Original Research Article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 10: 1–17
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076241286635
journals.sagepub.com/home/dhj
Sage

Exploring the influential factors and improvement strategies for digital information literacy among the elderly: An analysis based on integrated learning algorithms

Haiyan Kong[1,2] and Xinyu Wang[3]

Abstract

Objective: Despite previous research identifying factors such as age, education level, income, and interest in technology that influence digital literacy among the elderly, this study attempts to use machine learning algorithms, especially ensemble learning algorithms, to predict and identify the key factors that affect the digital information literacy of the elderly, so as to propose effective strategies to improve the elderly's ability to utilize digital information and better integrate into the digital society.

Methods: This study used primary data on older adults from the Digital Divide Survey 2022 conducted by the Korea National Information Society Agency. A predictive model was built, and 15 variables that were highly important in predicting digital information literacy were identified. Prediction accuracy was assessed using an ensemble of algorithms including Random Forest, LGBM, XGBoost, AdaBoost, and CatBoost.

Results: The study found that in addition to demographic factors and personal technology use ability factors, relationship support factors and social digital environment factors are also important predictors of digital information literacy for the elderly. Among different predictive models, the CatBoost model, based on boosting ensemble, exhibited the highest predictive accuracy at 86.2%, followed by Random Forest (85.5%), LGBM (85.2%), XGBoost (84.5%), and AdaBoost (83.8%). The predictive accuracies of these models were higher than those of traditional machine learning models, indicating the effectiveness of ensemble learning algorithms in predicting digital information literacy among the elderly.

Conclusions: The academic significance of this study lies in the application of artificial intelligence technologies to the social sciences, specifically demonstrating the effectiveness of ensemble learning algorithms in predicting factors influencing the digital literacy levels of the elderly. This approach provides a novel and powerful tool for addressing complex social issues. The practical significance lies in the proposed strategies for improving the digital literacy of the elderly based on the research results, including education and training, social relationship support, social participation, technical support, and policy formulation, aiming to help the elderly better adapt to the digital environment, narrow the digital divide, and enhance the elderly's sense of participation and happiness in the digital society.

Keywords

Elderly, digital information literacy, influencing factors, improvement strategies, machine learning, ensemble learning

Submission date: 11 April 2024; Acceptance date: 30 August 2024

[1]School of Business, Xinyang Normal University, Xinyang, China
[2]Dabie Mountain Economic and Social Development Research Center, Xinyang, China
[3]Department of Management Information Systems, Chungbuk National University, Cheongju, Korea

Corresponding author:
Xinyu Wang, Department of Management Information Systems, Chungbuk National University, Cheongju 28644, Republic of Korea.
Email: startxinyu@hotmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

Introduction

With the rapid development and popularization of information and communication technology (ICT) around the world, every aspect of human life has undergone a significant transformation. This technology has not only increased the convenience of life and expanded the opportunities for communication but has also largely changed the way information is accessed, processed, and used.[1] More importantly, ICT offers the elderly in accessing health information, engaging in social activities, and maintaining social connections, and this has become a key factor in improving the health and life quality for the older population.[1–3] Despite the opportunities that ICT offers, the digital divide becomes more significant between different generations in which the older population was not able to acquire, validate, and effectively use information through ICT easily. This issue has become even more pronounced in the quarantine period during the COVID-19 pandemic.

According to the digital information literacy levels of various age groups in "The Report on the Digital Divide of 2020" conducted by the National Information Society Agency (NIA) of South Korea, there is a significant gap in digital information literacy among older people compared with the general public, with digital information literacy scores in the 60 s (85.6) and 70 s (29.7) age groups well below the level of the general public (100).[4] This gap is more significant in countries with an aging population and an information society, such as South Korea. The variation in information technology capabilities among the elderly exacerbates social differentiation and exclusion, further leading to social inequality.[5] Particularly, during the quarantine period of the COVID-19 pandemic, elderly individuals with little knowledge of digital information had faced more difficulties in daily life and were benefited less from online services.

Previous studies on digital information literacy among the older population have mainly utilized traditional econometric models to analyze the causal relationship between dependent and independent variables. It has been found that age, income, education, social support, and the need to be close with family have been identified as key factors influencing digital information literacy among older adults.[6–10] In recent years, the development of artificial intelligence and machine learning techniques has provided new tools for social science research. These techniques are capable of assessing linear and nonlinear relationships between variables and accurately predicting influencing factors.[11] The applications of machine learning are not limited to areas such as medical diagnostics,[12] fraud detection,[13] and student satisfaction prediction,[14] but it also shows great potential in predicting the influencing factors of the digital divide and aging.[11,15,16]

A number of studies have reported a number of factors that influence digital information literacy among older adults, such as age, education level, income, and interest in technology.[6–10] In this study, we aim to develop a deeper understanding of the complex relationships between these aforementioned factors and their interactions with other variables that may not have been fully considered before, through machine learning techniques. This study aims to establish a model based on public data and machine learning techniques to predict and analyze the key factors affecting the digital literacy of the elderly. This method identifies the influencing factors of digital information literacy of the elderly, thereby providing a scientific basis for designing effective intervention measures. These interventions should comprehensively consider various aspects such as education and training, social relationship support, social participation, technical support, and policy formulation. This comprehensive approach will enable the elderly population to better adapt to the digital environment, enhance their digital literacy, bridge the digital divide, and improve their sense of participation and well-being in the digital society. The results of this study will contribute theoretically to the exploration of the application of machine learning and artificial intelligence technologies in improving the health and well-being of the elderly. Additionally, the findings provide empirical evidence for enhancing the digital literacy of the elderly and offer valuable references for policymakers, social service providers, and technology developers. This ensures that the elderly can fully utilize information technology and enjoy the various benefits brought by the digital age.

Literature review

Digital information literacy

Digital information literacy is a multidimensional concept that aimed at describing an individual's ability to access, understand, evaluate, and apply information in the digital age.[1,2] This concept encompasses terms related to information literacy, digital literacy, and ICT skills. In recent years, this concept has had an unprecedented impact on the quality of life of older adults as society enters the age of aging and digital information. Studies have shown that good digital information literacy is associated with positive health outcome in older adults and can reduce depression levels, alleviate loneliness, and enhance life satisfaction through newly built social relationships online.[5,17,18]

However, there is a significant digital divide among the elderly groups over 65 years old in South Korea, with their skills in using digital information being significantly lower than those of other age groups.[11] This phenomenon not only reflects the lack of training in digital skills in formal education but also highlights the deficiencies of older adults in using computers or smartphones for information retrieval, online shopping, online banking, administrative tasks, and social activities.[19] Despite the existence of arguments that the IT technology is too complex for older adults, the popularity of the Internet actually provides them opportunities to interact with society, enhance independence, and communicate across generations.[5,18]

Given the disadvantages faced by the elderly in digital information literacy, research in this field holds significant social importance. By identifying and understanding the key variables that influence the digital information literacy of the elderly, effective strategies can be proposed to enhance their digital information utilization capabilities. This not only helps bridge the digital divide but also significantly improves the quality of life for the elderly, thereby promoting overall societal progress and development.

Application of machine learning in digital information literacy research

Machine learning, a pivotal branch of artificial intelligence, is dedicated to mimicking the human learning process through a combination of mathematics, statistics, cognitive science, and computer science. This technique recognizes patterns by analyzing historical data and thus predicts future trends.[16,20] Its applications in various fields such as disease diagnosis,[12] fraud detection,[13] text classification,[21] and image recognition[22] have significantly impacted the research and industry. However, traditional machine learning algorithms face challenges such as data imbalance, which limits their effectiveness in certain contexts.[23] To address these challenges, researchers have turned to new approaches such as ensemble learning and deep learning, aiming to improve performance and generalization by combining the strengths of multiple models.

Ensemble learning, which enhances the overall performance by combining the prediction results of multiple base learners, has achieved excellent results in various application scenarios.[24] Although these methods have been widely used in several research fields, however, relatively limited research has been conducted in the area of digital information literacy. Past studies have primarily focused on the use of traditional machine learning algorithms to address specific problems, such as Internet usage and successful aging judgments.[15,16,25] Given the successful application of ensemble learning techniques in other domains and their capability to handle complex datasets, this study aims to bridge this gap, particularly regarding the factors influencing digital information literacy among the older population. We will focus on several widely used ensemble algorithms, including Random Forest, AdaBoost, XGBoost, LightGBM, and CatBoost, and explore how they can help improve digital information literacy among older adults.

The primary objective of this study is to utilize machine learning methods, particularly ensemble learning algorithms, to predict and analyze the key factors influencing the digital information literacy of elderly individuals. By doing so, the study aims to propose effective strategies to enhance the digital information utilization capabilities of the elderly. Unlike previous studies based on causal relationship models, this research leverages feature importance evaluation in the XGBoost algorithm to achieve more accurate predictions and analyses of the factors affecting elderly individuals' digital information literacy. Consequently, the study seeks to develop effective strategies to facilitate the elderly's better integration into the digital society, improve their quality of life, and address the challenges posed by an aging population.

Methods

This study aims to provide valuable insights to enhance digital literacy among the elderly by predicting the factors influencing their digital literacy and assisting them in adapting to the digital development environment. The core of the research lies in utilizing advanced machine learning algorithms to predict the digital literacy of the elderly based on the 2022 Digital Divide Survey data conducted by the NIA of South Korea. The specific research methodology steps are as follows: firstly, selecting the raw data concerning the elderly from the 2022 NIA Digital Divide Survey data; then, selecting 15 variables deemed highly important in predicting digital literacy based on feature importance; and visualizing the specific contributions of these 15 features in predicting the digital literacy of the elderly using the SHAP method; constructing prediction models using ensemble machine learning algorithms, including Random Forest, LGBM, XGBoost, AdaBoost, and CatBoost; and evaluating the performance and prediction accuracy of each model; and finally, based on the analysis of results, determining the variables that are highly important and influential in predicting digital literacy, as well as identifying which machine learning algorithm is more effective in predicting the digital literacy of the elderly. The research framework is illustrated in Figure 1.

Participants and data preprocessing

The data used in this study are a cross-sectional dataset from the 2022 Digital Divide Survey conducted between September and December 2022. The South Korean government has been conducting this survey since 2002 to measure digital usage and the digital divide among different population groups. This dataset includes general demographic characteristics, digital device usage proficiency (PC/mobile digital devices), digital literacy levels, attitudes toward and actual usage of digital information technology, as well as changes in usage and perceptions of digital transformation due to COVID-19, from 15,000 individuals residing in 16 cities across South Korea. The participants include general citizens, people with disabilities, the elderly, and low-income individuals, among others. In the interview, each participant answered about 160 to 180

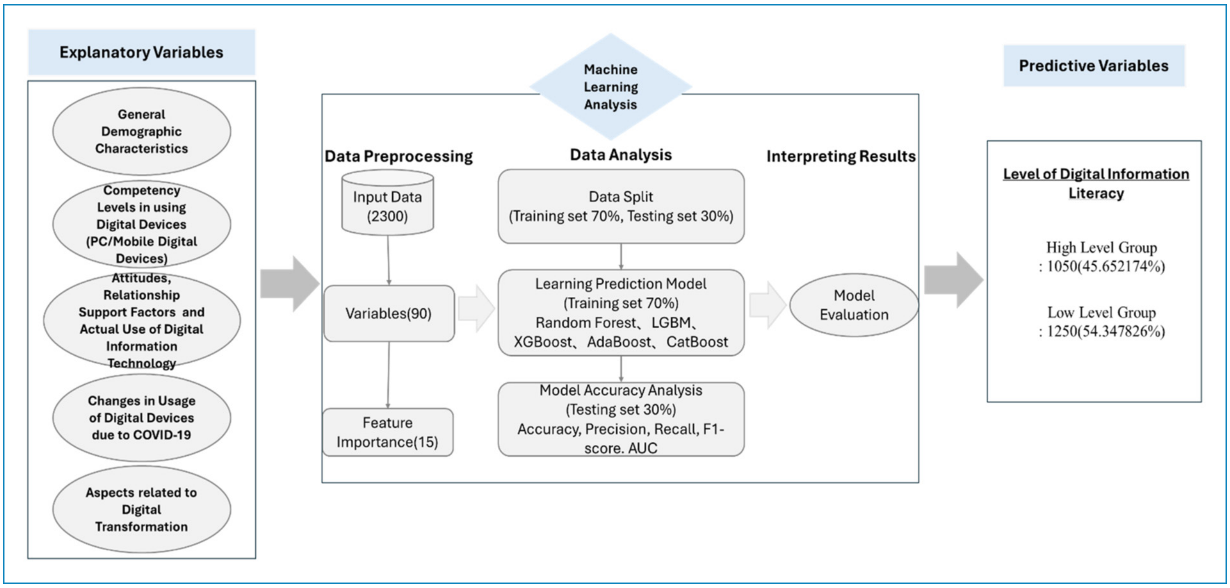

Figure 1. Framework.

questions. The survey questions were formulated based on the NIA (2022) questionnaire and referenced previous studies[5,11,19,26–28] to organize the variable names and classifications used in this study (see Table 1). The survey employed a square root proportional allocation method, based on sociodemographic characteristics collected through multistage stratified sampling (e.g., gender, age, and residence). Consequently, these data are appropriate for this study, which investigates a sample of 2300 individuals aged 55 and above. Table 2 reports the basic characteristics of the sample population.

This study utilized the open-source programming language Python (version 3.8.10) and Jupyter Notebook, along with machine learning libraries such as Scikit-learn (version 1.0.2), Pandas (version 1.4.1), Numpy (version 1.22.2), and Matplotlib (version 3.5.2) for analysis. The research process was divided into five steps: initial data selection, handling missing data and data structure processing, feature ranking to identify the most important features, machine learning algorithms to develop classifier models, and model evaluation.

To maximize the benefits from this approach, it is essential to minimize the number of discarded variables during the preprocessing stage. Fundamentally, data preprocessing includes data integration, handling of missing or noisy values, and technical processing, which enhances the authenticity and predictability of the data. Initially, we compared the questions included in the annual surveys to extract common questions and replaced them with identical variable names for the elderly sample. Due to the nature of the survey, variables with missing data due to lack of response were removed. Additionally, irrelevant or insignificant adjustment parameters used for formulating the dependent variable were also discarded. Subsequently, because the characteristics of the data structure were not uniform in dimensions and units, which could affect the model's evaluation of feature importance, we first converted categorical variables to numerical ones using label encoding (e.g., education level, income) and then applied min–max scaling. Finally, we created a discrete predictive variable based on the average value of the digital literacy index for the elderly, converting it into a binary classification problem. For the fifth step, we allocated the 2300 samples into 70% for the training dataset and 30% for the testing dataset.

Predictive variables

In this study, we utilized a binary dependent variable to predict the digital literacy levels of the elderly. This consisted of 19 questions encompassing both basic and advanced levels of digital literacy for the elderly. These questions were referenced from the NIA (2022) questionnaire and previous research analyses.[11,19] All questions were rated on a five-point Likert scale (strongly disagree, disagree, neutral, agree, and strongly agree). The specific content is detailed in Table 3. To improve prediction accuracy, we create a discrete predictive variable based on the average value of the digital information literacy indicators for the elderly, assigning values 0 and 1 to represent low and high levels of digital information literacy, respectively. Among the 2300 participants tested, 1250 (54.347826%) exhibited a lower level of digital literacy, whereas 1050 (45.652174%) demonstrated a higher level of digital literacy. Discretization is a key step in machine learning research as it simplifies data representation and enhances understanding.[11]

Table 1. Survey content on the digital information divide.

Classification	Measurement items	Metrics
Demographic factors	Gender, age, occupation, education, income, disability status, household type, housing type, etc.	Dummy, categorical
Digital device usage ability	Access level (ownership of wired and wireless information devices at home and the possibility of accessing the Internet at any time)	Categorical
	Skill level (PC usage ability and mobile digital device usage ability)	4-point scale (strongly disagree → strongly agree)
Digital information level	Information processing ability, information recognition and management, information production and sharing, economic activities, social participation, security settings and privacy protection, online rights protection, etc.	5-point scale (strongly disagree → strongly agree)
Digital information usage attitude and outcomes	Motivation for using digital devices, social relationship support, confidence and usage attitude, effects of using digital devices, satisfaction in various aspects of daily life, and overall life satisfaction	4-point scale (strongly disagree → strongly agree)
	Frequency of PC/mobile digital device usage, reasons for not using the Internet, willingness to use the Internet in the future	Categorical
Digital transformation and related to COVID-19	Changes in Internet/mobile service usage due to COVID-19, perception	5-point scale (strongly disagree → strongly agree)
	Experiences with Internet/mobile services related to COVID-19, reasons for not experiencing them	Categorical
	Perception, demand, satisfaction with digital transformation	5-point scale (strongly disagree → strongly agree)
	Perception, demand, participation in digital education courses	Categorical

Explanatory variables

To improve prediction accuracy, we utilized the feature importance algorithm of XGBoost to identify the most critical explanatory variables in the predictive model. The feature importance algorithm is capable of eliminating redundant and irrelevant features, thereby improving the performance of the classifier.[29] As illustrated in Figure 1, we initially removed variables related to personal identity information and missing data from the original dataset of 2,300, resulting in 90 explanatory variables. Subsequently, we employed the XGBoost algorithm to calculate the importance scores of these 90 explanatory variables. This enabled us to identify the most crucial factors affecting the prediction outcomes, with all feature importance scores depicted in Figure 2.

The data structure features impacting the digital information literacy levels of the elderly are not uniform in dimensions and units, which affects the model's evaluation of feature weights, thereby influencing its accuracy and convergence speed. Therefore, min–max normalization is applied for feature normalization, scaling the feature data to the [0,1] range. The transformation function is as follows:

$$x^{*} = \frac{\mathrm{x} - x_{min}}{x_{max} - x_{min}}$$

where max is the maximum value in the sample data and min is the minimum value in the sample data.

From the 90 explanatory variables, we derived the top 15 explanatory variables based on the importance scores, forming the optimal feature subset. The results of the feature selection are presented in Table 4.

Interpretability of SHAP method

To further understand the contribution of features to the model output, we calculated SHAP values. SHAP values provide specific contributions of each feature to the model's predictions, helping us identify which features have positive or negative impacts on the prediction outcomes for specific samples. By visualizing SHAP values

Table 2. Demographic characteristics.

Variables	Category	Frequency	%
Gender	Male	1118	48.61
	Female	1182	51.39
Age	55–64	1130	49.13
	65–74	872	73.91
	75+	298	12.96
Education level	Elementary School or less	312	13.57
	Middle school	483	21.00
	High school	1185	51.52
	College or higher	320	13.91
Monthly income (KRW)	0–1,990,000	596	25.91
	2,000,000–3,990,000	874	38.00
	4,000,000–5,990,000	523	22.74
	6,000,000–7,990,000	234	10.17
	Over 8,000,000	73	3.17
Occupation	Manager	45	1.96
	Professional and related workers	17	0.74
	Clerical workers	81	3.52
	Service workers	219	9.52
	Sales workers	357	15.52
	Skilled agricultural, forestry, and fishery workers	91	3.96
	Craft and related trade workers	184	8.00
	Plant and machine operators and assemblers	52	2.26
	Elementary occupations workers	191	8.30
	Military personnel	1	0.04
	Full-time housewife	694	30.17
	Elementary/middle/high school student	343	14.91
	University student (including junior college and graduate students)	25	1.09
	Unemployed	45	1.96
	Others	17	0.74
Residence	Urban	2037	88.57
	Rural	263	11.44
Total		2300	100

(as shown in Figure 3), we can intuitively observe the specific contributions of the top 15 features in predicting the digital literacy of the elderly. The dots represent the samples for each feature, colored according to their feature values, ranging from blue to red, indicating the ascending order of values. Features such as Q4A5 (ability to use a PC), Q15A2 (how to solve problems when using digital devices), ADQ4 (educational level), and ADQ1 (age) contribute significantly to the model's predictions. Q4A5 (ability to use a PC), Q15A2 (how to solve problems when using digital devices), and ADQ4 (educational level) have a positive impact on the model's predictions, whereas ADQ1 (age) has a negative impact.

Feature correlation test

This study selected the ETA coefficient to calculate the correlation between features and constructed a correlation matrix to test the degree of correlation among the selected features. The ETA coefficient (η^2) is a commonly used measure of feature correlation, particularly suitable for the relationship between categorical and numerical variables. Given a pair of variables (X, Y), the ETA coefficient is defined as follows:

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

where SS_{total} is the total sum of squares, representing the total variance of the numerical variable with respect to its mean. $SS_{between}$ is the sum of squares between groups, representing the variance between the means of each group (defined by the categorical variable) and the overall mean.

Table 3. Dependent variables.

Measurement items	
Information Literacy Indicators_1	I can install, uninstall, or upgrade programs on a PC, or copy, delete, move, or change files and folders.
Information Literacy Indicators_2	I can install, delete, or update apps on smart devices.
Information Literacy Indicators_3	I can use tool apps such as a calculator, schedule manager, address book, etc., on smart devices.
Information Literacy Indicators_4	I can host or participate in meetings using remote, non-face-to-face meeting apps (such as Google Meet, Zoom, etc.).
Information Literacy Indicators_5	I can utilize smart devices that sync with smartphones, such as smartwatches (Galaxy Watch, Apple Watch), smart refrigerators, and Internet of Things (IoT) devices.
Information Literacy Indicators_6	I can distinguish reliable information from search results by comparing various data.
Information Literacy Indicators_7	I can discern fake news by seeking and utilizing relevant reference materials or information.
Information Literacy Indicators_8	I know how to use smartphone settings changes (filtering) needed to filter out harmful information (pornography, criminal or violent content, etc.).
Information Literacy Indicators_9	I can convert existing video content such as dramas, animations, music videos, movies, etc., into other formats like videos or GIFs.
Information Literacy Indicators_10	I can work on assignments or projects with others using online collaboration programs (such as Google Docs).
Information Literacy Indicators_11	I can make purchases using online simple payment systems (such as Naver Pay, Kakao Pay, etc.).
Information Literacy Indicators_12	I can find directions using navigation, online map services (Kakao Map, Naver Map, Google Map, etc.), and traffic information.
Information Literacy Indicators_13	I can find and participate in communities (online cafes, etc.) on the Internet that match my interests.
Information Literacy Indicators_14	I can discuss political/social issues happening online or engage in signing petitions.
Information Literacy Indicators_15	I can set up security settings such as locks on PCs, smartphones, etc.
Information Literacy Indicators_16	I can delete cookies and browsing history on PCs, smartphones, tablet PCs.
Information Literacy Indicators_17	When posting on social media or forums, I can set the privacy scope of my posts.
Information Literacy Indicators_18	I know how to take provisional measures against online insults or defamation against me.
Information Literacy Indicators_19	I know how to report if someone infringes my rights (such as defamation of character, copyright infringement, etc.) on portals or social media.

(continued)

Table 3. Continued.

The value of the ETA coefficient ranges between [0, 1], with higher values indicating stronger correlations between variables.

Assuming mmm is the sample size in dataset D, and each sample dataset contains n features (with the nth feature being the prediction target), we calculate the ETA coefficient between each pair of features to form a correlation matrix. η_{ij}^2 represents the ETA coefficient between feature iii and feature j, defined as follows:

$$\eta_{ij}^2 = \frac{SS_{between}}{SS_{total}}$$

The results of the ETA coefficient calculations showed that the correlations among the selected 15 feature vectors were relatively weak, with the correlation coefficients ranging from a minimum of 0.02 to a maximum of 0.55. Thus, the selected features are not redundant. The computed feature correlation heatmap is shown in Figure 4.

Model training

In order to determine the most effective binary classifier for predicting the digital information literacy level of the

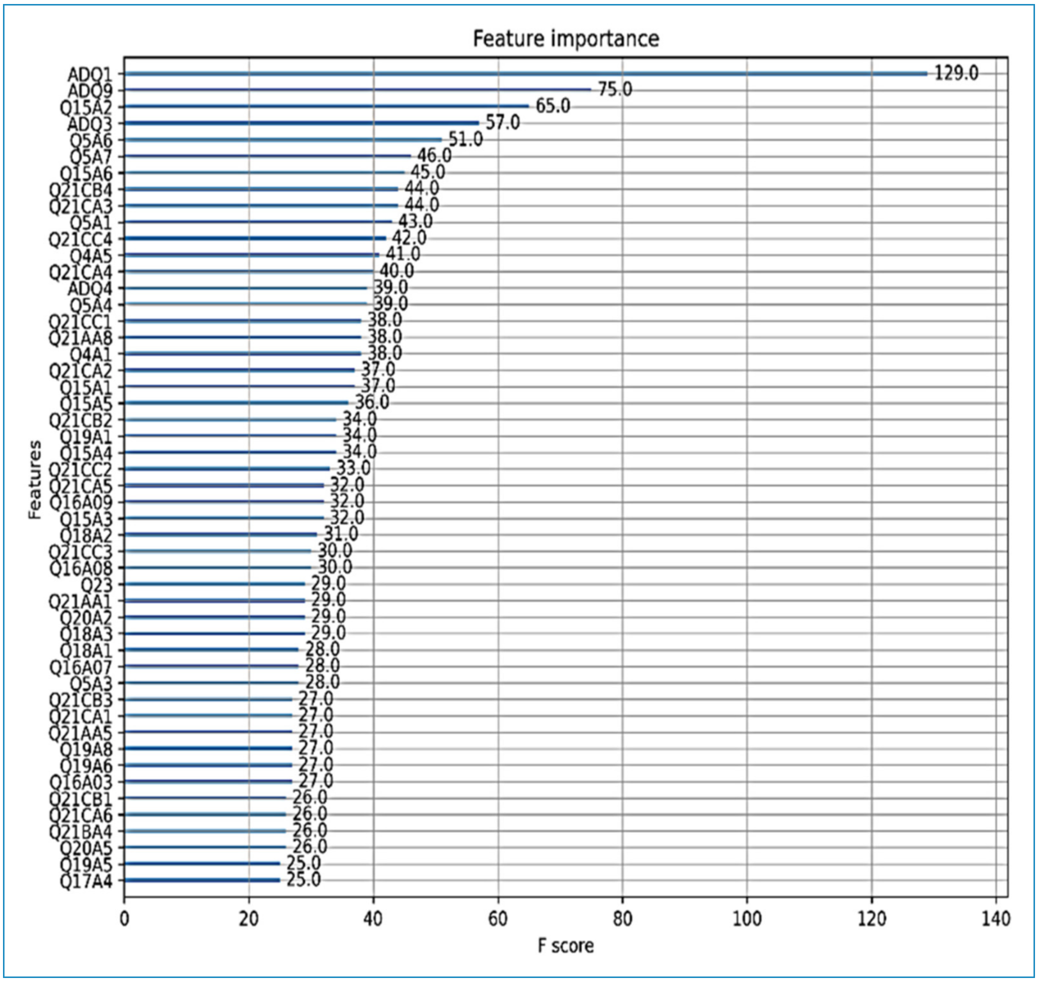

Figure 2. Feature importance.

elderly, we adopted classic and popular ensemble learning algorithms, namely, the bagging representative algorithm Random Forest and the boosting representative algorithm AdaBoost, XGBoost, LightGBM, and CatBoost.

Random Forest: This algorithm is suitable for high-dimensional datasets with multiple variables, utilizing a voting process for high-accuracy sample classification. The algorithm's performance is shared among its most capable subtrees in classification. By distributing samples across subtrees with different dataset attributes, this algorithm achieves high accuracy on noisy data and can prevent overfitting. Moreover, Random Forest typically performs excellently when the input data contains many features (i.e., high-dimensional data).[30] For a given input sample, the output of the Random Forest classifier can be represented as follows:

$$H(x) = model\{h_1(x)h_2(x),\ \ldots,\ h_n(x)\}$$

where $h_i(x)$ is the prediction result of the i tree and $H(x)$ is the prediction result of the entire Random Forest model.

AdaBoost: Adaptive Boosting is an algorithm within the ensemble category that uses weak classifiers to classify cases simultaneously, successively identifying and eliminating errors in classified cases within each classifier. The advantages of this algorithm include its high accuracy and generality, efficient computational capabilities, flexibility for various tasks on complex data, ease of adaptation, and the ability to integrate with other algorithms.[31,32] For a binary classification problem, the final model of AdaBoost can be represented as follows:

$$H(x) = sign\left(\sum_{t-1}^{T} a_t h_t(x)\right)$$

where $h_t(x)$ is the prediction of the weak classifier in the t round for the sample x, a_t is the weight of that weak classifier, and T is the total number of weak classifiers.

XGBoost: XGBoost is a model developed through an improved boosting method for decision trees, featuring internal functions for regularizing overfitting and

Table 4. Independent variables.

No.	Variable names	Independent variables
1	ADQ1	Age
2	ADQ9	Average monthly income
3	Q15A2	How to solve problems when using digital devices_2) Searching for information online for help.
4	ADQ3	Occupation
5	Q5A6	Ability to use mobile devices_6) I can scan/repair malicious codes (viruses, spyware, etc.) on smart devices.
6	Q5A7	Ability to use mobile devices_7) I can create documents (notepad, word) on smart devices
7	Q15A6	How to solve problems when using digital devices_6) Asking a professional personnel such as a service center for help.
8	Q21CB4	Desire for digital transformation_ Willing to take classes when receiving training related to digital transformation.
9	Q21CA3	Awareness of digital transformation _I know about digital transformation.
10	Q5A1	Ability to use mobile devices_1) I can configure display/sound/security/notification/input method settings on smart devices.
11	Q21CC4	Satisfaction with digital transformation change _Efficiency increased through non-face-to-face study and work.
12	Q4A5	Ability to use a PC_5) I can transfer files on my PC to others via the Internet.
13	Q21CA4	Perception of digital transformation_ The digital transformation will not have much impact on my life.
14	ADQ4	Highest educational level
15	Q5A4	Ability to use mobile devices_4) I can send files/photos from a smart device to others.

performing internal cross-validation for each trial.[32,33] Due to its exceptional classification performance, XGBoost is often used in competitions such as Kaggle. Most importantly, the greatest advantage of XGBoost lies in its high practicality. XGBoost allows for the derivation of important metrics, indicating the variables of greater importance among the independent variables, thereby enabling the examination of each independent variable's relative predictive power. Predictions are created by weak learners that iteratively improve upon the previous learner. The objective function of XGBoost consists of two parts: one is the loss function L of the training data, and the other is the regularization term Ω used to control the complexity of the model. For a given prediction $\hat{y}_i^{(t)}$ and target y_i, the objective function at the t round can be represented as follows:

$$Obj^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^{t} \Omega(f_k)$$

where l is the loss function, f_k is the k tree, and $\Omega(f_k)$ is the regularization term. Among all boosting algorithms, the XGBoost model is possibly the most frequently used, with applications in research predictions such as medical diagnosis[34] and fraud detection.[35]

LightGBM: LightGBM is an efficient implementation of the gradient boosting algorithm, developed by Microsoft researchers in 2017.[36] This histogram-based algorithm increases training speed and reduces memory consumption through a leaf-wise growth strategy and maximum depth limitation. According to the level-wise growth strategy, leaves on the same level are split simultaneously. Leaves on the same level are treated equally, even though they have different information gains. Information gain refers to the expected reduction in entropy caused by splitting nodes based on an attribute.[37]

$$IG(B, V) = En(B) - \sum_{v \in (V)} \frac{|B_V|}{B} En(B_v)$$

$$En(B) = \sum_{d=1}^{D} - p_d log_2 p_d,$$

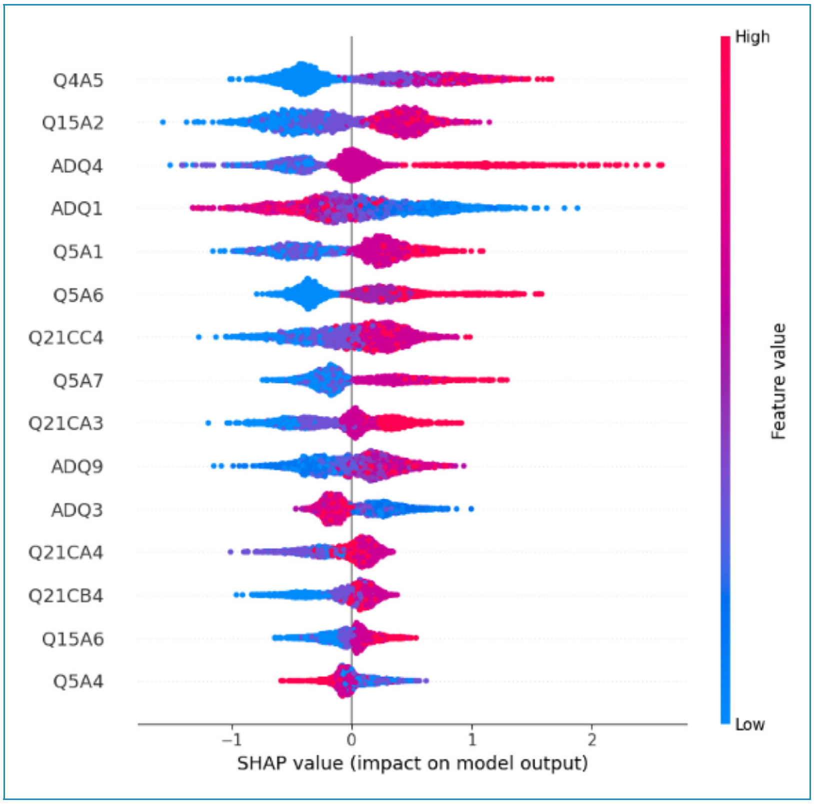

Figure 3. SHAP value.

where $En(B)$ is the information entropy of set B, p_d is the proportion of B belonging to category d, D is the number of categories, v is the value of attribute V, and B_v is the subset of B with attribute value v.

CatBoost: The CatBoost algorithm is an implementation of gradient boosting proposed by Prokhorenkova et al. in 2017.[38] CatBoost employs an improved and more robust approach that does not lead to overfitting and ensures that all examples in the training set are used for training the model. The algorithm effectively handles categorical features during the training phase. A significant improvement of CatBoost is its ability to perform unbiased gradient estimation, thereby reducing overfitting. This method involves randomly permuting the training set, and, for each sample, the algorithm calculates the average label value of samples with the same category value before the given sample in the permutation.[39] If $\sigma = (\sigma_1, \ \ldots, \sigma_n)$ is a permutation, then $x_{\sigma_{p,k}}$ is replaced with as follows:

$$\frac{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}]Y_{a_j} + a.p}{\sum_{j=1}^{p-1}[x_{\sigma_{j,k}} = x_{\sigma_{p,k}}]Y_{a_j} + a}$$

where P is the prior value and a is the weight of the prior value. Meanwhile, the parameter $a > 0$. Furthermore, adding prior values and prior weights in the CatBoost algorithm ensures the reduction of noise obtained from low-frequency categories.[40] Prokhorenkova et al.[38] compared its performance with XGBoost and LightGBM in their pioneering CatBoost article, indicating that CatBoost is less likely to overfit compared with XGBoost or LightGBM. They attribute the enhancement of performance to the aforementioned method used by the CatBoost algorithm for encoding categorical features. When the input is categorical data, the CatBoost algorithm achieves excellent performance, outperforming most machine learning algorithms, and it inherently can handle missing data.

Evaluation metrics and analysis results

The evaluation criteria of a model are crucial for the measurement of the final results. To determine the best model for assessing the digital information literacy of the elderly, we employ various metrics calculated based on the confusion matrix, such as accuracy, precision, recall, and F1 score, for measurement. A higher value indicates better model performance. The comparison results of different algorithm performance metrics are presented in Table 5. Based on the confusion matrix, several evaluation metrics can be calculated to assess model performance:

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

It represents the proportion of correctly predicted instances.

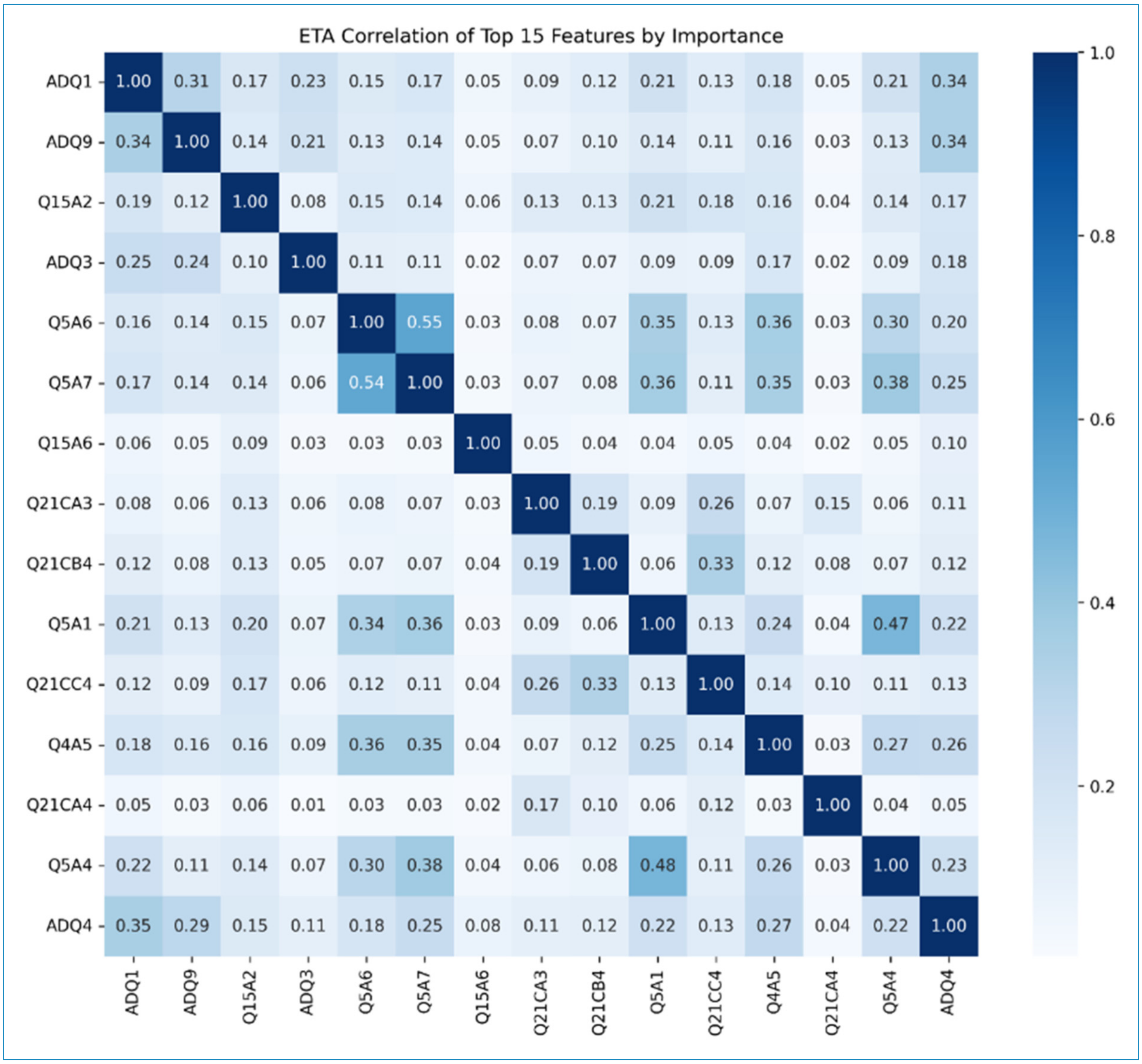

Figure 4. Feature correlation heat map.

Precision:

$$Precision = \frac{TP}{TP + FP}$$

It represents the proportion of instances predicted as positive that are actually positive.

Recall:

$$Recall = \frac{TP}{TP + FN}$$

It represents the proportion of actual positives that were correctly identified by the model.

F1 Score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The harmonic mean of precision and recall used to measure the model's overall effectiveness. In the equation, "TP" stands for true positive, "TN" for true negative, "FN" for false negative, and "FP" for false positive.

Based on the performance on the test set, the CatBoost model achieved the highest accuracy and AUC values, indicating its superior generalization ability and demonstrating its advantages in practical applications. In contrast, while the XGBoost and LGBM models performed excellently on the training set, they exhibited overfitting issues on the test set. The other models also performed comparably on the test set, but CatBoost had a slight edge. Considering the confusion matrix and other performance metrics (such as accuracy, precision, recall, F1 score, and AUC), the CatBoost model demonstrated better balance in classifying

Table 5. The results of the performance criteria of selected models.

		RandomForest	AdaBoost	XGBoost	LGBM	CatBoost
Best Params		max_depth: 10, min_samples_leaf: 4, min_samples_split: 10, n_estimators: 50	learning_rate: 0.1, n_estimators: 100	learning_rate: 0.1, max_depth: 5, n_estimators: 100	lambda_l1: 0.1, lambda_l2: 0.1, learning_rate: 0.1, min_data_in_leaf: 50, n_estimators: 100, num_leaves: 20	depth: 5, iterations: 100, learning_rate: 0.1
Accuracy	Train	0.935	0.858	0.991	0.997	0.953
	Test	0.857	0.859	0.855	0.857	0.861
Precision	Train	0.942	0.846	0.990	0.997	0.958
	Test	0.854	0.834	0.836	0.848	0.843
Recall	Train	0.917	0.845	0.989	0.997	0.941
	Test	0.830	0.868	0.836	0.841	0.858
F1 Score	Train	0.929	0.845	0.989	0.998	0.949
	Test	0.842	0.851	0.844	0.844	0.850
AUC	Train	0.988	0.940	0.998	0.998	0.993
	Test	0.942	0.939	0.937	0.937	0.943
Confusion matrix		[327, 45], [54, 264]	[317, 55], [42, 276]	[319, 53], [47, 271]	[324, 48], [51, 267]	[321, 51], [45, 273]

positive and negative samples and performed excellently on the test set, further proving the numerous advantages of the CatBoost classifier in predicting panel data. Therefore, it is recommended as the best model for practical applications. It is simple and efficient, can handle missing and noisy data, and shows good interpretability. The optimal parameters for each classifier and their performance on the training and test sets are shown in Table 5.

Differences in digital information literacy between urban and rural populations

To describe the basic situation of digital information literacy among urban and rural populations, we conducted statistical analyses on both groups separately and further examined whether the differences in digital information literacy between urban and rural populations are statistically significant through an independent samples t-test. The results are shown in Table 6.

Based on the comprehensive analysis, the mean digital literacy score for the urban population (0.477) is higher than that of the rural population (0.338). The t-test results indicate a t-statistic of 4.263 and a p-value much less than 0.05, demonstrating that the difference in digital literacy between urban and rural populations is statistically significant. This implies that the digital information literacy of the urban population is significantly higher than that of the rural population.

Discussion

Based on the results of the feature importance analysis, this study identified four dimensions that can predict the digital information literacy levels of the elderly: demographic factors, relational support factors, personal technology usage ability factors, and social digital environment factors. These dimensions are of great value in predicting the factors influencing digital information literacy among the elderly. This study was conducted to predict the digital information literacy of older adults in the current context, as many elderlies are keeping pace with technological developments amidst the acceleration of digital transformation. In particular, this study utilizes public data and machine learning, which have not been commonly used in previous digital information literacy research, to establish a predictive model of digital information literacy and perform related analyses. In this study, 15 variables that are significant in predicting digital information literacy among older adults were derived, which can be categorized into (1) demographic perspectives (ADQ1, ADQ9, ADQ3, ADQ4), (2) relational support perspective (Q15A2, Q15A6), (3) personal technology usage ability perspective (Q5A6, Q5A7, Q5A1, Q4A5, Q5A4), and (4) social digital environment perspective (Q21CB4, Q21CA3, Q21CC4, Q21CA4). The importance of these four perspectives for predicting the digital information literacy of the elderly will be discussed in the following text.

Table 6. *t*-Test results.

Residence type	Mean	SD (standard deviation)	*T*-value	*P*-value
Urban	0.477	0.500	4.263	0.000
Rural	0.338	0.474		

Firstly, the discussion of digital information literacy has largely focused on demographic factors such as age, income, occupation, and education. Research has shown that with increasing age, individual perceptual functions and cognitive abilities may decline, thereby affecting the ability of the elderly to use ICT.[41,42] In contrast, older adults with higher-income levels usually have better living condition and better quality of life tend to use the internet for communication more frequently, thereby accessing more information technology resources.[43,44] Moreover, the professional environment's demands for technology directly influence the digital information literacy levels of individuals in various occupational categories, with those working in information technology-related jobs usually having a higher level of technical proficiency and information processing capability compared with people in other professions.[15] The education level is directly linked to an individual's ability to acquire and utilize Internet-based ICT, with higher-educated individuals often possessing higher digital information literacy.[45] Therefore, age, income, occupation, and education are considered to be the main influencing factors in predicting the digital information literacy of the elderly, and this argument may have been reflected in the results of this study.

Secondly, relationship support is also an important influencing factor in predicting the digital information literacy of the elderly. When the elderly encounter issues using digital technology and can obtain assistance from others, this immediate problem resolution significantly improves their motivation to learn and use Internet technology.[46,47] Compared with older adults who cannot obtain technical help, those who receive support show higher digital capabilities.[48] When the elderly face problems using information technology, they typically first seek help from family and friends to overcome technical barriers. This assistance not only resolves the difficulties they encounter in meeting digital needs but also further enhances their digital literacy.

Thirdly, an individual's ability to use technology directly determines the ease with which the elderly can navigate the digital world. According to "The Report on the Digital Divide" Survey results published by the NIA in 2022 and 2023, the digital information level of the older population is generally low, reflecting the severity of the information alienation phenomenon.[19,49] This situation primarily stems from the elderly often not receiving sufficient training during conventional education to enhance their abilities to use digital tools such as computers or smartphones, affecting their skills in searching for information, conducting e-commerce, online banking, handling administrative tasks, and using social networks. The Internet usage ability of the older population is a quality guarantee for accessing digital information;[50,51] therefore, in the parallel phase of digitalization and aging society, improving the digital technological skills of the elderly is an important consideration in addressing the challenges of a digital society.

Fourthly, the COVID-19 pandemic has accelerated the digital transformation of society, which not only changed people's lifestyles but also expanded the market for digital technology applications. Especially for the middle-aged and elderly group, there was a significant change in the awareness of digital information technology and the use of applications related to daily life during the pandemic.[5] According to the NIA's report in 2023, during COVID-19, the elderly had paid more attention to the Internet and mobile technology, with a significant increase in their use of digital technology.[19] This indicates that the change in the social digital environment is also a key factor in predicting the digital information literacy of the elderly.

Additionally, we compared the differences in digital literacy between urban and rural populations. The results indicated that urban residents exhibit significantly higher levels of digital literacy compared with their rural counterparts. This finding may be attributed to the availability of educational resources, the prevalence of information technology, and the opportunities for exposure in urban areas. Urban regions typically have better network infrastructure and more opportunities for digital technology training, which may contribute to the superior performance of urban residents in digital literacy. This disparity could affect the ability of rural populations to access information, participate in social activities, and engage in economic activities, thereby exacerbating the digital divide.

Based on the analysis results, this study proposes a comprehensive approach to improve digital literacy among the elderly, considering education and training, social relationship support, social participation, technical support, and policy formulation.

Firstly, to enhance the elderly's individual technical skills, specifically designed digital literacy courses should be developed. These courses should cover smartphone applications, computer operations, Internet usage, and security measures. The content should be presented in simple, understandable language and involve practical, hands-on methods, enabling the elderly to effectively grasp and apply these skills. Secondly, in terms of social relationship support, family members and friends should be encouraged to actively participate in the elderly's digital learning process. Community organizations and nonprofit institutions should offer specialized digital education programs for the elderly, organize digital technology study groups, and establish volunteer service teams. Young people or technology enthusiasts can provide regular one-on-one technical guidance, helping the elderly solve specific problems and enhancing their digital skills and confidence. Thirdly, from the perspective of social participation, relevant departments should engage in activities aimed at improving digital literacy among the elderly. This includes offering free training courses, developing applications and services tailored for the elderly, and emphasizing the usability and safety of these products during promotion. Additionally, traditional media and social media should be used to publicize the positive impact of digital technology on the quality of life of the elderly, motivating more elderly individuals to try and learn new digital skills. Fourthly, in terms of technical support, electronic devices designed specifically for the elderly should be developed and promoted. These devices should feature large fonts, simplified interfaces, and easy-to-use characteristics. Dedicated technical support hotlines or online help services should be provided to promptly address any issues the elderly encounter while using digital devices and the Internet. Based on feedback from the elderly, digital products and services should be continuously improved and optimized to better meet their needs and enhance their user experience. Finally, the government should formulate differentiated policies to address urban–rural disparities. For urban areas, policy support and financial investment should be used to promote the development and sharing of public digital education resources and to enhance the digitalization of public services. For rural areas, infrastructure investment should be increased to ensure widespread network coverage and access, and financial subsidies should be provided to help rural residents purchase necessary digital devices and participate in training.

Through these comprehensive strategies, the elderly will be better equipped to adapt to the digital environment, improve their digital literacy, enhance their sense of participation and well-being in the digital society, narrow the digital divide, and achieve broader social inclusion.

Conclusion

The implications of this study are as follows. From an academic perspective, the research results demonstrate the effectiveness of machine learning techniques, particularly ensemble learning algorithms, in predicting factors influencing the digital literacy levels of elderly. By combining XGBoost feature importance and SHAP value analysis, this study reveals the significant impact of key features on model predictions. This comprehensive analysis not only enhances the transparency of the model but also provides important insights for subsequent feature engineering and model optimization. Compared with previous research, the ensemble learning algorithms employed in this study exhibit remarkable performance, achieving a classification accuracy exceeding 80%, highlighting a significant advantage over traditional methods. Furthermore, the analysis of variable importance confirms that age, income, occupation, educational background, and individual technological proficiency are crucial factors influencing the digital literacy of elderly. In addition, the study identifies relational support and the social digital environment as predictive factors of digital literacy levels among the elderly. This finding aligns with existing structural equation model research, further validating the application effectiveness of ensemble learning techniques in this research domain. Lastly, the developed and optimized ensemble learning model in this study demonstrates potential for broader application in similar research across other fields, showcasing its theoretical value in practical applications.

The practical significance of this study lies in the targeted intervention recommendations based on the research findings. These recommendations encompass educational training, technical support, policy formulation, and social engagement to enhance the digital literacy of the elderly and bridge the digital divide. These suggestions can provide data support and scientific evidence for governments and relevant institutions to formulate policies, thereby promoting the development of digital literacy education programs and infrastructure for the elderly. Additionally, in the process of enhancing digital literacy among the elderly, this study recommends going beyond a single demographic perspective and individual skill enhancement. It emphasizes the impact of relational support and rapid digital transformation on the digital literacy of the elderly. Furthermore, the study's results indicate that urban populations exhibit significantly higher digital literacy levels compared with rural populations, reflecting an imbalance in the development of digital literacy between urban and rural areas. The overall advantage of urban residents highlights disparities in the dissemination of information technology and educational resources between urban and rural regions. This finding is significant for policymakers, underscoring the need to take measures to reduce the digital divide, improve the digital literacy of rural populations, and promote integrated urban–rural development.

This study has certain limitations and suggests future research directions. Firstly, the study relies on qualitative variables in the database to predict factors influencing the digital literacy of the elderly, which constrains our ability to confirm causal relationships between variables. Future research should adopt longitudinal methods to clarify the causal relationships between these variables. Secondly, the current study focuses on the elderly population without comparing it with other age groups. Subsequent research needs to explore the differences in digital literacy across different generations to comprehensively understand the digital literacy performance and needs of various age groups. Lastly, future research should further investigate the specific factors affecting the digital literacy gap between urban and rural areas, such as education, economic conditions, and social support, to develop more targeted policy measures that help rural populations better adapt to the digital age.

Acknowledgments: All authors would like to thank the NIA, which provided the data.

Contributorship: KH and WX conceived and designed the study; KH developed the methodology; WX acquired the data; KH and WX analyzed and interpreted the data; KH and WX wrote, reviewed, and revised the manuscript.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: Ethical approval was not required as this study data are open-sourced, and the original research has already been conducted.

Funding: The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was supported in part by the Postgraduate Education Reform and Quality Improvement Project of Henan Province (YJS2022JD30).

Guarantor: XW.

ORCID iDs: Haiyan Kong https://orcid.org/0000-0002-6047-5820
Xinyu Wang https://orcid.org/0009-0008-6004-2209

References

1. Sparks JR, Katz IR and Beile PM. Assessing digital information literacy in higher education: a review of existing frameworks and assessments with recommendations for next-generation assessment. *ETS Res Rep Ser* 2016; 2016: 1–33.
2. Weber H, Hillmert S and Rott KJ. Can digital information literacy among undergraduates be improved? Evidence from an experimental study. *Teach High Educ* 2018; 23: 909–926.
3. Van Dijk J. *The deepening divide: Inequality in the information society*. London: Sage, 2005.
4. National Information Society Agency(NIA). *The Report on the Digital Divide of* 2020; National Information Society Agency: Seoul, Republic of Korea, 2021.
5. Oh EA and Bae SM. The relationship between the digital literacy and healthy aging of the elderly in Korea. *Curr Psychol* 2024; 43: 16160–16169.
6. Neves BB, Amaro F and Fonseca JR. Coming of (old) age in the digital age: ICT usage and non-usage among older adults. *Sociol Res Online* 2013; 18: 22–35.

7. Haight M, Quan-Haase A and Corbett BA. Revisiting the digital divide in Canada: the impact of demographic factors on access to the internet, level of online activity, and social networking site usage. *Inf Commun Soc* 2014; 17: 503–519.
8. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. *New Media Soc* 2016; 18: 313–331.
9. Puspitasari L and Ishii K. Digital divides and mobile internet in Indonesia: impact of smartphones. *Telemat Inform* 2016; 33: 472–483.
10. Hwang EH, Shin SJ and Jung DY. A study of the pattern of elderly's internet usage, self-efficacy, and self-esteem. *J Korean Public Health Nurs* 2011; 25: 118–128.
11. Park JR and Feng Y. Trajectory tracking of changes in digital divide prediction factors in the elderly through machine learning. *PLoS One* 2023; 18: e0281291.
12. Aruleba RT, Adekiya TA, Ayawei N, et al. COVID-19 diagnosis: a review of rapid antigen, RT-PCR and artificial intelligence methods. *Bioengineering* 2022; 9: 153.
13. Khan AT, Cao X, Li S, et al. Fraud detection in publicly traded US firms using beetle antennae search: a machine learning approach. *Expert Syst Appl* 2022; 191: 116148.
14. Ho IMK, Cheong KY and Weldon A. Predicting student satisfaction of emergency remote learning in higher education during COVID-19 using machine learning techniques. *PLoS One* 2021; 16: e0249423.
15. Hidalgo A, Gabaly S, Morales-Alonso G, et al. The digital divide in light of sustainable development: an approach through advanced machine learning techniques. *Technol Forecast Soc Change* 2020; 150: 119754.
16. Ahmadi M, Nopour R and Nasiri S. Developing a prediction model for successful aging among the elderly using machine learning algorithms. *Digit Health* 2023; 9: 1–22, Article 20552076231178425.
17. Kong H and Liu H. The relationship between ICT use and perceived life satisfaction among older people in Korea: the mediating effect of social capital. *Sustainability* 2023; 15: 9353.
18. Cui K, Zou W, Ji X, et al. Does digital technology make people healthier: the impact of digital use on the lifestyle of Chinese older adults. *BMC Geriatr* 2024; 24: 1–12.
19. National Information Society Agency(NIA). *The Report on the Digital Divide of* 2022; National Information Society Agency: Seoul, Republic of Korea, 2023.
20. Al-Hashem MA, Alqudah AM and Qananwah Q. Performance evaluation of different machine learning classification algorithms for disease diagnosis. *Int J E-Health Med Commun* 2021; 12: 1–28.
21. Chen H, Wu L, Chen J, et al. A comparative study of automated legal text classification using random forests and deep learning. *Inf Process Manag* 2022; 59: 102798.
22. Graham CA, Shamkhalichenar H, Browning VE, et al. A practical evaluation of machine learning for classification of ultrasound images of ovarian development in channel catfish (*Ictalurus punctatus*). *Aquaculture* 2022; 552: 738039.
23. Mienye ID and Sun Y. Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. *Inform Med Unlocked* 2021; 25: 100690.
24. Mienye ID, Sun Y and Wang Z. Improved predictive sparse decomposition method with densenet for prediction of lung cancer. *Int J Comput* 2020; 1: 533–541.
25. Coria SR, Mondragón-Becerra R, Pérez-Meza M, et al. CT4RDD: classification trees for research on digital divide. *Expert Syst Appl* 2013; 40: 5779–5786.
26. Kang H, Baek J, Chu SH, et al. Digital literacy among Korean older adults: a scoping review of quantitative studies. *Digit Health* 2023; 9: 20552076231197334.
27. Jun W. A study on cause analysis of digital divide among older people in Korea. *Int J Environ Res Public Health* 2021; 18: 8586.
28. Cho M and Kim KM. Effect of digital divide on people with disabilities during the COVID-19 pandemic. *Disabil Health J* 2022; 15: 101214.
29. Jiang X, Zhang Y, Li Y, et al. Forecast and analysis of aircraft passenger satisfaction based on RF-RFE-LR model. *Sci Rep* 2022; 12: 11174.
30. Capitaine L, Genuer R and Thiébaut R. Random forests for high-dimensional longitudinal data. *Stat Methods Med Res* 2021; 30: 166–184.
31. Wang F, Li Z, He F, et al. Feature learning viewpoint of AdaBoost and a new algorithm. *IEEE Access* 2019; 7: 149890–149899.
32. Mienye ID and Sun Y. A survey of ensemble learning: concepts, algorithms, applications, and prospects. *IEEE Access* 2022; 10: 99129–99149.
33. Liang W, Luo S, Zhao G, et al. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics* 2020; 8: 765.
34. Al-Sarem M, Saeed F, Boulila W, et al. Feature selection and classification using CatBoost method for improving the performance of predicting Parkinson's disease. In: *Advances on smart and soft computing: proceedings of ICACIn 2020*. Singapore: Springer; 2021. pp.189–199.
35. Alam MN, Podder P, Bharati S, et al. Effective machine learning approaches for credit card fraud detection. In: *Innovations in bio-inspired computing and applications: proceedings of the 11th international conference on innovations in bio-inspired computing and applications (IBICA 2020)*. Cham: Springer International Publishing; 2021. pp.154–163.
36. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 2017: 30. https://dl.acm.org/doi/abs/10.5555/3294996.3295074
37. Cui Z, Qing X, Chai H, et al. Real-time rainfall-runoff prediction using light gradient boosting machine coupled with singular spectrum analysis. *J Hydrol* 2021; 603: 127124.
38. Prokhorenkova L, Gusev G, Vorobev A, et al. Catboost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst* 2018: 31. https://dl.acm.org/doi/abs/10.5555/3327757.3327770
39. Dorogush AV, Ershov V and Gulin A. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363. 2018.
40. Lu C, Zhang S, Xue D, et al. Improved estimation of coalbed methane content using the revised estimate of depth and CatBoost algorithm: a case study from southern Sichuan Basin, China. *Comput Geosci* 2022; 158: 104973.
41. Zadražilová I and Vizváry P. Digital literacy competencies and interests of elderly people. In: *European conference on information literacy*. Cham: Springer International Publishing; 2021. pp.137–146.

42. Chen K and Chan AHS. A review of technology acceptance by older adults. *Gerontechnology* 2011; 10: 1–12.
43. Wallinheimo AS and Evans SL. More frequent internet use during the COVID-19 pandemic associates with enhanced quality of life and lower depression scores in middle-aged and older adults. *Healthc (Basel)* 2021; 9: 393.
44. Gaia A, Sala E and Cerati G. Social networking sites use and life satisfaction. A quantitative study on older people living in Europe. *Eur Societies* 2021; 23: 98–118.
45. Sun X, Yan W, Zhou H, et al. Internet use and need for digital health technology among the elderly: a cross-sectional survey in China. *BMC Public Health* 2020; 20: 1–8.
46. Xie XL, Chen Y, Lao YX, et al. Status quo, influencing factors and coping strategies of internet use among the elderly. *Chin J Gerontol* 2017; 13: 1.
47. Liu Z, Zhang H, Zhang Y, et al. Current situation and influencing factors of e-health literacy among rural older adults in Zhengzhou. *Modern Prevent Med* 2020; 47: 283–309.
48. Shi Y, Ma D, Zhang J, et al. In the digital age: a systematic literature review of the e-health literacy and influencing factors among Chinese older adults. *J Public Health (Berl)* 2023; 31: 679–687.
49. National Information Society Agency(NIA). *The Report on the Digital Divide of* 2021; National Information Society Agency: Seoul, Republic of Korea, 2022.
50. Sabo RM. Lifelong learning and library programming for third agers. *Libr Rev* 2017; 66: 39–48.
51. Xiong J and Zuo M. How does family support work when older adults obtain information from mobile internet? *Inf Technol People* 2019; 32: 1496–1516.