Methods and Applications

Advancing Underlying Cause of Death Inference Through Wide and Deep Model

Xin Fang[1]; Shaofen Huang[1]; Yanrong Yin[1]; Tiehui Chen[1]; Zhijun Liao[2,#]; Wenling Zhong[1,#]

ABSTRACT

Introduction: Accurately filling out death certificates is essential for death surveillance. However, manually determining the underlying cause of death is often imprecise. In this study, we investigate the Wide and Deep framework as a method to improve the accuracy and reliability of inferring the underlying cause of death.

Methods: Death report data from national-level cause of death surveillance sites in Fujian Province from 2016 to 2022, involving 403,547 deaths, were analyzed. The Wide and Deep embedded with Convolutional Neural Networks (CNN) was developed. Model performance was assessed using weighted accuracy, weighted precision, weighted recall, and weighted area under the curve (AUC). A comparison was made with XGBoost, CNN, Gated Recurrent Unit (GRU), Transformer, and GRU with Attention.

Results: The Wide and Deep achieved strong performance metrics on the test set: precision of 95.75%, recall of 92.08%, F1 Score of 93.78%, and an AUC of 95.99%. The model also displayed specific F1 Scores for different cause-of-death chain lengths: 97.13% for single causes, 95.08% for double causes, 91.24% for triple causes, and 79.50% for quadruple causes.

Conclusions: The Wide and Deep significantly enhances the ability to determine the root causes of death, providing a valuable tool for improving cause-of-death surveillance quality. Integrating artificial intelligence (AI) in this field is anticipated to streamline death registration and reporting procedures, thereby boosting the precision of public health data.

The surveillance of causes of death in populations is vital for public health decision-making (*1*). A crucial aspect of this surveillance is accurately completing death certificates. Challenges such as inadequate medical knowledge, unfamiliarity with cause-of-death inference rules, and limited practical experience can result in inconsistent quality of cause-of-death registration. Quality assessments have revealed that the accuracy of manually deduced underlying causes of death varies from 55% to 84% (*2*–*3*).

In recent years, there has been a focus on utilizing artificial intelligence (AI) techniques to automatically identify the primary causes of death. Despite deep neural network models achieving 97.8% accuracy in this task (*4*), their practical applications are restricted by performance limitations. This study employs the Wide and Deep framework to enhance the accuracy and stability of deep learning models, aiming to better predict the underlying causes of death and improve cause of death surveillance.

METHODS

The data set utilized in our analysis comprises mortality records collected through nation-wide cause-of-death surveillance systems in Fujian Province, spanning from 2016 to 2022 (Supplementary Table S1, available at https://weekly.chinacdc.cn/). We employed the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10), to systematically categorize the causes of death into 25 broad classifications, while preserving the specificity of 4-digit codes. Any records with inconsistencies such as inaccuracies within the cause-of-death sequence, incorrect deduction of the primary cause of death, ICD-10 coding mistakes as identified by specialists at the Fujian Provincial CDC, or missing information regarding the deceased's sex and age, were omitted from our study. Upon the elimination of redundant entries via cross-referencing identity numbers, names, and dates of demise, the final cohort entailed 403,547 individual death records. Identifying particulars were redacted, with the preservation of birth and death dates, sex, the complete ICD-10 codes for the causes of death, and the duration between disease

onset and mortality. Analyses focused on the primary cause of death codes.

Age was divided into four groups: ≤14 years, 15–44 years, 45–64 years, and ≥65 years. Label encoding was applied to classify all ICD-10 codes, age groups, and genders into integer classes. The durations between illnesses were determined, with any missing durations recorded as null; the collected data was normalized using z-score normalization.

In light of considerable variation in the prevalence of different primary causes of death (Supplementary Table S2, available at https://weekly.chinacdc.cn/), we utilized a subsampling strategy for the over-represented categories C and I, whereby 40,000 cases were chosen at random. Conversely, categories with fewer instances underwent up sampling. We constructed a multi-directed graph using the cause-of-death sequences extracted from our data. This graph featured causes as nodes and incorporated both their sequential relationships and the temporal intervals between successive diseases as edge attributes. From each category, random samples of primary causes of death, along with other associated causes within their sequence, were selected. We used the graph to reconstruct cause-of-death sequences with lengths varying from one to four causes. To ensure uniformity across all categories classified by ICD-10 codes, we adjusted the sample size to 40,000 cases per category through oversampling, creating a research dataset comprising one million cases. We allocated 10% of the original dataset for each category to create a test set, ensuring that the number of samples from each category did not surpass 4,000 (Supplementary Table S2). The data remainder, after the test set extraction, served for model training and validation, applying a 5-fold cross-validation technique (Supplementary Figure S1, available at https://weekly.chinacdc.cn/). Notably, statistical analysis revealed significant disparities in both gender and age profiles between the test set and the training dataset (Supplementary Table S3, available at https://weekly.chinacdc.cn/).

The Wide and Deep model integrates a linear component, Wide, for memorization, and a neural network, Deep, for generalization. It incorporates a Convolutional Neural Networks (CNN) into the deep component to detect complex patterns in structured data while maintaining essential rule-based patterns. This model aims to accurately predict causes of death by balancing linear and non-linear interactions of features (Figure 1). It is well-suited for tasks with categorical and continuous data and has shown efficacy in recommendation systems (*5*).

The research compared the performance of various models including XGBoost, CNN, Gated Recurrent Unit (GRU), Transformer, and GRU with an attention mechanism (Attention GRU) for the task of inferring the underlying cause of death. Python 3.11.5 and PyTorch 2.1.0 were used for analysis. The evaluation on the test set was based on metrics such as weighted accuracy, precision, recall, F1-Score, and area under the curve (AUC). The outcomes of the model tests were presented as means and standard deviations ($\bar{x}\pm s$).

RESULTS

The Wide and Deep model outperformed other models, attaining a recall of 92.08, an F1 score of 93.78, and an AUC of 95.99. XGBoost demonstrated superior precision to Wide and Deep, while CNN showed a comparable AUC. However, their recall and F1 scores were slightly lower compared to Wide and Deep. The performance difference between the other models and Wide and Deep was notable (Table 1).

The Wide and Deep model consistently demonstrated superior and stable performance across all chain lengths, exhibiting the highest accuracy, precision, recall, and F1 score, notably at chain length 2 (accuracy: 92.62, precision: 97.11, recall: 93.91, F1 score: 95.08). However, the model's performance slightly declined with longer chain lengths (Table 2).

In the original dataset, the ICD-10 categories I, C, J, W, and E represented 83.13% of the causes of death (Supplementary Table S2), and Wide and Deep demonstrated effective predictive performance for these common categories (Figure 2 and Supplementary Table S4–S7, available at https://weekly.chinacdc.cn/). The models achieved the highest accuracy for categories I (96.72%), J (97.92%), and E (94.00%) compared to other models, with the highest F1 scores for all these categories (I: 93.07, C: 98.42, J: 97.62, W: 89.42, E: 92.72). Wide and Deep accurately identified the underlying cause of death, even in cases of incorrect numerical codes (Supplementary Figure S2, available at https://weekly.chinacdc.cn/).

DISCUSSION

This study illustrates the effectiveness of the Wide and Deep framework-based deep learning model for predicting potential causes of death. The model

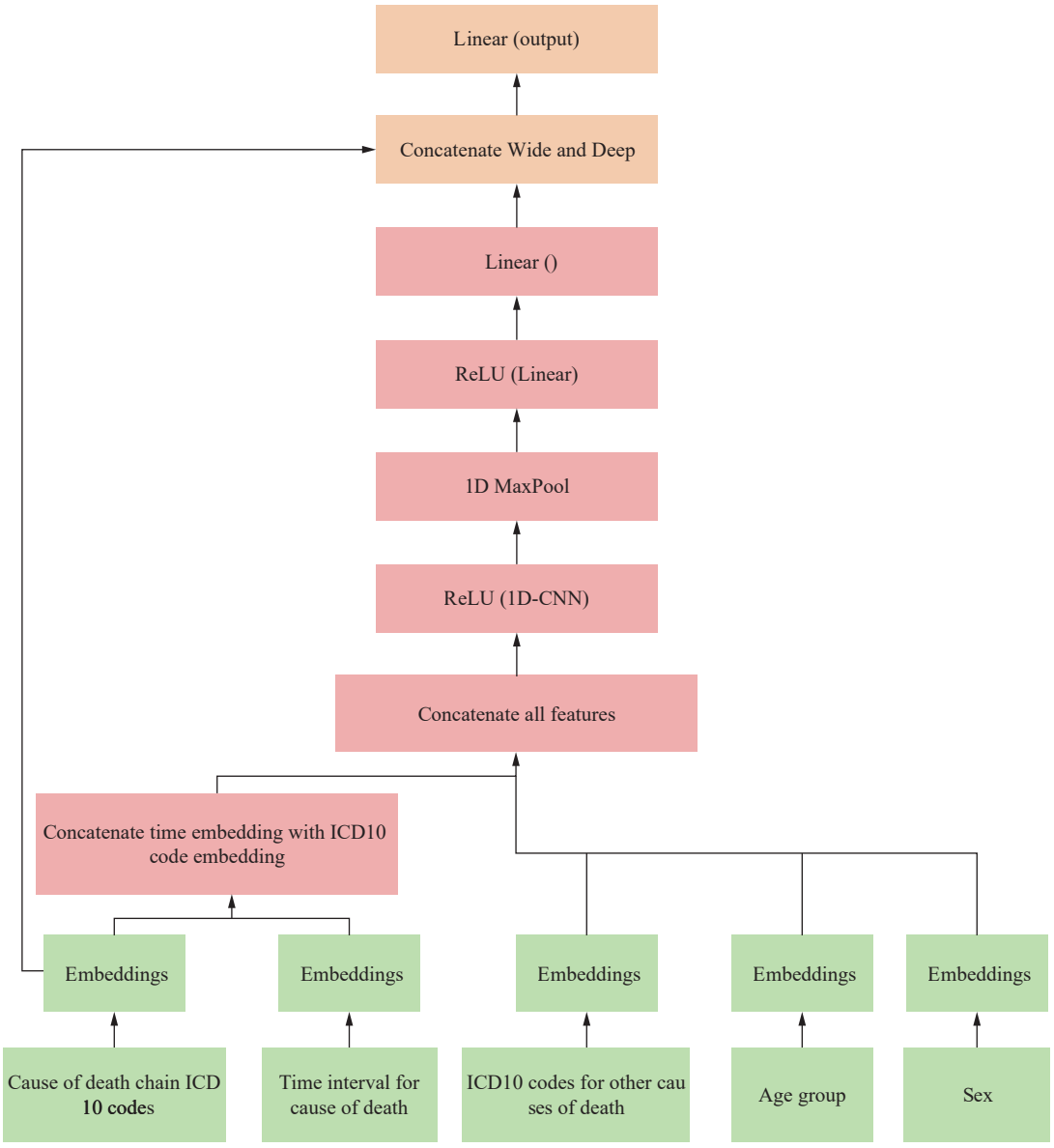


FIGURE 1. Schematic diagram of the Wide and Deep model structure.
Abbreviation: ICD=International Classification of Diseases; CNN=Convolutional Neural Networks.

surpasses manual inference levels at Chinese national surveillance points (*3*,*6*), showing practical applicability with a weighted F1 Score of 93.78 and a weighted AUC of 95.99.

Previously, the China CDC collaborated with the United States to develop a rule-based automated coding tool for determining the underlying cause of death (*7*), achieving an accuracy of 84.8% (*8*). In 2018, US researchers employed sequence rule mining to create a common cause of death inference model, with an error rate of 20.1% compared to human-expert determined death certificates (*9*). In a separate study in 2020, researchers from France and Italy utilized a deep neural network for inferring the underlying cause of death, achieving over 97% accuracy (*10*). However, due to country-specific variations, validating the model for cross-country applicability remains an ongoing research challenge (*4*). Using an attention mechanism, the model exhibited an accuracy range of 80.9% to 81.7% (*11*). Researchers from the Beijing Institute of Technology and China CDC developed a hybrid inference model with the Sink-CF algorithm, improving precision and recall for determining the fundamental cause of death to 93.8% and 90.1%, respectively (*12*).

Understanding the intricate protocols of determining the cause of death necessitates extensive medical expertise. Traditional training involves

TABLE 1. Comparative performance metrics of deep learning models in cause-of-death analysis.

Model	Precision	Recall	F1 Score	Weighted AUC
XGBoost	96.52±0.08	73.02±0.22	82.68±0.16	86.34±0.11
CNN	93.65±0.34	89.62±0.40	91.38±0.36	94.75±0.20
GRU	94.07±0.37	87.29±0.57	89.62±0.40	93.60±0.29
Transformer	94.06±1.21	83.93±0.81	87.47±1.10	91.90±0.41
Attention GRU	94.34±0.56	87.63±0.35	90.06±0.46	93.76±0.18
Wide and Deep	95.75±0.06	92.08±0.10	93.78±0.09	95.99±0.05

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit; AUC=area under the curve.

TABLE 2. Model performance across different lengths of cause-of-death chains.

Length of chain	Model	Accuracy	Precision	Recall	F1 Score
Length=1	XGBoost	67.24±0.73	97.65±0.19	74.75±0.61	83.51±0.51
Samples=5,281	CNN	90.22±1.20	97.61±0.32	96.50±0.21	97.26±0.26
	GRU	86.28±1.44	97.31±0.96	93.55±0.52	95.61±0.61
	Transformer	83.10±1.63	97.56±1.23	92.21±0.89	94.14±1.14
	Attention GRU	87.23±0.93	97.94±0.59	94.21±0.55	95.46±0.46
	Wide and Deep	92.90±0.20	98.63±0.21	97.34±0.07	97.13±0.13
Length=2	XGBoost	72.53±0.41	97.04±0.11	70.97±0.19	81.16±0.16
Samples=9,450	CNN	89.63±0.56	94.70±0.36	91.38±0.48	92.43±0.43
	GRU	86.50±0.66	94.27±0.32	88.03±0.65	90.47±0.47
	Transformer	82.76±1.60	95.09±1.37	85.05±1.28	89.35±1.35
	Attention GRU	87.02±0.40	94.98±0.82	88.72±0.40	91.52±0.52
	Wide and Deep	92.62±0.12	97.11±0.07	93.91±0.09	95.08±0.08
Length=3	XGBoost	68.76±0.20	96.24±0.15	74.86±0.29	83.22±0.22
Samples=6,095	CNN	84.81±0.88	91.60±0.45	85.20±0.52	87.34±0.34
	GRU	83.75±0.52	93.42±0.72	84.10±0.60	87.37±0.37
	Transformer	80.86±0.65	93.72±0.97	79.86±0.25	84.73±0.73
	Attention GRU	84.22±0.69	93.48±0.61	84.22±0.42	87.48±0.48
	Wide and Deep	88.53±0.35	94.20±0.14	88.49±0.30	91.24±0.24
Length=4	XGBoost	56.32±2.39	93.75±0.33	72.30±0.42	79.33±0.33
Samples=1,384	CNN	75.07±1.40	83.51±1.20	70.84±1.08	75.95±0.95
	GRU	79.63±0.83	88.44±0.52	72.31±1.12	76.80±0.80
	Transformer	75.27±2.13	86.04±1.46	62.64±2.27	63.84±1.84
	Attention GRU	76.17±0.33	86.00±0.77	70.06±0.79	73.40±1.40
	Wide and Deep	81.00±0.49	86.66±0.76	75.33±0.55	79.50±0.50

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit.

learning these protocols through practice over time to enhance proficiency in completing death certificates (*13*). Conversely, an AI-based model offers a cost-effective solution for grassroots personnel. Basic computer skills enable staff to utilize the chain-of-cause-of-death data for inferring the underlying cause of death.

Compared to prior studies, our model enhances the overall performance in the task of cause-of-death speculation. The study was carried out using a test dataset that mirrored real-world data. Detailed performance results for different lengths of cause-of-death chains and ICD-10 classifications are presented. A comparative analysis with XGBoost, CNN, GRU, Transformer, and Attention GRU models showed that Wide and Deep achieved the best performance.

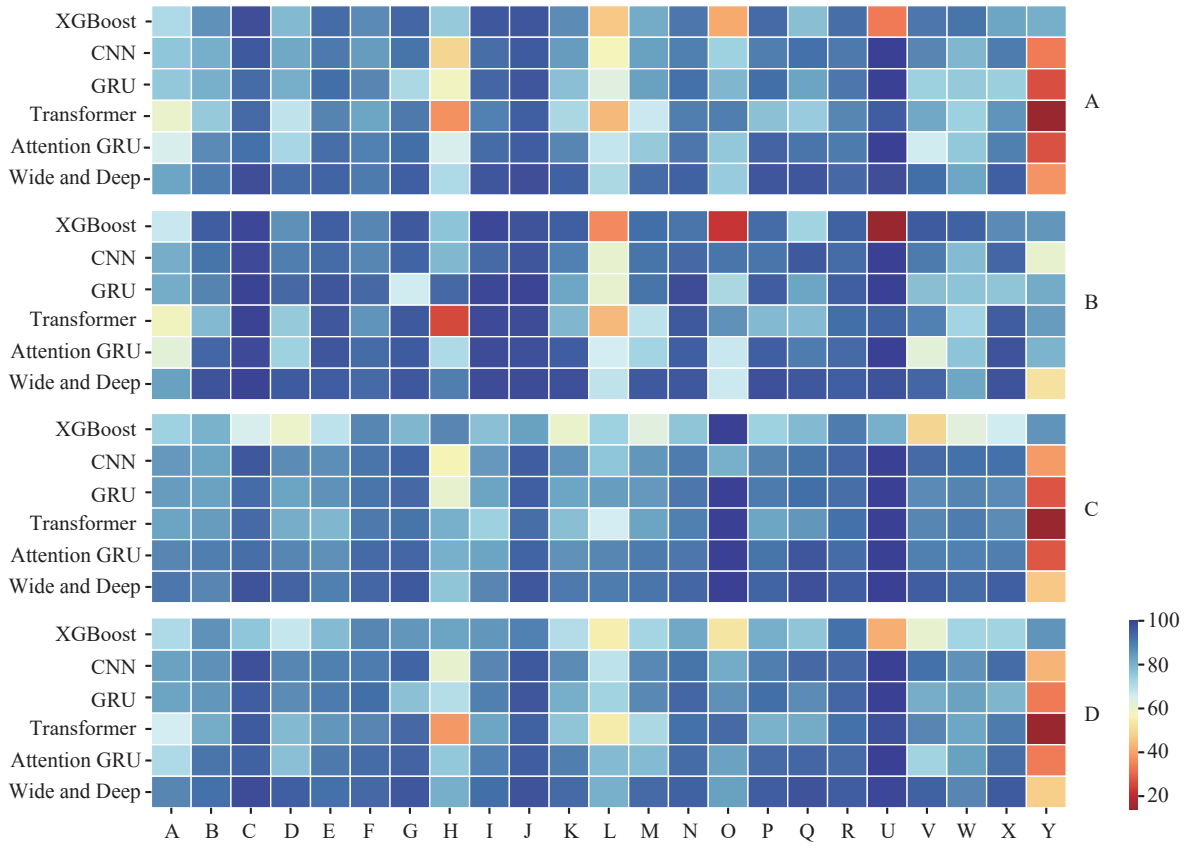


FIGURE 2. Comparison of the performance of models for the underlying cause of death speculation. (A) Accuracy; (B) Precision; (C) Recall; (D) F1 Score.
Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit.

Additionally, the composition of cause-of-death types in the test dataset closely resembles the actual scenario in Fujian Province (*14*), offering a test setting that aligns with real-world data.

However, this study is subject to some limitations. First, a notable dissimilarity exists in the distribution of primary causes of death between the original dataset, which predominantly includes diseases related to the circulatory, neoplasms, and respiratory systems. To address this issue, a technique of up sampling and subsampling was employed to balance the training data. Second, while the training data originate from scrutinized records of national mortality surveillance sites, and though these records have undergone expert verification, certain chains of causes of death may still be illogical, potentially impacting the efficacy of model training.

CONCLUSION

The developed cause-of-death inference model in this study demonstrated superior performance, suggesting that the deep learning model utilizing the Wide and Deep framework has the potential to enhance the accuracy of cause-of-death surveillance. Employing AI technology is projected to boost the quality of cause-of-death registration reports efficiently and mitigate the burden of manual review, thereby optimizing time and resource allocation compared to traditional training methods for registration staff.

Conflicts of interest: No conflicts of interest.

Funding: Supported by the Fujian Provincial Health Youth Project (2020QNB017), the Fujian Province Pilot Project (2020Y0060), the National Natural Science Foundation of China (62072107), and the Natural Science Foundation of Fujian Province of China (2020J01610).

doi: 10.46234/ccdcw2024.094

[#] Corresponding authors: Wenling Zhong, mbzwl@163.com; Zhijun Liao, liaozhijun@fjmu.edu.cn.

[1] Department for Chronic and Noncommunicable Disease Control and Prevention, Fujian Provincial Center for Disease Control and Prevention, Fuzhou City, Fujian Province, China; [2] Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Fujian Medical University, Fuzhou City, Fujian Province, China.

Submitted: February 05, 2024; Accepted: May 13, 2024

REFERENCES

1. Brolan CE, Gouda HN, Abouzahr C, Lopez AD. Beyond health: five global policy metaphors for civil registration and vital statistics. Lancet 2017;389(10074):1084 – 5. https://doi.org/10.1016/S0140-6736(17)30753-5.
2. Qaddumi JAS, Nazzal Z, Yacoup ARS, Mansour M. Quality of death notification forms in North West Bank/Palestine: a descriptive study. BMC Res Notes 2017;10(1):154. https://doi.org/10.1186/s13104-017-2469-0.
3. He Q, Liu ZR, Chen YJ, Xing XY, Li R. Sampling audit evaluation on quality of the network reporting death data in Anhui national disease surveillance points from 2013 to 2017. Anhui J Prev Med 2019;25(3):165-9. https://d.wanfangdata.com.cn/periodical/ahyfyx201903002. (In Chinese).
4. Falissard L, Morgand C, Roussel S, Imbaud C, Ghosn W, Bounebache K, et al. A deep artificial neural network-based model for prediction of underlying cause of death from death certificates: algorithm development and validation. JMIR Med Inform 2020;8(4):e17125. https://doi.org/10.2196/17125.
5. Ma YH, Jiang JT, Dong S, Li CM, Yan XY. Book recommendation model based on wide and deep model. In: Proceedings of the 2021 IEEE international conference on artificial intelligence and industrial design (AIID). Guangzhou, China: IEEE. 2012; p. 247-54. http://dx.doi.org/10.1109/AIID51893.2021.9456524.
6. Xuan SL, Di XJ, Li SF, Yang WJ, Kang K, Ma SW. Evaluation on the reliability of mortality surveillance system in Henan from 2015 to 2017. Henan J Prev Med 2021;32(10):781 – 4. https://doi.org/10.13515/j.cnki.hnjpm.1006-8414.2021.10.013.
7. Lu TH. Using ACME (Automatic Classification of Medical Entry) software to monitor and improve the quality of cause of death statistics. J Epidemiol Community Health 2003;57(6):470-1. https://jech.bmj.com/content/57/6/470.info.
8. Ji YB, Wang LJ, Zhou MG. Analysis on coding examples of automated coding software on underlying death cause in death surveillance. Chin J Dis Control Prev 2013;17(9):813-7. http://qikan.cqvip.com/Qikan/Article/Detail?id=47346254. (In Chinese).
9. Hoffman RA, Venugopalan J, Qu L, Wu H, Wang MD. Improving validity of cause of death on death certificates. ACM BCB 2018;2018:178-83. https://pubmed.ncbi.nlm.nih.gov/32558825/.
10. Mea VD, Popescu MH, Roitero K. Underlying cause of death identification from death certificates via categorical embeddings and convolutional neural networks. In: Proceedings of the 2020 IEEE international conference on healthcare informatics (ICHI). Oldenburg, Germany: IEEE. 2020; p. 1-6. http://dx.doi.org/10.1109/ICHI48887.2020.9374316.
11. Zhu YD, Sha Y, Wu H, Li M, Hoffman RA, Wang MD. Proposing causal sequence of death by neural machine translation in public health informatics. IEEE J Biomed Health Inform 2022;26(4):1422 – 31. https://doi.org/10.1109/JBHI.2022.3163013.
12. Yang X, Ma HS, Gao KY, Ge H. An automated method of causal inference of the underlying cause of death of citizens. Life 2022;12(8):1134. https://doi.org/10.3390/LIFE12081134.
13. Hart JD, Sorchik R, Bo KS, Chowdhury HR, Gamage S, Joshi R, et al. Improving medical certification of cause of death: effective strategies and approaches based on experiences from the Data for Health Initiative. BMC Med 2020;18(1):74. https://doi.org/10.1186/s12916-020-01519-8.
14. Lin XQ, Zhong WL, Li WY, Huang SF, Zhu Y, Yin YR. Study on the changes of death cause spectrum and the loss of life expectancy in Fujian province in 2015. Chronic Pathematol J 2019;20(12):1795 – 8. https://doi.org/10.16440/j.cnki.1674-8166.2019.12.011.

SUPPLEMENTARY MATERIAL

Method for Completing the Death Certificate

Part I of the cause of death: Fill in the disease or condition that directly caused the death, which is mandatory for every death.

Severe illnesses, injuries, or complications leading directly to death should be recorded in row (a). Subsequently, starting from line (b), preceding factors that potentially contributed to the condition in row (a) or the preceding line must be documented until the primary cause is identified, creating a logical sequence.

e.g., (d)→(c)→(b)→(a)

Part II on causes of death complements the contents of Part I.

Non-fatal comorbid conditions that could influence mortality should be documented. All diagnosed chronic illnesses like mental disorders, diabetes, hypertension, tumors, and coronary artery disease must be recorded, prioritized by severity without a specific limit. If no cause is identified, it should not be listed.

The primary cause of death is the first disease or injury that led to all subsequent conditions listed on the death certificate. Determining the primary cause follows the information provided in Parts I and II of the death certificate and complies with the coding regulations established by the World Health Organization (WHO).

SUPPLEMENTARY TABLE S1. Medical certificate of death (presumption) for Chinese residents.

Name:	Age:	Sex:
Diagnosis of major diseases causing death	Disease Name	Estimated time between onset of illness and death
Ⅰ. (a) Direct cause of death		
(b) Disease or condition causing (a)		
(c) Disease or condition causing (b)		
(d) Disease or condition causing (c)		
Ⅱ. Other disease diagnoses (other significant conditions that contributed to the death, but were not related to causing the death)		
Underlying cause of death:		ICD-10 Codes:

SUPPLEMENTARY TABLE S2. Distribution of ICD-10 classification in the original dataset and test dataset.

ICD10 classification	Original dataset		Test dataset	
	Count (*n*)	Ratio (%)	Count (n)	Ratio (%)
I (Diseases of the circulatory system)	140,895	34.91	4,000	18.0
C (Neoplasms)	120,440	29.85	4,000	18.0
J (Diseases of the respiratory system)	38,386	9.51	3,838	17.3
W (External causes of morbidity and mortality)	22,283	5.52	2,228	10.0
E (Endocrine, nutritional and metabolic diseases)	13,468	3.34	1,346	6.1
R (Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified)	13,385	3.32	1,338	6.0
K (Diseases of the digestive system)	9,204	2.28	920	4.1
G (Diseases of the nervous system)	7,701	1.91	770	3.5
Y (External causes of morbidity and mortality)	6,887	1.71	688	3.1
V (External causes of morbidity and mortality)	6,594	1.63	659	3.0
X (External causes of morbidity and mortality)	5,746	1.42	574	2.6
N (Diseases of the genitourinary system)	4,289	1.06	428	1.9
F (Mental and behavioral disorders)	3,264	0.81	326	1.5
D (Neoplasms; Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanisms)	2,545	0.63	254	1.1
M (Diseases of the musculoskeletal system and connective tissue)	2,373	0.59	237	1.1
A (Certain infectious and parasitic diseases)	1,807	0.45	180	0.8
B (Certain infectious and parasitic diseases)	1,384	0.34	138	0.6
P (Certain conditions originating in the perinatal period)	1,032	0.26	103	0.5
Q (Congenital malformations, deformations and chromosomal abnormalities)	1,022	0.25	102	0.5
L (Diseases of the skin and subcutaneous tissue)	646	0.16	64	0.3
U (Codes for special purposes)	83	0.02	8	0
H (Diseases of the eye and adnexa; Diseases of the ear and mastoid process)	58	0.01	5	0
O (Pregnancy, childbirth and the puerperium)	47	0.01	4	0
T (Injury, poisoning and certain other consequences of external causes)	7	0	0	0
S (Injury, poisoning and certain other consequences of external causes)	1	0	0	0
Total	403,547	100	22,210	100.0

SUPPLEMENTARY TABLE S3. Sex and age distribution of the training, validation, and test sets.

Groups	Training and validation sets n (%)	Testing sets n (%)	*P* value
Sex			
Male	509,645 (52.1)	9,605 (43.2)	<0.001
Female	468,145 (47.9)	12,605 (56.8)	
Age, years			
≤14	78,262 (8.0)	370 (1.7)	<0.001
15–44	109,343 (11.2)	1,081 (4.9)	
45–64	173,723 (17.8)	3,785 (17.0)	
≥65	616,462 (63.0)	16,974 (76.4)	

Note: Statistical differences between groups were assessed using the chi-square test.

SUPPLEMENTARY TABLE S4. Accuracy of deep learning models for ICD-10 classification of underlying causes of death (%).

Models	A	B	C	D	E	F	G	H
XGBoost ($\bar{x}$±s)	70.92±1.79	85.78±0.30	97.70±0.07	77.77±1.71	92.66±0.33	86.87±0.52	93.67±0.49	74.76±16.97
CNN ($\bar{x}$±s)	75.46±1.63	80.33±2.69	96.06±0.39	81.70±4.25	90.53±0.45	84.06±2.12	91.28±0.94	48.67±11.93
GRU ($\bar{x}$±s)	75.08±2.82	79.78±3.98	92.83±0.56	80.16±3.50	92.19±0.53	88.03±0.39	71.04±28.93	58.00±4.47
Transformer ($\bar{x}$±s)	59.57±28.56	74.77±11.18	92.97±0.96	68.26±4.01	88.39±1.56	82.31±6.23	90.25±1.34	36.80±33.98
Attention GRU ($\bar{x}$±s)	64.85±27.60	87.04±1.93	91.87±1.71	71.73±27.65	92.45±0.38	88.76±0.89	91.96±1.65	64.89±12.80
Wide and Deep ($\bar{x}$±s)	82.22±0.77	89.81±0.57	97.68±0.18	92.59±0.71	94.00±0.57	90.10±1.05	95.20±0.56	70.67±8.94
	I	**J**	**K**	**L**	**M**	**N**	**O**	**P**
XGBoost ($\bar{x}$±s)	96.60±0.19	96.82±0.23	86.72±1.30	46.05±1.47	81.20±1.60	90.79±0.31	40.89±1.99	93.12±0.94
CNN ($\bar{x}$±s)	92.42±0.70	95.81±0.50	83.91±5.26	56.97±7.84	83.30±2.23	88.94±1.06	73.33±18.07	89.23±1.27
GRU ($\bar{x}$±s)	93.77±0.75	96.87±0.40	76.88±19.29	63.13±4.56	83.30±4.00	91.90±0.74	78.67±13.66	92.09±2.45
Transformer ($\bar{x}$±s)	88.57±0.61	95.17±0.69	71.57±18.33	43.69±9.90	66.66±29.11	89.23±1.45	89.33±15.35	76.48±16.43
Attention GRU ($\bar{x}$±s)	92.78±0.31	95.49±0.77	87.69±1.28	67.76±2.67	74.89±31.23	90.64±0.52	75.33±18.50	94.49±4.27
Wide and Deep ($\bar{x}$±s)	96.72±0.17	97.92±0.13	94.71±0.37	71.05±1.70	92.80±1.78	94.89±0.95	74.67±7.30	96.81±1.07
	Q	**R**	**U**	**V**	**W**	**X**	**Y**	
XGBoost ($\bar{x}$±s)	77.18±2.73	92.55±0.22	32.96±4.52	90.59±0.45	91.32±0.12	82.27±0.38	80.26±0.58	
CNN ($\bar{x}$±s)	91.77±1.78	90.38±1.06	100.00±0.00	87.89±1.80	78.48±1.46	89.64±1.04	33.29±2.31	
GRU ($\bar{x}$±s)	82.75±16.46	90.75±0.36	100.00±0.00	73.64±19.35	74.76±1.03	73.92±24.71	25.98±1.24	
Transformer ($\bar{x}$±s)	74.70±13.70	87.82±1.43	95.56±6.09	81.53±1.23	73.38±0.36	85.26±1.22	13.65±0.48	
Attention GRU ($\bar{x}$±s)	91.19±2.23	89.96±0.96	100.00±0.00	65.68±24.21	75.13±0.80	89.09±0.84	26.24±2.46	
Wide and Deep ($\bar{x}$±s)	96.70±1.09	93.58±0.61	97.78±4.97	92.16±0.30	82.10±0.37	95.02±0.09	37.25±1.29	

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit.

SUPPLEMENTARY TABLE S5. Precision of deep learning models for ICD-10 classification of underlying causes of death (%).

Models	A	B	C	D	E	F	G	H
XGBoost ($\bar{x}$±s)	75.38±1.77	96.67±0.88	99.18±0.02	89.59±1.97	96.35±0.33	90.62±0.63	97.02±0.57	82.00±11.93
CNN ($\bar{x}$±s)	85.44±1.28	93.62±2.26	98.94±0.41	91.89±5.12	94.58±0.93	90.84±1.83	95.44±0.82	83.57±25.20
GRU ($\bar{x}$±s)	85.63±3.47	91.25±4.49	99.71±0.15	95.14±3.93	97.60±0.72	95.14±0.88	74.31±31.71	95.00±11.18
Transformer ($\bar{x}$±s)	68.28±33.89	83.36±13.03	99.60±0.16	80.97±7.39	97.37±0.88	88.88±6.86	97.09±1.42	43.56±40.55
Attention GRU ($\bar{x}$±s)	71.11±30.98	95.40±2.23	98.81±1.35	79.85±32.21	97.63±0.43	94.57±0.68	96.72±1.53	78.00±17.89
Wide and Deep ($\bar{x}$±s)	87.41±0.93	97.75±1.04	99.66±0.11	96.78±0.73	96.56±0.51	94.85±0.91	97.27±0.28	92.00±10.95
	I	**J**	**K**	**L**	**M**	**N**	**O**	**P**
XGBoost ($\bar{x}$±s)	99.04±0.10	97.75±0.30	96.26±0.40	51.43±0.50	93.96±1.26	93.40±0.38	40.89±1.99	94.52±1.28
CNN ($\bar{x}$±s)	94.87±0.87	97.56±0.71	91.47±6.46	70.16±9.85	93.53±2.54	95.15±0.60	93.33±14.91	93.28±1.95
GRU ($\bar{x}$±s)	99.07±0.20	99.31±0.01	86.45±24.23	69.89±5.23	93.45±4.24	98.59±0.90	78.67±13.66	96.49±1.50
Transformer ($\bar{x}$±s)	98.84±0.91	98.73±0.69	83.87±24.56	57.87±18.21	76.23±34.36	97.00±1.13	89.33±15.35	83.46±19.36
Attention GRU ($\bar{x}$±s)	98.70±0.17	98.13±0.50	96.60±1.13	74.02±2.63	79.40±33.60	96.34±0.51	75.33±18.50	96.36±3.27
Wide and Deep ($\bar{x}$±s)	98.61±0.24	98.64±0.13	98.04±0.23	76.12±0.97	96.96±1.40	97.33±0.53	74.67±7.30	98.18±1.30
	Q	**R**	**U**	**V**	**W**	**X**	**Y**	
XGBoost ($\bar{x}$±s)	79.48±2.63	96.11±0.08	35.05±4.50	96.81±0.32	96.14±0.10	90.11±0.57	88.71±0.51	
CNN ($\bar{x}$±s)	97.11±1.48	94.44±0.59	100.00±0.00	92.47±1.40	83.16±1.24	95.43±0.22	70.24±3.78	
GRU ($\bar{x}$±s)	86.62±17.90	96.43±0.38	100.00±0.00	82.81±24.27	82.04±0.49	81.80±29.11	85.90±5.23	
Transformer ($\bar{x}$±s)	83.06±16.88	93.97±2.22	95.56±6.09	91.61±1.30	79.20±0.23	96.49±1.54	87.86±9.20	
Attention GRU ($\bar{x}$±s)	92.27±3.36	94.89±1.01	100.00±0.00	71.32±29.14	82.06±0.68	97.87±0.80	84.48±0.91	
Wide and Deep ($\bar{x}$±s)	97.27±0.79	96.23±0.73	97.78±4.97	95.31±0.17	86.38±0.50	97.78±0.20	63.79±5.19	

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit

SUPPLEMENTARY TABLE S6. Recall of deep learning models for ICD-10 classification of underlying causes of death (%).

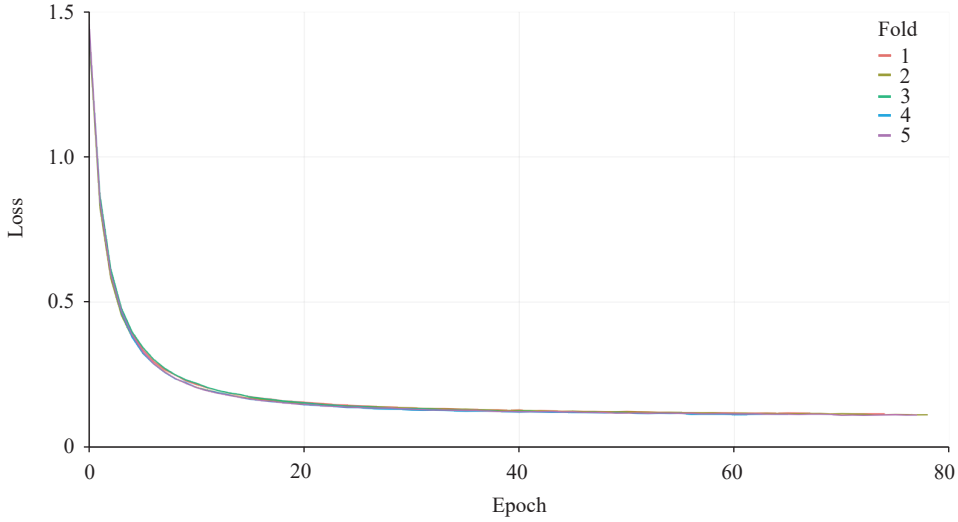
Models	A	B	C	D	E	F	G	H
XGBoost ($\bar{x}$±s)	73.11±1.15	79.57±1.39	65.01±0.22	59.45±1.28	68.60±0.30	87.67±0.59	78.70±1.88	88.00±10.95
CNN ($\bar{x}$±s)	84.67±1.50	82.61±1.15	96.54±0.41	86.61±1.34	86.45±0.84	91.04±0.93	94.00±0.21	56.00±8.94
GRU ($\bar{x}$±s)	84.00±0.61	82.90±1.96	92.69±0.69	82.52±1.32	85.77±0.45	91.10±0.89	93.38±0.30	60.00±0.00
Transformer ($\bar{x}$±s)	82.67±0.61	84.06±1.70	92.97±0.94	80.71±1.78	78.51±3.52	90.55±0.55	91.30±0.98	80.00±0.00
Attention GRU ($\bar{x}$±s)	88.22±1.54	89.28±0.61	92.27±1.45	87.72±1.19	86.06±0.65	92.94±0.65	93.66±0.45	80.00±0.00
Wide and Deep ($\bar{x}$±s)	90.89±1.01	88.12±0.65	97.22±0.14	94.41±0.51	89.17±0.39	93.74±0.47	96.29±0.40	76.00±8.94
	I	**J**	**K**	**L**	**M**	**N**	**O**	**P**
XGBoost ($\bar{x}$±s)	77.10±0.46	83.38±0.13	59.85±0.80	73.13±3.01	62.95±1.17	76.03±1.05	100.00±0.00	73.40±2.24
CNN ($\bar{x}$±s)	84.72±0.46	95.41±0.30	85.24±0.69	75.63±6.59	84.89±1.50	89.77±0.86	80.00±20.92	88.35±1.82
GRU ($\bar{x}$±s)	82.83±1.01	95.14±0.55	82.37±1.27	83.75±2.37	84.73±1.90	90.61±0.93	100.00±0.00	90.10±2.21
Transformer ($\bar{x}$±s)	73.52±2.71	92.39±1.02	77.17±1.85	65.63±6.81	82.78±1.47	88.97±0.85	100.00±0.00	82.33±2.51
Attention GRU ($\bar{x}$±s)	82.77±0.48	94.11±0.86	85.57±0.58	87.81±2.80	89.96±0.96	90.89±0.44	100.00±0.00	91.46±1.87
Wide and Deep ($\bar{x}$±s)	88.12±0.33	96.62±0.08	90.35±0.29	89.69±1.78	91.14±1.16	93.64±0.38	100.00±0.00	93.98±0.81
	Q	**R**	**U**	**V**	**W**	**X**	**Y**	
XGBoost ($\bar{x}$±s)	78.04±1.64	89.72±0.34	80.00±6.85	48.77±0.52	62.60±0.40	65.96±1.04	85.44±0.48	
CNN ($\bar{x}$±s)	91.57±1.12	93.93±0.80	100.00±0.00	92.99±0.81	91.77±1.54	91.64±0.95	38.60±3.71	
GRU ($\bar{x}$±s)	91.96±1.89	92.39±0.39	100.00±0.00	87.07±0.72	88.42±1.23	87.11±0.41	27.12±1.87	
Transformer ($\bar{x}$±s)	84.90±3.77	91.63±0.74	100.00±0.00	87.65±0.49	89.73±0.46	86.55±0.75	13.90±0.39	
Attention GRU ($\bar{x}$±s)	96.67±1.12	92.84±0.26	100.00±0.00	89.04±1.05	88.70±0.73	89.30±0.64	27.44±2.61	
Wide and Deep ($\bar{x}$±s)	97.45±0.88	95.62±0.36	100.00±0.00	95.60±0.24	92.68±0.76	95.12±0.21	46.45±2.13	

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit

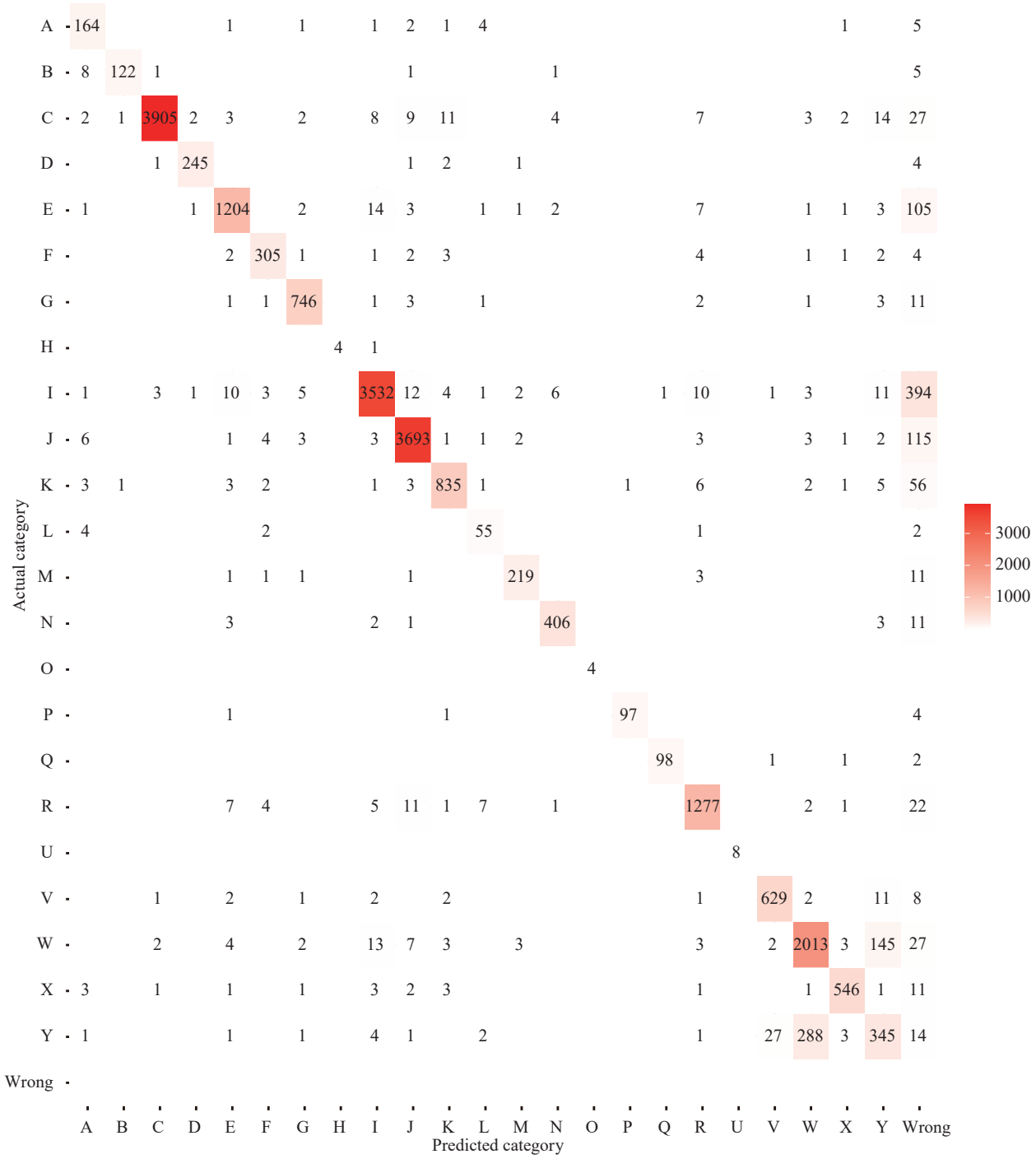
SUPPLEMENTARY TABLE S7. F1 Score of deep learning models for ICD-10 classification of underlying causes of death.

Models	A	B	C	D	E	F	G	H
XGBoost ($\bar{x}$±s)	74.23±1.45	87.28±0.60	78.54±0.16	71.46±1.17	80.14±0.22	89.12±0.40	86.90±1.21	84.73±10.73
CNN ($\bar{x}$±s)	85.04±0.91	87.77±1.51	97.72±0.21	89.12±2.75	90.33±0.63	90.94±1.26	94.71±0.42	64.76±11.06
GRU ($\bar{x}$±s)	84.78±1.81	86.83±2.52	96.07±0.36	88.35±2.05	91.30±0.44	93.07±0.18	79.42±21.70	73.33±3.73
Transformer ($\bar{x}$±s)	69.70±29.14	83.32±7.22	96.17±0.51	80.64±2.81	86.90±2.21	89.60±3.77	94.10±0.75	45.60±41.38
Attention GRU ($\bar{x}$±s)	74.38±26.59	92.22±1.07	95.42±1.16	79.66±24.46	91.48±0.32	93.75±0.49	95.16±0.90	78.09±10.02
Wide and Deep ($\bar{x}$±s)	89.11±0.42	92.68±0.72	98.42±0.12	95.58±0.38	92.72±0.33	94.29±0.63	96.78±0.31	82.56±6.13
	I	**J**	**K**	**L**	**M**	**N**	**O**	**P**
XGBoost ($\bar{x}$±s)	86.70±0.31	89.99±0.18	73.80±0.66	60.37±1.12	75.39±1.03	83.82±0.59	58.02±1.97	82.61±1.38
CNN ($\bar{x}$±s)	89.51±0.45	96.47±0.27	88.15±3.14	72.33±6.41	88.99±1.68	92.38±0.46	83.62±12.02	90.73±1.21
GRU ($\bar{x}$±s)	90.22±0.64	97.18±0.28	82.71±14.68	76.11±3.35	88.84±2.35	94.42±0.55	87.56±8.26	93.17±1.21
Transformer ($\bar{x}$±s)	84.29±1.78	95.45±0.83	78.64±14.52	59.77±9.95	74.72±27.61	92.81±0.84	93.78±9.08	82.00±12.06
Attention GRU ($\bar{x}$±s)	90.04±0.23	96.08±0.64	90.74±0.57	80.29±1.98	80.16±27.04	93.53±0.24	84.89±12.41	93.84±2.42
Wide and Deep ($\bar{x}$±s)	93.07±0.19	97.62±0.10	94.04±0.19	82.35±1.26	93.95±0.98	95.45±0.30	85.33±4.87	96.03±0.98
	Q	**R**	**U**	**V**	**W**	**X**	**Y**	
XGBoost ($\bar{x}$±s)	78.74±1.98	92.80±0.17	48.66±5.22	64.86±0.44	75.83±0.28	76.16±0.61	87.04±0.26	
CNN ($\bar{x}$±s)	94.25±0.75	94.18±0.62	100.00±0.00	92.72±1.05	87.24±0.89	93.49±0.58	49.65±2.58	
GRU ($\bar{x}$±s)	88.38±11.24	94.37±0.25	100.00±0.00	82.99±15.99	85.11±0.65	81.55±20.56	41.13±1.53	
Transformer ($\bar{x}$±s)	83.40±10.12	92.77±0.83	97.65±3.22	89.58±0.72	84.14±0.27	91.24±0.65	23.97±0.75	
Attention GRU ($\bar{x}$±s)	94.38±1.33	93.85±0.48	100.00±0.00	76.22±19.59	85.25±0.54	93.39±0.46	41.38±3.06	
Wide and Deep ($\bar{x}$±s)	97.36±0.56	95.92±0.40	98.82±2.63	95.45±0.11	89.42±0.17	96.43±0.05	53.62±1.32	

Abbreviation: CNN=Convolutional Neural Networks; GRU=Gated Recurrent Unit



SUPPLEMENTARY FIGURE S1. Loss curve across 5-fold cross validation.
Note: The original dataset contained 5,516 underlying cause of death codes, plus one unknown code and one missing code, making a total of 5,518 dimensions as the input and output layers for the ICD-10 codes. To validate the model's performance robustly, we employed a 5-fold cross-validation approach. After exploratory experiments, the embedding vector dimension of the WideAndDeep model was set to 24, and the hidden layer dimension to 32. The model used a cross-entropy loss function and an Adam optimizer with a learning rate of 0.0001 and L2 regularization of 0.00001. A learning rate scheduler with a factor of 0.5 was set. The dropout rate was 0.5, the batch size was 32. In the training phase of our model, we implemented an early stopping mechanism. The training process was terminated if the validation loss failed to decrease over 5 consecutive epochs. The best model stopped training early at 66 epochs.



SUPPLEMENTARY FIGURE S2. Confusion matrix of Wide and Deep model.
Note: The ICD-10 code comprises an alphabetic classification and a numeric code, such as “I21.1”. In cases where the model accurately predicts the alphabetic part but provides an incorrect three-digit code, it is classified as “Wrong”.