Deep-learning architecture for PM 2.5 concentration prediction: A review

PM 2.5浓度预测的深度学习架构：综述

主要内容

Main Contents

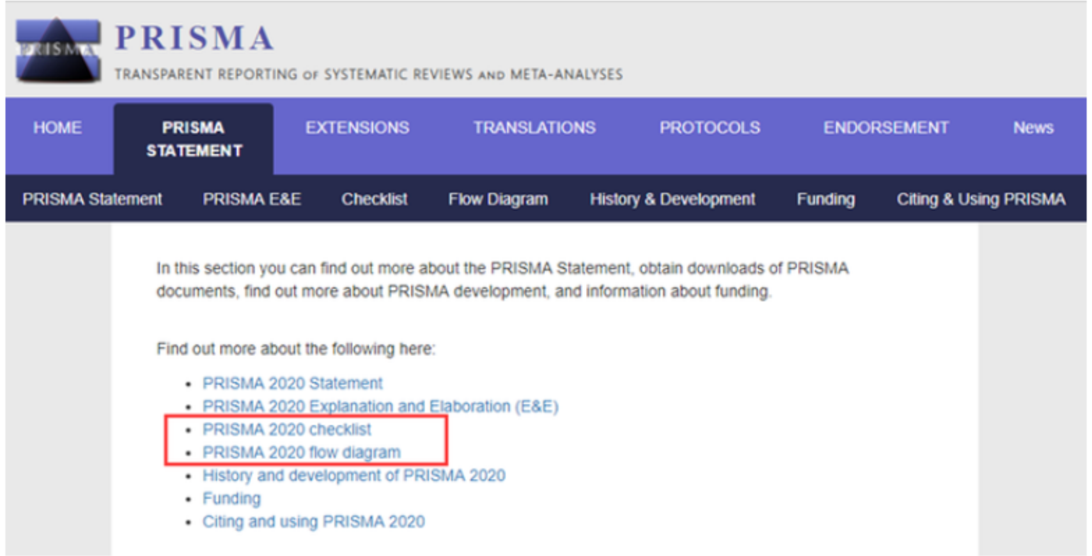
摘要

PM 2.5浓度预测的深度学习架构

- **问题**
  - 缺乏统一的标准化框架来评估基于 DL 的 PM 2.5 预测模型的性能
- **方法**
  - PRISMA
- **分类**
  - 基于DL的模型
    - 效率高，可解释性差
  - 混合学习模型
    - 与常规方法结合
      - 如确定性模型和统计模型可解释性高，但在准确性和速度上有所妥协
    - 与深度学习方法结合
      - 创新性强，可解释性低
- **成果**
  - 提出了一种新的三维评价框架，即数据集-方法-实验标准(DMES)，以统一和规范基于DL模型的pm2.5预测评价

背景

PRISMA指南的官方网站请见prisma-statement.org/。在“PRISMA Statements”（PRISMA声明）标签下面，可找到PRISMA指南的两份重要材料。



一份是指南的清单，英文名PRISMA checklist（即PRISMA检查表）。检查表中总计27个检查项目，以一篇研究型论文的正文组织结构为序，从标题（title）到摘要（abstract）、前言（introduction）、方法（methods）、结果（results）、讨论（discussion）及其他附加信息，描述了一项科学严谨且具备可重复性的综述性研究所应当遵循的每个步骤。

指南不仅指导作者如何设计一篇高质量的综述报告，也指导作者如何进行数据分析、写作中如何表述。建议研究者在进行系统综述之前先熟悉本指南。**在撰写系统综述时遵循该指南，有利于控制和降低研究的不可重复性和研究人员的偏倚。**

第二份材料是一张流程图，英文名PRISMA flow diagram。这张流程图用“如果……那么……”的结构，将收集到的论文按照预设的标准分类（分类情形详见PRISMA checklist文件）。

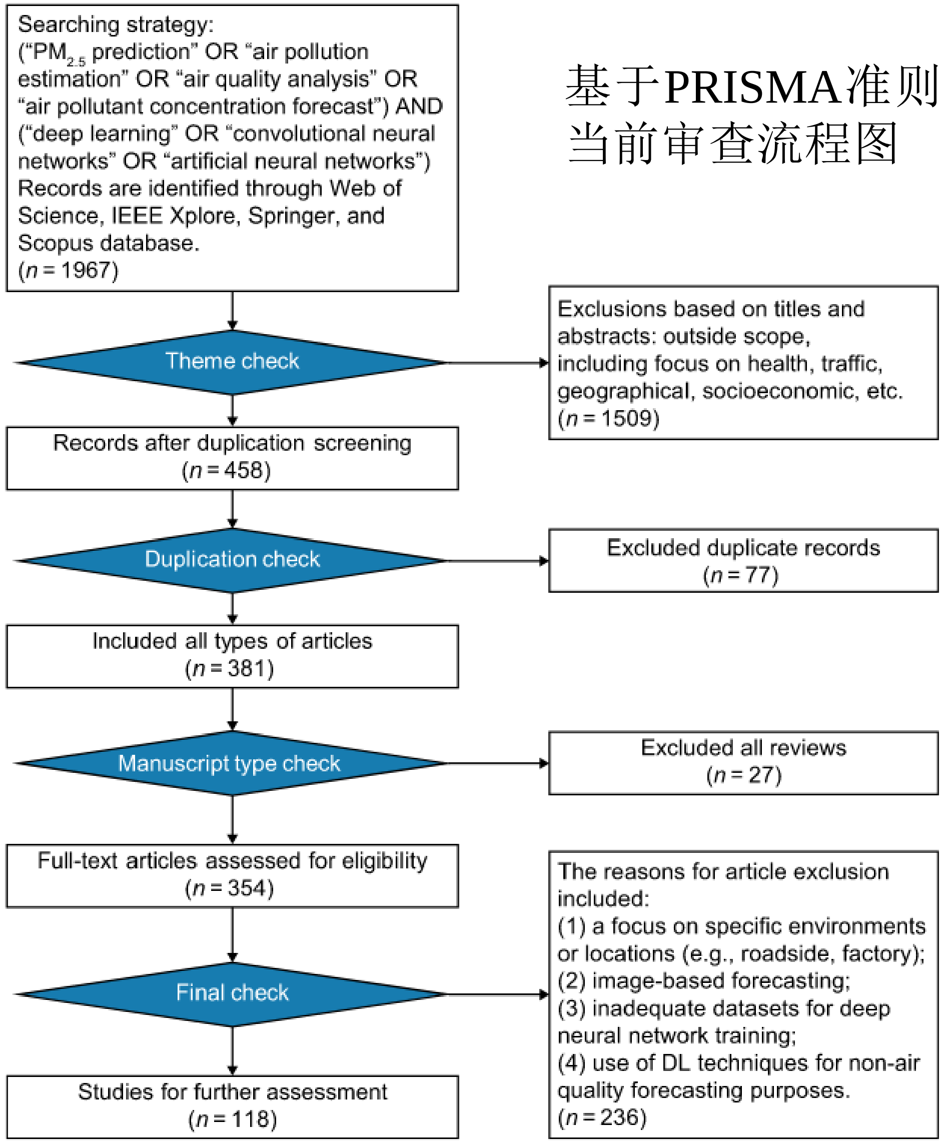
PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	

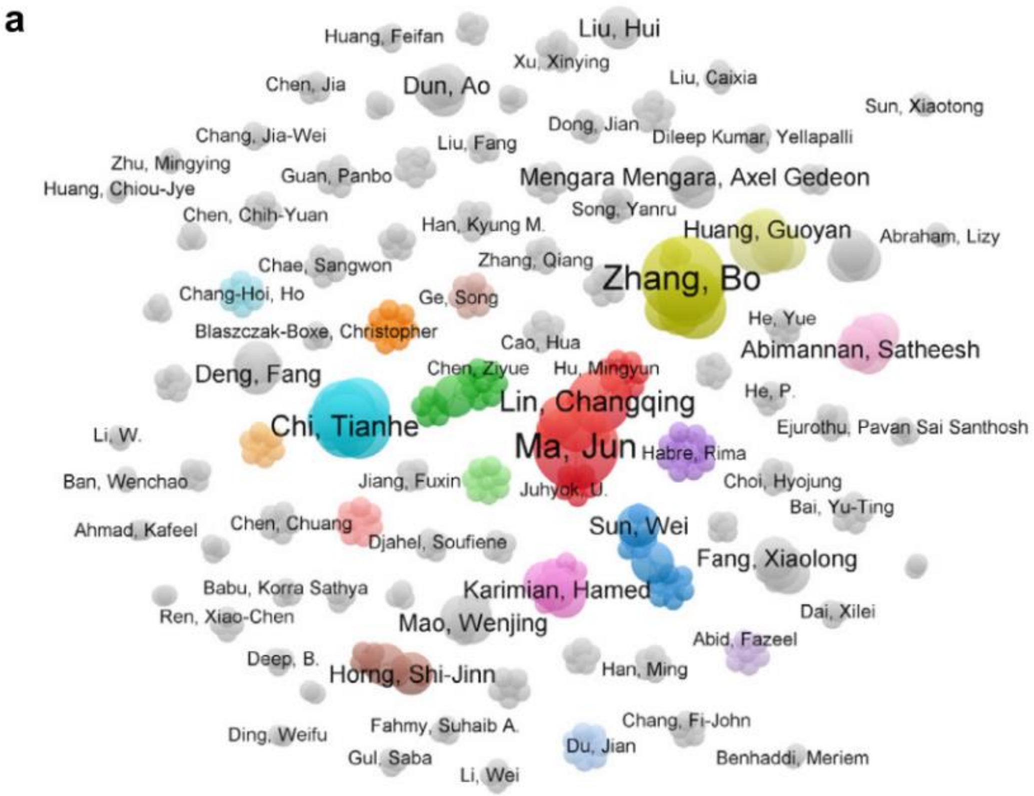
PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	
	23b	Discuss any limitations of the evidence included in the review.	
	23c	Discuss any limitations of the review processes used.	
	23d	Discuss implications of the results for practice, policy, and future research.	
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	
Competing interests	26	Declare any competing interests of review authors.	
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	

方法

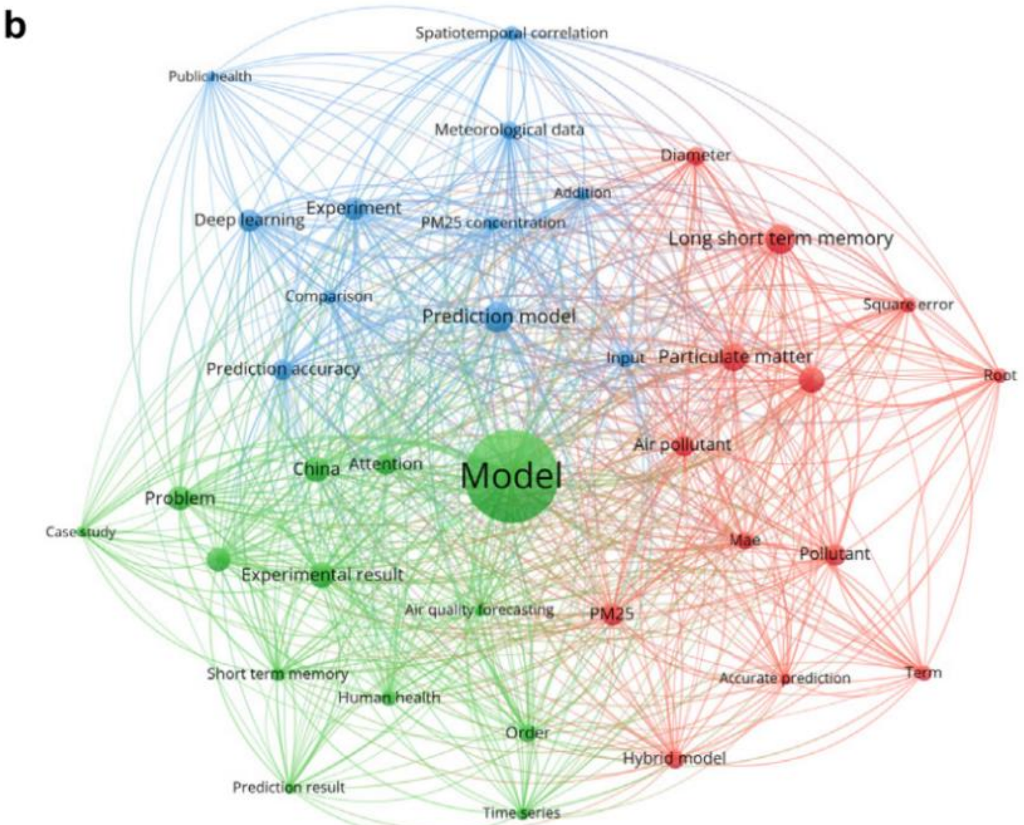


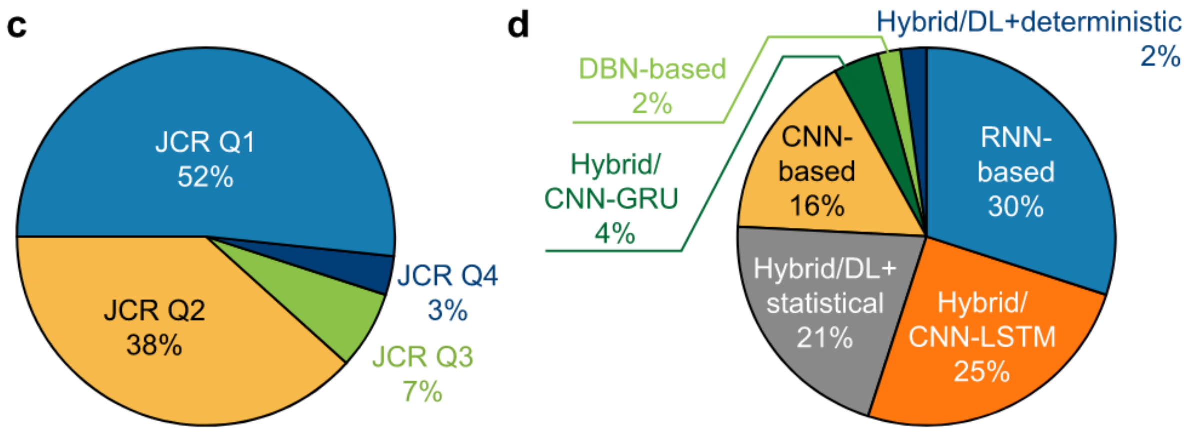
基于PRISMA准则的当前审查流程图



a，被评审论文作者之间的关系。显示区域的大小表示作者所写文章的数量，作者之间的距离表示作者之间的交流程度，颜色表示使用聚类算法获得的相关程度。

文章的作者之间相对独立，这可能解释了基于深度学习的预测和评价缺乏统一的标准。





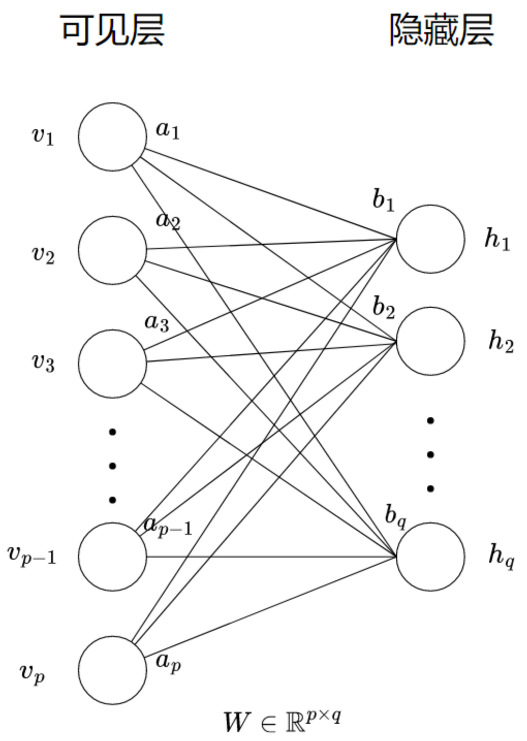
b，关键词趋势。圆圈大小表示文章中出现的总体频率。每当两个单词同时出现时，就会在它们之间画一条线段，线段的数量反映了单词之间的关系。同时出现的高频词通过聚类进行分组和汇总，颜色表示该词的频率分类。没有明显的关键词聚类，表明目前的研究谱系缺乏清晰度。

c，期刊引文报告所选论文的划分，分别显示四分位数Q1、Q2、Q3和Q4中文章的百分比。
文章的质量很高，其中一半以上的文章发表在科学引文索引(SCI) Q1。
d，根据文章的结构对所选文章进行分类，显示文章在各种模型上的比例。
最近基于dl的PM2.5浓度预测研究大多使用了与LSTM网络相关的算法。此外，许多研究采用混合模型算法。

文章归类
基于深度学习的方法
深度信念网络DBN
卷积神经网络CNN
基于CNN/DNN的改进
基于图卷积神经网络(GCNNs)的改进
基于时空(ST)相关性的改进
循环神经网络RNN
基于RNN
基于LSTM
Transformer
基于深度学习的混合方法
深度学习与传统方法相结合
DL 加确定性方法
DL 加统计方法
经典的统计方法
自回归综合移动平均(ARIMA)
小波变换(WT)
经验模态分解(EMD)
传统的机器学习方法
深度神经网络集成方法
CNN + LSTM
CNN + GRU

深度信念网络（Deep Belief Network, DBN）由 Geoffrey Hinton 于2006年提出。DBN 可以看作是一系列受限玻尔兹曼机（RBM）的堆叠。

受限玻尔兹曼机（restricted Boltzmann machine, RBM）是一种**无向概率图模型**，并且受限为**二分图**。整个模型有两层——可见层（包含可见单元）和隐藏层（包含隐单元），满足**层内无连接，层间全连接**。



标准的受限玻尔兹曼机由二值隐单元和可见单元组成。权重矩阵$W=(w_{ij})$中的每个元素指定了隐单元 h_i 和可见层单元v_i之间边的权重。此外对于每个可见层单元v_i有偏置项a_i，对每个隐层单元 h_i 有偏置项 b_i。

$$v_i \in \{0,1\}, 1 \le i \le p$$
$$h_j \in \{0,1\}, 1 \le j \le q$$
$$W \in \mathbb{R}^{p\times q}$$

能量函数定义为：
$$\begin{aligned} E(v,h) &= -\sum_{i=1}^{p} a_i v_i - \sum_{j=1}^{q} b_j h_j - \sum_{i,j} w_{ij} v_i h_j \\ &= -a^T v - b^T h - v^T W h \end{aligned}$$

给定训练集的 n 个样本，RBM 的关键就是计算模型中的参数 $\theta=(W,a,b)$

RBM 一般采用对数损失函数，考虑最大化$log(L)=\sum_{i=1}^{n} logL(x^{(i)};\theta)$

$$L(x^{(1)},x^{(2)},\ldots,x^{(n)};\theta)=\prod_{i=1}^{n}L(x^{(i)};\theta)$$

$$\text{with } L(x;\theta)=P_\theta(v=x)=\sum_h P_\theta(x,h)=\sum_h e^{-E_\theta(x,h)}$$

梯度计算：2002年，Geoffrey Hinton 提出对比散度 (contrastive divergence, CD) 算法来训练 RBM，成为 RBM 的标准训练算法。

k 步 CD 算法:

给定样本 x，取初始值 $v^{(0)}:=x$。然后执行 k 步 Gibbs 采样。其中第 t 步（$t=1,2,\ldots,k$）先后采样:

- 根据 $P(h|v^{(t-1)})$，采样得到 $h^{(t-1)}$
- 根据 $P(v|h^{(t-1)})$，采样得到 $v^{(t)}$

利用 k 步之后得到的 $v^{(k)}$ 估算 (1), (2), (3) 中 $\sum_v P(v)\ldots$ 对应的期望项:

$$\frac{\partial}{\partial w_{ij}}logL(x;\theta)\simeq v_i^{(0)}P(h_j=1|v^{(0)})-P(h_j=1|v^{(k)})v_i^{(k)} \tag{4}$$

$$\frac{\partial}{\partial a_i}logL(x;\theta)\simeq v_i^{(0)}-v_i^{(k)} \tag{5}$$

$$\frac{\partial}{\partial b_j}logL(x;\theta)\simeq P(h_j=1|v^{(0)})-P(h_j=1|v^{(k)}) \tag{6}$$

估算得到梯度的近似值后，在每一步利用梯度上升法进行参数更新即可：

$w_{ij} = w_{ij} + \epsilon \Delta w_{ij}$，其中 $\Delta w_{ij} = \frac{1}{n_{\text{batch}}} \sum_x \frac{\partial}{\partial w_{ij}} logL(x;\theta)$ （一组 batch 数据的平均梯度）

参数 a_i, b_j 同理。

> Remark: 实际应用中，**1 步 CD 算法的效果就很不错了。**

例如，在推荐系统中，我们可以把每个用户对各个物品的评分作为可见层神经元的输入，训练模型时，对于每个样本，我们仅仅用有用户数值的可见层神经元来训练模型。
对于可见层输入的训练样本和随机初始化的 W, a，我们可以用 sigmoid 激活函数得到隐藏层的神经元的 0,1 值，这就是编码。然后反过来从隐藏层的神经元值和 W, b ,可以得到可见层输出，这就是解码。对于每个训练样本，我们期望编码解码后的可见层输出和我们的之前可见层输入的差距尽量的小，即上面的对数似然损失函数尽可能小。按照这个损失函数，我们通过迭代优化得到 W, a, b。对于某个没有评分的物品，我们用解码的过程可以得到一个预测评分，取最高的若干评分对应的物品即可作用户物品推荐。

DBN 的训练思路是**贪心逐层训练**。具体而言，从下到上分别将每层当做 RBM 进行训练；然后固定当前层权值，取样当前层的“隐层”作为下一层的输入。

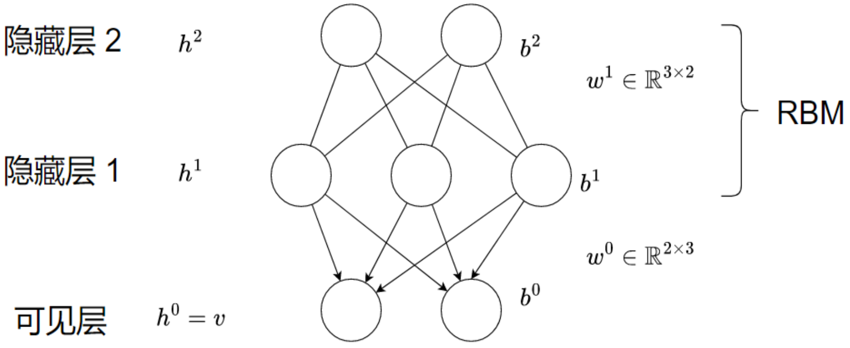
DBN 属于**概率图模型**，并且是有向图与无向图的混合。只有最后两个隐藏层之间是无向图（这是一个 RBM），其余的都是有向图。

我们的目标是最大化 $p(v) = \sum_h p(v,h) = \sum_h p(h)p(v|h)$

一个想法是，能不能把 $p(v|h)$ 固定住，然后最大化 $p(h)$，以期让 $p(v)$ 更大呢？从这个想法出发，我们把 h 作为另一个 RBM 的输入；训练这个 RBM 不就让 $p(h)$ 最大化了吗？

从这个角度来看，DBN 的结构设计就很合理了。还有一个问题，为什么要把前面的层设计为有向图，而不是像 RBM 那样的无向图？

shuhuai 大佬在这里（深度信念网络：叠加 RBM 的动机）给出的解释是：由于要固定 $p(v|h)$，所以要把 v 到 h 的方向去掉，以保持该层已训练好的权重不变。（对此我其实不是很理解



1. 从 $p(h^2)$ 中采样得到 h^2；
2. 从 $p(h^1|h^2)$ 中采样得到 h^1；
3. 从 $p(v|h^1)$ 中采样得到 v

基于深度学习的方法的总结

模型		考虑
基于DBN		DBN参数优化、影响PM2.5浓度的全方面因素
基于CNN	基于CNN/DNN	结合气象数据和卫星图像、过去和未来的时间信息、数据集扩大、地理区域和环境的迁移性、时空信息、插值实现实时预测、注意力机制进行特征选择
	基于GCNN	时空关系、学习到的特征未知、模型的可解释性、聚类、复杂网络分析以确定最重要的空气质量监测站、研究地域尽量广
	基于ST	时空注意和残差学习、长期依赖关系、时空特征、体系结构复杂而仅限于特定地区或污染物
基于RNN	基于RNN	多种输入数据（包含历史PM 2.5浓度、气象数据和空气质量指数(AQI)数据）
	基于LSTM	模型的可解释性、BiLSTM模型处理空间和时间相关性、CNN和AutoEncoder (AE)与BiLSTM集成、平衡采样、将主成分分析(PCA)、注意机制和LSTM相结合进行特征提取、捕获局部和全局依赖关系、结合多个数据源、迁移学习、加权LSTM
基于Transformer		时间嵌入、探针稀疏的自注意力机制可以对最重要的查询进行排序、ResInformer、残差连接

Table 1
The research used DBN-based methods.

Study	Year	Location	Model	Time step	RMSE ($\mu g\ m^{-3}$)	MAE ($\mu g\ m^{-3}$)	R^2
H.Xing et al. [16]	2021	Beijing, China	TDBN	D/S/T+1	11.197	12.298	0.862
Y.Xing et al. [17]	2019	Baoding, China	MGWO + DBN	-	20.260	17.604	0.884

Table 2
The research used CNN-based methods.

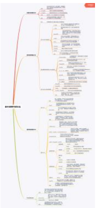
Study	Year	Location	Model	Time step	RMSE ($\mu g\ m^{-3}$)	MAE ($\mu g\ m^{-3}$)	MAPE (%)	R^2
Zhang et al. [33]	2023	Yangtze River Delta Region, China	STA-ResCNN	H/M/T+1	6.98	3.91	12.62	-
Yu et al. [34]	2023	Los Angeles, US	ST-Transformer	H/S/T+12	6.92	4.00	-	-
Choudhury et al. [26]	2022	Delhi, India	AGCTCN	H/M/T+(1-3)	11.76	8.75	-	0.64
Li et al. [18]	2016	Beijing, China	STDL	S/S/-	14.96	9.00	21.75	-
Li et al. [19]	2020	California, US	ensemble-based DL	D/S/T+1	2.70	-	-	-
Samal et al. [20]	2021	Talcher, India	MTCAN	D/S/T+14	9.00	7.00	-	-
Luo et al. [19]	2020	Shanghai, China	CNN + GBM	H/S/T+1	10.02	-	-	0.85
Chae et al. [24]	2021	South Korea	ICNN	H/S/T+24	1.64	-	-	0.97
Zhang et al. [32]	2021	Beijing, China	ST-CausalConvNet	H/M/-	17.43	11.74	-	0.93
Shi et al. [25]	2021	Beijing, China	DSTP-FC(Encoder-Decoder]	H/M/T+(1-6)	32.51	19.50	-	-
Xiao et al. [27]	2022	-	DP-DDGCN	H/S/T+9	11.75	14.53	-	-
Zhao et al. [28]	2021	Jing-Jin-Ji Region, China	AQSTN-GCN	H/S/T+1	19.00	12.03	0.30	0.94
Wang et al. [36]	2022	China	STWC-DNN	H/S/T+1	12.70	-	-	0.92
Wang et al. [23]	2020	Shanghai, China	Sequence-to-Sequence	D/S/T+7	22.32	-	-	0.52
Ni et al. [22]	2022	Beijing/Tianjin, China	TL-DSTP-DANN	H/S/T+3	15.97	11.75	20.00	-
Dun et al. [35]	2022	Fushun, China	DGRA-STCN	H/S/T+2	12.50	8.21	88.40	-
Ouyang et al. [29]	2022	Beijing, China/London, UK	DC-STDGN	H/M/T+(1-3)	30.58/13.42/4.28	29.63/12.15/3.03	-	-
Ejurothu [31].	2022	New Delhi, India	HGNN	H/S/T+8	19.83	16.61	-	-
Dun et al. [30]	2022	Beijing/Fushun, China	DGC-MTCN	H/S/T+1	9.77/12.96	5.54/8.39	-	0.95/0.91

Table 3
The research used RNN-based methods.

Study	Year	Location	Model	Time step	RMSE (μg m^{-3})	MAE (μg m^{-3})	MAPE (%)	R^2
Ayturan et al. [38]	2018	Ankara, Turkey	RNN	H/S/T+1	6.28	4.21	-	-
Dai et al. [37]	2021	Tianjin, China	RNN	H/-/T+1	11.87	-	-	-
Ma et al. [51]	2019	Guangdong, China	BiLSTM	H/S/T+1	8.24	4.80	9.01	-
Li et al. [39]	2017	Beijing, China	LSTM	H/S/T+1	12.60	5.46	11.93	-
Zhao et al. [65]	2019	Beijing, China	LSTM-FC	H/M/T+(1-6)	35.82	23.97	-	-
Wen et al. [66]	2019	Beijing, China	STCLSTM	H/M/T+1	12.08	5.82	17.09	-
Zhou et al. [67]	2019	Taiwan, China	DM-LSTM	H/S/T+1	4.49	-	-	-
				H/S/T+4	9.31	-	-	-
Ma et al. [70]	2019	Guangdong, China	TL-BLSTM	H/S/T+1	8.54	4.95	22.32	-
Chang et al. [40]	2020	Taiwan, China	LSTM	H/S/T+1	-	-	-	-
Xayasouk et al. [41]	2020	Seoul, South Korea	LSTM	H/S/T+1	11.11	-	-	-
Karimian et al. [42]	2019	Tehran, Iran	LSTM	H/S/T+12	10.32	7.41	-	0.74
Tong et al. [52]	2019	Florida, US	BiLSTM	H/S/T+1	3.65	1.62	18.48	-
Mao et al. [43]	2021	Jing-Jin-Ji Region, China	LSTM	H/M/T+(1-24)	20.68	14.56	-	0.74
Ma et al. [61]	2020	Wayne, US	Lag-LSTM	H/S/T+1	3.48	1.85	25.63	-
Zhang et al. [58]	2020	Beijing, China	AE + BiLSTM	H/S/T+24	2.19	-	-	-
Zou et al. [59]	2021	Yangtze River Delta Region, China	FDN (AE + LSTM)	H/S/T+1	4.32	3.31	-	-
Xu et al. [57]	2021	Beijing, China	AE + LSTM	H/S/T+1	14.52	8.22	45.40	-
				H/M/T+(1-3)	24.87	15.60	64.72	-
Qadeer et al. [44]	2020	Seoul, South Korea	LSTM	H/-/-	4.82	3.58	-	0.87
Zhang et al. [53]	2021	Beijing, China	BiLSTM	H/-/T+1	17.20	14.15	-	-
Wang et al. [64]	2021	Beijing, China	CR-LSTM	H/S/T+24	8.96	12.89	-	0.74
Shi et al. [60]	2022	Beijing, China	BS-LSTM	H/S/T+3	32.32/12.42	18.36/9.75	-	-
Kristiani et al. [45]	2022	-	LSTM	H/S/T+1	1.90	1.27	11.12	-
Deep et al. [54]	2022	Delhi, India	BiLSTM	H/S/T+1	15.59	-	-	-
Sun et al. [68]	2019	Liaoning, China	LSTM-DRSL	H/S/T+1	10.53	9.09	20.05	-
Lin et al. [46]	2020	Taiwan, China	LSTM	H/S/T+1	4.46	-	30.00	0.86
Park et al. [47]	2021	Seoul, South Korea	LSTM	H/S/T+3	-	-	-	-
Mengara et al. [56]	2022	Seoul, South Korea	AE + BiLSTM	H/S/T+1	7.48	5.02	30.48	-
Mengara et al. [55]	2020	Busan, South Korea	CNN + BiLSTM	H/S/T+1	6.93	5.07	30.90	-
Ding et al. [62]	2022	Ningxia, China	PCA-Attention-LSTM	D/S/T+1	7.57	4.93	-	0.91
Peralta et al. [48]	2022	Santiago, Chile	LSTM	H/S/T+1	9.85	4.40	-	0.74
Liu et al. [71]	2022	Jing-Jin-Ji Region, China	MGC-LSTM	H/S/T+1	2.91	2.16	12.96	-
Hu et al. [63]	2022	Beijing, China	Conv1D-LSTM	H/S/T+1	20.76	11.20	-	0.96
Wu et al. [69]	2022	Beijing, China	CE-AGA-LSTM	H/S/T+1	21.88	14.49	-	0.95
Waseem et al. [49]	2022	Lahore/Karachi/Islamabda, Pakistan	LSTM	H/S/T+1	-	-	11.70/7.40/9.50	-
				D/S/T+1	-	-	28.2/42.1/15.1	-
Gul et al. [50]	2022	Punjab, India	LSTM	H/S/T+1	0.19	-	-	-
				H/S/T+(1-24)	0.73	-	-	-
Xiao et al. [72]	2020	Jing-Jin-Ji Region, China	WLSTME	D/S/T+1	40.67	26.10	-	-

Table 4
The research used Transformer-based methods.

Study	Year	Location	Model	Time step	RMSE (μg m^{-3})	MAE (μg m^{-3})	MAPE (%)	R^2
Zhou. H et al. [73]	2021	Beijing, China		D/S/T+1	0.2852	0.2159	0.80	0.8285
		Shijiazhuang, China	Informer	D/S/T+1	0.5112	0.2890	1.79	0.6433
		Wuhan, China		D/S/T+1	0.5225	0.4180	1.44	0.5329
	2021	Beijing, China		D/S/T+1	0.2692	0.2012	0.89	0.8472
		Shijiazhuang, China	InformerStack	D/S/T+1	0.5408	0.3081	2.25	0.6020
		Wuhan, China		D/S/T+1	0.3716	0.2911	1.49	0.7621
M.A.A.A.-q et al. [74]	2023	Beijing, China		D/S/T+1	0.2822	0.2130	0.85	0.8320
		Shijiazhuang, China	ResInformer	D/S/T+1	0.4646	0.3138	2.00	0.5857
		Wuhan, China		D/S/T+1	0.4706	0.3782	1.54	0.6142
	2023	Beijing, China		D/S/T+1	0.2623	0.1964	0.75	0.8549
		Shijiazhuang, China	ResInformerStack	D/S/T+1	0.5343	0.3055	1.91	0.4937
		Wuhan, China		D/S/T+1	0.3712	0.2982	1.39	0.7656



基于深度学习的混合方法的总结

模型				特点	考虑
深度学习与传统方法相结合	DL 加确定性方法			采用气象学原理和数学方程，以大气物理化学反应为基础，模拟污染物的排放、转化、扩散和去除过程，如天气研究和预测(WRF)模型、社区多尺度空气质量(CMAQ)	确定性CMAQ模型与RNN相结合同时考虑各种气象和环境因素；数值模拟模型WRF和CMAQ模拟PM2.5浓度的空间分布，而LSTM用于预测这些浓度的时间变化
	DL 加统计方法	经典统计方法	小波变换(WT)	将信号或数据分解为称为小波的较小分量，这些分量基本上是不同频率的小波状函数和持续时间。然后使用这些小波以更容易分析的形式表示原始信号或数据。WT对于分析具有非平稳特性的信号特别有用。	小波变换与SAE和LSTM混合模型相结合、将小波变换与基于CNN的模型相结合，使用扩张残差卷积神经网络 (DRCN) 、将聚类、特征选择和经验小波变换(EWT)集成到基于LSTM的框架中，分解和预测
			经验模态分解(EMD)	EMD（经验模态分解）是一种完全自适应的信号分解算法；EEMD（集成经验模态分解）是 EMD 的改进版本，它将随机性引入分解过程中以提高鲁棒性和减少模式混合；CEEMD（具有自适应噪声的完整集成经验模态分解）是对EEMD的进一步改进，将自适应降噪技术融入到分解过程中；变分模态分解 (VMD)是一种时间序列分析中常用的信号分解方法	将EMD与GRU、CNN和BiLSTM相结合、应用EEMD对LSTM的输入数据进行分解、将CEEMD与deep-TCN相结合、结合CEEMD、BP、LSTM和ARIMA构建了一个综合混合框架、VMD和BiLSTM、将多个DL模型(LSTM ESN和TCN)与统计算法(VMD和GBDT)相结合
		传统机器学习方法		机器学习的特征可解释，特征通常是手工标注的，以帮助提高模型的可解释性。DL特征学习能力强大，处理大规模、高维数据方面具有优势，但缺乏模型的可解释性。	DNN和DBSCAN聚类算法相结合、基于RF和BiLSTM、集成了卡尔曼滤波注意机制和LSTM神经网络、基于回声状态网络(ESN)和粒子群算法、基于LSTM、RF和粒子群算法的模型、Q-learning强化学习与GCN-LSTM-GRU深度学习模型
深度神经网络集成方法	CNN + LSTM			利用CNN获取输入数据的特征，然后利用LSTM对特征序列的时间依赖性进行建模	基于resnet的CNN-LSTM模型、基于图卷积网络 (GCN) 和 LSTM 的混合、基于注意力的CNNLSTM模型、由ANN、LSTM和CNN组成的模型ST-DNN、结合长短期记忆网络、卷积神经网络和随机森林算法的EDPF模型
	CNN + GRU			cnn能够从原始像素数据中学习分层特征。cnn可以用于时间序列数据，以学习值的时间序列模式，例如每日或每小时的污染物浓度。gru是一种循环神经网络(rnn)，gru能够有选择地更新和重置他们的内部状态，使他们更有效地学习长期依赖。	1D CNN和BiGRU以捕获本地和时间模式、3D CNN与GRUs、基于堆叠自编码器(AE)、CNN和GRU的混合深度学习模型

Table 5
The research used DL plus deterministic methods.

Study	Year	Location	Model	Time step	RMSE ($\mu g\ m^{-3}$)	MAE ($\mu g\ m^{-3}$)	MAPE (%)	R^2
Chang-Hoi et al. [76]	2021	South Korea	RNN-CMAQ	H/S/T+6	-	5.10	-	0.862
Sun et al. [77]	2021	China	LSTM-WRF-CMAQ	H/S/T+48	11.03	8.08	68.83	0.884

Table 6
The description of statistical methods.

Category	Model	Details
Classic method	ARIMA	Autoregressive integrated moving average model
	WT	Wavelet transform
	EWT	Empirical wavelet transforms
	EMD	Empirical mode decomposition
	EEMD	Ensemble empirical mode decomposition
	CEEMD	Complementary ensemble empirical mode decomposition
	VMD	Variational mode decomposition
Machine learning	RF	Random forest
	SVR	Support vector machine-based regression
	GBDT	Gradient boosting decision tree
	ANN	Artificial neural networks
	PSO	Particle swarm optimization
	DBSCAN	Density-based spatial clustering of applications with noise
	FE	Fuzzy entropy
	Kalman-filter	Kalman filter
	GWO	Grey wolf optimizer
	mRMR	Max-relevance and min-redundancy
	Q	Q-learning

Table 7
The research used DL plus statistical methods.

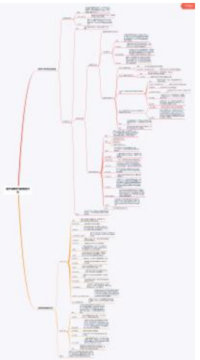
Study	Year	Location	Model	Time step	RMSE (μg m^{-3})	MAE (μg m^{-3})	MAPE (%)	R^2
Qiao et al. [80]	2019	China	WT-SAE-LSTM	D/S/T+1	-	3.88	-	-
Huang et al. [83]	2021	Beijing, China	EMD-GRU	H/M/T+(1-4)	11.37	6.53	25.79	0.98
Jin et al. [84]	2020	Beijing, China	EMD-CNN-GRU	H/M/T+(1-24)	42.26	34.95	65.30	0.67
Zaini et al. [85]	2022	Cheras/Batu Muda, Malaysia	EEMD-LSTM	H/S/T+1	4.21/4.89	2.81/2.77	14.15/14.64	0.97/0.96
Zhang et al. [86]	2021	Beijing, China	VMD-BiLSTM	H/S/T+1	9.39	5.35	16.40	0.99
Chang et al. [87]	2020	Taiwan, China	GBDT-SVR-LSTM	H/S/T+1	7.67	5.00	-	-
Liu et al. [88]	2021	Changsha, China	GCN-LSTM-GRU-Q	M/-/-	17.63	14.24	2.91	-
Liu et al. [89]	2020	Shanghai, China	CEEMD-LSTM	H/S/T+3	3.28	2.23	5.74	0.99
Jiang et al. [90]	2021	Beijing, China	CEEMD + DeepTCN	H/S/T+1	1.11	0.65	2.65	-
Kim et al. [82]	2021	Beijing, China	FC-DTWD-EWT-CBLSTM	H/S/T+1	2.29	1.51	4.03	0.94
				H/S/T+10	5.17	3.37	8.98	0.85
Lu et al. [91]	2021	Yangtze River Delta Region, China	DBSCAN-DNN	H/S/T+1	13.29	-	0.90	-
Teng et al. [92]	2022	Shanghai, China	EMD-SE-BiLSTM	H/S/T+1	2.77	1.88	-	0.98
				H/S/T+3	5.04	3.56	-	0.95
Fu et al. [93]	2021	Hangzhou, China	CEEMD-LSTM	H/S/T+1	6.48	4.76	15.76	-
Zhang et al. [94]	2020	Gansu, China	ESN-PSO	H/S/T+1	8.73	5.47	8.20	0.93
Wang et al. [95]	2022	China	LSTM-RF-PSO	H/S/T+1	4.93	2.91	24.36	-
Wang et al. [96]	2022	-	CEEMD-FE-mRMR-GWO-LSTM	H/-/-	8.26	6.60	19.77	0.95
Wei Sun et al. [97]	2022	Jing-Jin-Ji Region, China	LSTM-CEEMADN	D/S/T+1	3.52	2.73	-	0.97
Xu et al. [98]	2022	China	CEEMD-CNN-LSTM	H/S/T+2	12.67	9.60	-	0.87
Zhou et al. [99]	2022	Chongqing, China	Kalman-Filter-LSTM	H/S/T+1	8.45	7.30	-	0.96
Zhao et al. [100]	2022	Beijing/Guangzhou, China	RF-BiLSTM	H/S/T+1	7.26/1.77	3.73/1.33	-	1.00/0.99
		Xi'an/Shenyang, China			2.75/4.51	1.37/2.20	-	1.00
Zhang et al. [101]	2023	Pingqiao/South bay/Brewing, China	CEEMD-FCN-LSTM	-	3.81/5.39/4.02	2.47/2.81/4.55	4.48/6.48/2.37	0.98/0.97/0.98
Masood et al. [102]	2023	Delhi, India	ANN	-	24.12	-	-	0.94
Liu et al. [103]	2022	Shenyang/Changsha/Shenzhen, China	VMD-LSTM-ESN-TCN-GBDT	-	1.98/2.20/1.68	1.58/1.71/1.32	3.95/4.11/4.53	-
Benhaddi et al. [81]	2021	Marrakesh, Morocco	WT-CNN	H/-/-	0.01	-	99.10	-
Ban et al. [104]	2022	Hangzhou, China	CEEMD-LSTM-BP-ARIMA	D/S/T+1	4.55	3.66	-	0.79
M.A.A.A.-q et al. [105]	2021	Wuhan, China	PSO-SMA-ANFIS	H/S/T+1	22.39	17.50	16.83	0.51

Table 8
The research used CNN-LSTM methods.

Study	Year	Location	Model	Time step	RMSE (μg m^{-3})	MAE (μg m^{-3})	MAPE (%)	R^2
Huang et al. [106]	2018	Beijing/Shanghai, China	CNN-LSTM	H/S/T+1	24.22	14.63	-	-
Qin et al. [107]	2019	Shanghai, China	CNN-LSTM	H/S/T+24	14.30	-	-	-
Li et al. [108]	2020	Beijing, China	CNN-LSTM	D/S/T+1	18.99	16.81	-	-
Zhang et al. [109]	2020	Shijiazhuang, China	CNN-LSTM	H/S/T+1	14.94	-	-	-
Yang et al. [110]	2021	Beijing, China	CNN-LSTM	H/S/T+1	19.09	-	-	0.92
Wei et al. [111]	2021	Beijing, China	CNN-LSTM	H/S/T+6	-	19.54	-	0.62
Bekkar et al. [112]	2021	Beijing, China	CNN-LSTM	D/S/T+1	12.92	6.74	-	0.98
Wardana et al. [113]	2021	Beijing, China	CNN-LSTM	H/-/T+1	15.26	8.77	-	-
Tsokov et al. [114]	2022	Beijing, China	CNN-LSTM	H/S/T+1	14.95	8.48	-	-
Teng et al. [92]	2022	Beijing, China	CNN-LSTM	H/S/T+1	8.93	6.52	-	0.92
Kim et al. [115]	2022	South Korea	CNN-LSTM	H/S/T+1	10.52	-	-	0.37
Shao et al. [116]	2022	Seoul, South Korea	SCNN-LSTM	H/M/T+(1-10)	8.05	5.04	23.96	0.70
Choi et al. [117]	2022	Beijing, China	ResNet-LSTM	H/S/T+1	0.02	0.01	9.02	-
Zhang et al. [118]	2022	Yangtze River Delta Region, China	ResNet-LSTM	H/S/T+1	5.47	3.89	-	-
Cheng et al. [119]	2022	Beijing, China	SResCNN-LSTM	D/S/T+5	40.67	23.74	-	0.80
Zhao et al. [120]	2019	Beijing/Tianjin, China	STCNN-LSTM	H/S/T+6	19.36	15.53	26.00	0.70
Qi et al. [121]	2019	Jing-Jin-Ji Region, China	GCNN-LSTM	H/S/T+1	22.41	13.72	-	-
Soh et al. [123]	2018	Taiwan/Beijing, China	ANN-CNN-LSTM	H/S/T+6	-	-	-	-
Yang et al. [124]	2019	Beijing, China	DWFD-CNN-LSTM	H/M/T+(1-6)	43.90	29.17	-	-
Li et al. [122]	2020	Taiyuan, China	Attention-CNN-LSTM	H/M/T+(1-24)	14.83	8.98	-	0.99
Li et al. [125]	2022	Beijing, China	CBAM-CNN-BiLSTM	H/M/T+(13-18)	31.47	21.86	-	0.81
				H/M/T+(25-48)	32.34	22.30	-	0.79
Moursi et al. [126]	2022	Beijing, China	NARX-CNN-LSTM	H/S/T+1	23.64	-	-	0.92
Zhu et al. [127]	2023	Shanghai, China	1D-CNN + BiLSTM	H/S/T+1	3.88	2.52	-	0.94
Pak et al. [128]	2020	Beijing, China	PM predictor	D/S/T+1	2.99	2.21	3.90	-
Du et al. [129]	2021	Beijing, China	DAQFF	H/S/T+1	8.20	6.19	-	-
Zhu et al. [130]	2021	Jing-Jin-Ji Region, China	APNet	H/-/T+1	17.93	9.93	-	0.95
				H/-/T+72	29.11	20.07	-	0.87
Zhang et al. [131]	2022	Hong Kong/Beijing, China	Deep-AIR	H/S/T+1	-	-	21.10/23.90	-
Mohan et al. [132]	2022	Kerala, India	EDPF	H/M/T+24	12.96	9.28	56.73	0.44
Li et al. [133]	2022	Beijing, China	FPHFA	H/M/T+(1-12)	28.15	19.19	56.10	0.87
				H/M/T+(13-24)	22.12	15.27	43.80	0.93
Gunasekar et al. [134]	2022	Chennai, Tamandu	ARTOCL	NA	0.50	0.32	-	0.69

Table 9
The research used CNN + GRU methods.

Study	Year	Location	Model	Time step	RMSE (μg m^{-3})	MAE (μg m^{-3})	MAPE (%)	R2
Tao et al. [135]	2019	Beijing, China	CBGRU	H/S/T+2	14.53	10.47	34.09	-
Zhang et al. [136]	2020	Lanzhou, China	MTD-CNN-GRU	H/S/T+1	7.96	4.54	-	-
Faraji et al. [137]	2022	Tehran, Iran	3D CNN-GRU	H/S/T+1	-	-	-	0.84
				D/S/T+1	-	-	-	0.78
Chiang et al. [138]	2021	Taiwan, China	AE + CNN + GRU	D/S/T+1	5.03	3.10	-	-
Mao et al. [139]	2022	Taiwan, China	CNN + GRU	H/S/T+1	4.78	3.56	-	0.89
		Kennedy/Simon Bolivar, US		D/S/T+1	6.83/6.15	5.29/4.58	-	0.44/0.56



分析

Table 10
Description of the proposed indicators.

Standard	Indicator		Description
Dataset	Open source		Given the available dataset link or declare the process of data collection.
	Data feature	Predict step	The step size of the prediction task, i.e., single step or multi-step.
		Time resolution	The time resolution of dataset, i.e., hourly, daily, or monthly.
		Data size	The size of the using dataset.
	Data dimensions		Multiple data inputs include meteorological data.
	Dataset split		The division of training, valid, and testing set. The test set needs to contain all kinds of samples.
	Pre-processing	Normalize	Min-max scaling, z-score normalization, and decimal scaling, etc.
		Missing value	Describes how to handle missing or outlier values to ensure data continuity.
Method	Open source		Provide a link where the code will be available.
	Architecture		Whether to describe the network structure and give the parameters.
	Training process		The design or trend of loss or the learning objectives.
	Visual analysis		Visual visualization of predicted and ground truth.
	Novelty		Whether to innovate or apply the model to a domain for the first time.
Experiments	Experimental setting	Model config	The setting of design parameters, such as the convolution kernel size.
		Computation setup	Basic information about the used CPU and GPU in the experiment.
	Results metrics		The evaluation metrics, usually RMSE, MAE, and MAPE, also have SSIM, ACC, R^2, etc.
	Modeling metrics	FLOPs	The Floating Point Operations (FLOPs).
		Params	The number of trainable parameters.
	Comparison with SOTAs		The results are compared with the advanced algorithm under the same experimental setting.
	Ablation study		Removing or disabling different components or features to see how it affect the model's performance.

结论

本文对基于dl的PM2.5浓度预测体系结构进行了综述。严格按照PRISMA指南进行筛选了118篇论文。通过论文中的方法进行分类，我们以表格形式呈现了各种模型的结果。通过对基于dl的模型的详细分类，我们对各种PM 2.5预测方法的性能指标和应用条件进行了严格的检查和综合，我们的分析深入探讨了这些模型的优缺点。此外，我们引入了一个新的评估框架DMES，专门用于评估和标准化对类似主题的文章。该框架代表了在增强基于dl的研究论文的一致性和可比性方面迈出的重要一步，最终促进了更可靠和公平的评估。

谢谢