536	0.93
	GRU	ALL	32731.49	**22.41**	25005.00	0.9175	**0.92**
	Transformer	ALL	42209.57	27.98	29329.62	0.9123	1.07
	DNN	ALL	34340.94	23.31	25768.25	**0.9896**	1.81

Table 8
DM results of different prediction models.

Model 1	Model 2	Framework	Features	DM	*p* value
LSTM_QD+_SUB	RNN	IPIV	SUB	-3.87	<0.001
	LSTM	IPIV	SUB	-5.33	<0.001
	GRU	IPIV	SUB	-4.07	<0.001
	Transformer	IPIV	SUB	-6.25	<0.001
	DNN	IPIV	SUB	-6.52	<0.001
RNN_QD+	RNN	QD+	SUB	-3.79	<0.001
	GRU	QD+	SUB	-4.33	<0.001
	Transformer	QD+	SUB	-4.89	<0.001
	DNN	QD+	SUB	-3.94	<0.001
RNN_QD+	RNN	QD+	ALL	-4.89	<0.001
LSTM_QD+	LSTM	QD+	ALL	-7.85	<0.001
	GRU	QD+	ALL	-6.65	<0.001
	Transformer	QD+	ALL	-6.79	<0.001
	DNN	QD+	ALL	-7.36	<0.001

Table 9
Comparing metrics over different values of pop size of NSGAII.

Lag	Pop size	RMSE	MAPE	MAE
1	50	23738.18	43.86	45161.26
	100	23832.55	41.63	43028.72
	150	24037.28	44.59	455792.38
	200	23951.49	42.38	431746.08
3	50	25630.52	48.66	47108.03
	100	22818.64	41.23	42822.68
	150	22824.48	47.16	47506.69
	200	22276.99	42.06	43595.30
5	50	22557.57	46.32	46647.71
	100	22545.33	45.13	45965.33
	150	23889.51	43.66	44310.70
	200	24082.09	46.24	46606.99
7	50	27072.87	49.08	47772.38
	100	22627.74	43.75	43953.37
	150	21811.17	44.32	44508.66
	200	25328.53	47.40	47063.49

under the QD+ framework is the LSTM model, with indicator values of 0.9845 and 0.65 for PICP and MPIW, respectively.

Furthermore, the optimal model for the QD+ framework using all features as input for prediction is LSTM, with indicator values of 32697.95, 22.95, 24230.69, 0.9536, and 0.931 for RMSE, MAPE, MAE, PICP, and MPIW, respectively. This result also demonstrates that the same model performs significantly better when using feature subsets as input compared to using the entire feature set, indicating the effectiveness of feature selection in improving prediction performance.

Table 10
Comparing metrics over different values of crossover probability of NSGAII.

Lag	Crossover	RMSE	MAPE	MAE
1	0.85	23932.55	44.52	45028.72
	0.90	23832.55	41.63	43028.72
	0.95	23612.61	43.89	43312.37
3	0.85	21982.04	41.50	42871.77
	0.90	22818.64	41.23	42822.68
	0.95	21808.21	41.63	42994.11
5	0.85	22837.00	46.12	45679.05
	0.90	22545.33	45.13	45965.33
	0.95	22346.19	45.67	45486.07
7	0.85	23997.50	46.28	46490.23
	0.90	22627.74	43.75	43953.37
	0.95	23171.51	45.98	45793.57

Table 11
Comparing metrics over different values of α.

α	RMSE	MAPE	MAE	PICP	MPIW
0.05	**19496.88**	**10.82**	**12491.39**	**0.9845**	**0.650**
0.10	20290.95	12.74	13980.51	0.9484	0.542
0.15	19870.26	11.25	13032.53	0.8969	0.480
0.20	20281.13	11.52	13317.04	0.8505	0.388
0.25	20022.58	12.06	13493.56	0.9432	0.509
0.30	20337.97	11.33	13210.63	0.8298	0.447

Table 12
Comparing metrics over different values of λ_2.

λ_2	RMSE	MAPE	MAE	PICP	MPIW
0.1	21285.03	11.69	13914.80	0.9484	0.553
0.3	**19496.88**	**10.82**	**12491.39**	0.9845	0.650
0.5	19903.60	10.97	12583.08	0.9793	0.730
0.7	20931.55	13.14	14207.27	0.9845	0.902
0.99	24879.85	14.79	16675.47	**0.9948**	**1.573**

Table 13
Comparing metrics over different noise levels.

Noise level	RMSE	MAPE	MAE	PICP	MPIW
0.05	19893.48	10.63	12650.84	0.9845	0.673
0.10	19878.04	10.46	12343.41	0.9793	0.626
0.20	20112.80	10.69	12466.76	0.9484	0.562

Figs. 7, 8, 9 depict the optimal prediction results under the two frameworks after feature selection. Here, “lstm_qd+_v” and “lstm_ipiv” represent the point predictions of the LSTM model under the QD+ and IPIV frameworks, respectively. “Confidence_interval_lstm_qd+_v” and “Confidence_inter-val_lstm_ipiv_v” represent the interval prediction results of the LSTM model under the QD+ and IPIV frameworks, respectively. These results demonstrate that the LSTM model based on the QD+ loss function achieves more accurate point predictions while ensuring minimum coverage rates.

Furthermore, to elucidate the stability of the QD+ framework, we employed the Diebold-Mariano (DM) test method Wang et al. (2024b) to show the superiority of the LSTM model under this framework compared to other versions of the model. The DM test is a statistical method used to compare the accuracy of two time series prediction models, assessing whether there is a significant difference in their predictions of future values. The specific comparison results are exhibited in Table 8.

Table 8 indicates that the LSTM_QD+ model significantly outperforms any model under the IPIV framework, as well as models under the QD+ framework such as RNN and the QD+ framework models using all features as input. This demonstrates the effectiveness of the QD+ framework combined with feature selection methods.

5.3. Robustness analysis

To further validate the effectiveness of the proposed framework, we conducted robustness tests by tuning parameters in both the NSAGII and QD+ framework. Specifically, for the NSGA-II algorithm, robustness was evaluated through systematic parameter variations: population sizes spanning $[50, 100, 150, 200]$ and crossover probabilities across $[0.85, 0.90, 0.95]$, with quantitative results detailed in Tables 8 and 9. Within the QD+ framework, we adjusted parameters α and λ_2, where α controls the width of the interval prediction-the larger its value, the greater the contribution of the interval width to the loss function-and λ_2 balances the emphasis between point prediction and interval prediction-the larger its value, the more the interval prediction dominates the loss function. The tested ranges for α and λ_2 were $[0.05, 0.1, 0.15, 0.2, 0.25, 0.3]$ and $[0.1, 0.3, 0.5, 0.7, 0.99]$, respectively. Additionally, to assess the framework's robustness, we introduced Gaussian random noise to 10 % of the randomly selected feature columns, with noise levels set at $[0.05, 0.1, 0.2]$. Table 10 compares performance metrics across different crossover probabilities of NSGAII, the algorithm demonstrates varying sensitivity to this parameter across different lag periods. In addition, the sensitivity analysis results for α are presented in Table 11 and Fig. 10.

The robustness analysis demonstrates that adjusting α and λ_2 affects the performance metrics, with $\alpha = 0.05$ yielding the best overall results for RMSE, MAPE, MAE, PICP, and MPIW.

The sensitivity analysis results for α indicate that indicators related to interval prediction generally show a decreasing trend, while indicators related to point prediction fluctuate around a certain threshold. This behavior is associated with α and its relation to interval prediction. The sensitivity analysis results for λ_2 are presented in Table 12 and Fig. 11.

λ_2 controls the weight ratio between point prediction and interval prediction. As the value of λ_2 in Table 12 increases, the proportion of the interval component in the loss function decreases, leading to a general upward trend in PICP and MPIW. The metrics related to point prediction exhibit a trend of initially decreasing and then increasing. The optimal balance between interval prediction and point prediction performance is achieved when $\lambda_2 = 0.3$.

Table 13 presents the performance of different models under varying noise levels. The model demonstrates strong noise robustness, as evidenced by only marginal performance degradation despite increasing noise levels, with all metrics eventually stabilizing at acceptable values.

6. Conclusion

This study proposes an innovative public sentiment prediction framework specifically designed for major infectious disease outbreaks. Building upon conventional approaches, our framework uniquely integrates information ecology theory with the NSGA-II optimization algorithm to identify critical predictive features. The methodological advancement lies in the novel incorporation of the QD + loss function within deep learning architectures, enabling simultaneous point and interval predictions with enhanced accuracy. This comprehensive approach provides real-time analytical capabilities for public sentiment monitoring while supporting evidence-based decision-making in epidemic containment strategies.

Through rigorous validation using COVID-19 as a representative case study with Weibo data - one of China's most influential social media platforms that serves as a reliable barometer for public opinion - three key findings emerge: (1) The NSGA-II algorithm configured with a 3-day lag period demonstrated superior performance across all evaluation metrics. (2) Notably, both epidemic progression indicators (e.g., Daily new deaths) and geographical factors (e.g., Huadong, Huabei, Xibei) persistently emerged as dominant features regardless of lag period duration, underscoring their pivotal role in driving online discussion intensity. (3) Comprehensive comparative analyses validated the QD+ framework's significant predictive advantages, with Diebold-Mariano (DM) tests statistically confirming the model's robustness ($p<0.01$).

Several important research directions merit further investigation. First, application of this framework to other infectious disease scenarios would help establish its generalizability across different epidemiological contexts, though we acknowledge our current findings are most applicable to China's digital ecosystem. Second, when implementing real-time prediction systems, careful consideration must be given to exogenous variables including policy interventions, economic fluctuations, and natural disasters due to their potential confounding effects on public sentiment formation. Additionally, from a methodological perspective, exploring alternative advanced deep learning architectures for time-series forecasting (e.g., transformer-based models or temporal fusion transformers) could potentially enhance performance. Future work could also benefit from incorporating data from additional platforms to provide more comprehensive insights into public discourse patterns.

CRediT authorship contribution statement

Futian Weng: Idea generation, Writing – original draft, methodology; **Meng Su:** Writing – original draft, methodology; **Petr Hajek:** Revision, validation; **Mohammad Zoynul Abedin:** Supervision, administration, validation, revision.

Data availability

Data will be made available on request.

Declaration of competing interest

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

Appendix A. Feature names for different LAG periods

LAG	Feature Name
1	'baidu_index_0', 'user_ver_4_0', 'middle_num_0', 'xibei_0', 'haiwai_0', 'Daily new confirmed cases_0', 'Daily new deaths_0', 'Daily new suspected cases_0', 'Daily new number of recoveries_0'
3	'baidu_index_2', 'user_ver_1_2', 'user_ver_4_2', 'older_num_1', 'huadong_0', 'huadong_1', 'xibei_2', 'happy_2', 'angry_2', 'sad_0', 'surprise_2', 'total_post_num_1', 'total_post_num_2', 'like_num_1', 'review_num_0', 'fans_num_0', 'Daily new confirmed cases_2', 'Daily new deaths_2', 'Daily new suspected cases_0', 'Daily new number of recoveries_2'
5	'baidu_index_1', 'baidu_index_2', 'baidu_index_3', 'baidu_index_4', 'user_ver_0_2', 'user_ver_1_0', 'user_ver_1_1', 'user_ver_1_2', 'user_ver_2_0', 'user_ver_2_3', 'user_ver_4_0', 'user_ver_4_3', 'female_num_2', 'female_num_3', 'youth_num_2', 'youth_adult_num_3', 'middle_num_0', 'middle_num_2', 'middle_num_4', 'older_num_0', 'huadong_0', 'huadong_4', 'huabei_2', 'dongbei_0', 'xibei_0', 'haiwai_2', 'positive_1', 'happy_2', 'happy_4', 'neutral_0', 'angry_0', 'angry_1', 'sad_1', 'fear_4', 'surprise_1', 'surprise_2', 'review_num_3', 'fans_num_2', 'Daily new deaths_0', 'Daily new deaths_1', 'Daily new suspected cases_1', 'Daily new number of recoveries_2', 'Daily new number of recoveries_3', 'Daily new number of recoveries_4'
7	'baidu_index_0', 'baidu_index_1', 'baidu_index_2', 'baidu_index_6', 'user_ver_0_0', 'user_ver_0_3', 'user_ver_0_6', 'user_ver_1_1', 'user_ver_2_0', 'user_ver_2_3', 'user_ver_2_6', 'user_ver_3_3', 'user_ver_3_6', 'user_ver_4_5', 'user_ver_4_6', 'female_num_0', 'female_num_3', 'male_num_0', 'male_num_4', 'male_num_6', 'kong_num_1', 'kong_num_2', 'kong_num_5', 'youth_num_0', 'youth_num_1', 'youth_adult_num_0', 'middle_num_6', 'older_num_0', 'older_num_2', 'older_num_6', 'huadong_1', 'huadong_3', 'huadong_4', 'huadong_5', 'huazhong_0', 'huazhong_2', 'huazhong_5', 'huabei_1', 'huabei_3', 'huabei_4', 'huabei_5', 'huanan_1', 'huanan_2', 'dongbei_4', 'dongbei_5', 'xinan_2', 'xinan_3', 'xibei_2', 'xibei_3', 'xibei_4', 'haiwai_5', 'negative_1', 'happy_0', 'happy_2', 'happy_4', 'neutral_0', 'neutral_5', 'angry_0', 'angry_1', 'sad_1', 'sad_3', 'fear_0', 'fear_1', 'fear_3', 'fear_4', 'fear_6', 'surprise_0', 'surprise_2', 'surprise_4', 'total_post_num_1', 'total_post_num_6', 'like_num_5', 'repost_num_0', 'repost_num_1', 'repost_num_3', 'repost_num_4', 'review_num_1', 'review_num_3', 'review_num_4', 'fans_num_4', 'Daily new confirmed cases_3', 'Daily new deaths_0', 'Daily new deaths_1', 'Daily new deaths_3', 'Daily new deaths_5', 'Daily new suspected cases_0', 'Daily new suspected cases_2', 'Daily new suspected cases_4', 'Daily new suspected cases_5', 'Daily new suspected cases_6', 'Daily new number of recoveries_4', 'Daily new number of recoveries_6'

References

Ameur, H. B., Boubaker, S., Ftiti, Z., Louhichi, W., & Tissaoui, K. (2024). Forecasting commodity prices: empirical evidence using deep learning tools. *Annals of Operations Research*, *339*, 349–367.

Bishop, C. M. (1995). Neural networks for pattern recognition. London: Oxford University Press.

Chen, C. Y. T., Sun, E. W., & Lin, Y. B. (2024). Reconciling spatiotemporal conjunction with digital twin for sequential travel time prediction and intelligent routing. (pp. 1–46). https://doi.org/10.1007/s10479-024-05990-x

Chen, Y., & Zhang, Z. (2022). An easy numeric data augmentation method for early-stage covid-19 tweets exploration of participatory dynamics of public attention and news coverage. *Information Processing & Management*, *59*(6), 103073.

Du, S. Y., Dai, Y. X., Li, P. W., Zhao, N., Li, S., & Zheng, Y. (2022). Vaccinated or not? survey on attitude toward 'approach-avoidance conflict' under uncertainty. *Human Vaccines & Immunotherapeutics*, *18*(1), 1–6.

Eshkiti, A., Sabouhi, F., & Bozorgi-Amiri, A. (2023). A data-driven optimization model to respond to covid-19 pandemic: A case study. *Annals of Operations Research*, *328*(1), 337–386.

Espinosa, R., Jiménez, F., & Palma, J. (2023). Multi-surrogate assisted multi-objective evolutionary algorithms for feature selection in regression and classification problems with time series data. *Information Sciences*, *622*, 1064–1091.

Ferchichi, A., Chihaoui, M., & Ferchichi, A. (2024). Spatio-temporal modeling of climate change impacts on drought forecast using generative adversarial network: A case study in africa. *Expert Systems with Applications*, *238*, 122211.

Gizelis, T.-I., & Karim, S. M. (2024). How epidemics affect marginalized communities in war-torn countries: Ebola, securitization, and public opinion about the security forces in liberia. *World Development*, *179*, 106587.

Gonzalez-Vidal, A., Jimenez, F., & Gomez-Skarmeta, A. F. (2019). A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. *Energy and Buildings*, *196*, 71–82.

Heskes, T. (1996). Practical confidence and prediction intervals. In *Advances in neural information processing systems 9*. https://papers.nips.cc/paper/1306-practical-confidence-and-prediction-intervals.

Jiménez, F., Sánchez, G., García, J. M., et al., (2017). Multi-objective evolutionary feature selection for online sales forecasting. *Neurocomputing*, *234*, 75–92.

Kellner, D., Lowin, M., & Hinz, O. (2023). Improved healthcare disaster decision-making utilizing information extraction from complementary social media data during the covid-19 pandemic. *Decision Support Systems*, *172*, 113983.

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Lower upper bound estimation method for construction of neural network-based prediction intervals. *IEEE Transactions on Neural Networks*, *22*(3), 337–346.

Kumar, A., & Taylor, J. W. (2024). Feature importance in the age of explainable ai: Case study of detecting fake news & misinformation via a multi-modal framework. *European Journal of Operational Research*, *317*(2), 401–413.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems*, *30*, 1–12.

Li, Y. T., Chen, M. L., & Lee, H. W. (2024). Health communication on social media at the early stage of the pandemic: Examining health professionals' covid-19 related tweets. *Social Science & Medicine*, *347*, 116748.

Liu, L., & Fu, Y. (2022). Study on the mechanism of public attention to a major event: The outbreak of covid-19 in china. *Sustainable Cities and Society*, *81*, 103811.

Liu, L., Tu, Y., & Zhou, X. (2022). How local outbreak of covid-19 affect the risk of internet public opinion: A Chinese social media case study. *Technology in Society*, *71*, 102113.

Liu, M., Tang, H., Chu, F., et al., (2024). A signomial programming-based approach for multi-echelon supply chain disruption risk assessment with robust dynamic bayesian network. *Computers & Operations Research*, *161*, 106422.

Luo, H., Meng, X., Zhao, Y., et al., (2023). Rise of social bots: The impact of social bots on public opinion dynamics in public health emergencies from an information ecology perspective. *Telematics and Informatics*, *85*, 102051.

Mahdikhani, M. (2022). Predicting the popularity of tweets by analyzing public opinion and emotions in different stages of COVID-19 pandemic. *International Journal of Information Management Data Insights*, *2*(1), 100053.

Mikulic, J., & Baumgartner, R. M. (2025). Google trends and baidu index data in tourism demand forecasting: A critical assessment of recent applications. *Tourism Management*, *110*, 105164.

Nardi, B. A., & O'day, V. (2000). Information ecologies: Using technology with heart. MIT Press.

Nix, D. A., & Weigend, A. S. (1994). Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE international conference on neural networks (ICNN'94* (pp. 55–60). IEEE.

Pearce, T., Brintrup, A., Zaki, M., et al., (2018). High-quality prediction intervals for deep learning: A distribution-free, ensembled approach. In *International conference on machine learning* (pp. 4075–4084). PMLR.

Salem, T. S., Langseth, H., & Ramampiaro, H. (2020). Prediction intervals: Split normal mixture from quality-driven deep ensembles. In *Conference on uncertainty in artificial intelligence* (pp. 1179–1187).

Simhayev, E., Katz, G., & Rokach, L. (2022). Integrated prediction intervals and specific value predictions for regression problems using neural networks. *Knowledge-Based Systems*, *247*, 108685.

Su, M., Cheng, D., Xu, Y., & Weng, F. (2023). An improved Bert method for the evolution of network public opinion of major infectious diseases: Case study of covid-19. *Expert Systems with Applications*, *233*, 120938.

Subramanian, H., Angle, P., Rouxelin, F., & Zhang, Z. (2024). A decision support system using signals from social media and news to predict cryptocurrency prices. *Decision Support Systems*, *178*, 114129.

Veaux, R. D. D., Schumi, J., Schweinsberg, J., & Ungar, L. H. (1998). Prediction intervals for neural networks via nonlinear regression. *Technometrics*, *40*(4), 273–282.

Wang, Z., Wang, Q., Liu, Z. et al., (2024b). A deep learning interpretable model for river dissolved oxygen multi-step and interval prediction based on multi-source data fusion. *Journal of Hydrology*, *629*. 130637.

Weng, F., Zhu, J., Yang, C., et al., (2022). Analysis of financial pressure impacts on the health care industry with an explainable machine learning method: China versus the usa. *Expert Systems with Applications*, *210*, 118482.

Xu, Z., Zhan, B., & Wang, S. (2023). Public opinion risk on major emergencies: A textual analysis. *Procedia Computer Science*, *221*, 833–838.

Xue, L., Zhou, K., & Zhang, X. (2024). Continuous optimization for construction of neural network-based prediction intervals. *Knowledge-Based Systems*, *293*, 111669.

Yan, R., Wang, S., Zhen, L., & Jiang, S. (2024). Classification and regression in prescriptive analytics: Development of hybrid models and an example of ship inspection by port state control. *Computers & Operations Research*, *163*, 106517.

Yan, S., Su, Q., Gong, Z., & Zeng, X. (2022). Fractional order time-delay multivariable discrete grey model for short-term online public opinion prediction. *Expert Systems with Applications*, *197*, 116691.

Yang, L., Yu, N., Li, X., & Wang, J. (2025). What motivates the intention of seeking confirmed cases' activity trajectory information in public health emergencies? An information ecology theory analysis. https://doi.org/10.1108/K-04-2023-0627

Yao, X., Xu, Z., Škare, M., & et al., (2024). Aftermath on COVID-19 technological and socioeconomic changes: A meta-analytic review. *Technological Forecasting and Social Change*, *202*, 123322.

Zhang, C., Tian, Y. X., & Fan, Z. P. (2022). Forecasting sales using online review and search engine data: A method based on PCA-DSFOA-BPNN. *International Journal of Forecasting*, *38*(3), 1005–1024.

Zhang, L., Su, C., Jin, Y., Goh, M., & Wu, Z. (2018). Cross-network dissemination model of public opinion in coupled networks. *Information Sciences*, *451*, 240–252.

Zhao, J., He, H., Zhao, X., & et al., (2022). Modeling and simulation of microblog-based public health emergency-associated public opinion communication. *Information Processing & Management*, *59*(2), 102846.

Zhong, C., Du, W., Xu, W., Huang, Q., Zhao, Y., & Wang, M. (2023). Lstm-regat: A network-centric approach for cryptocurrency price trend prediction. *Decision Support Systems*, *169*, 113955.

Zhu, B., Zheng, X., Liu, H., & et al., (2020). Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. *Chaos, Solitons & Fractals*, *140*, 110123.