Journal of Big Data

RESEARCH Open Access

CRNet: a multimodal deep convolutional neural network for customer revisit prediction

Eunil Park[1,2,3*]

*Correspondence:
eunilpark@skku.edu

[1] Department of Interaction Science, Sungkyunkwan University, Seoul, Republic of Korea
[2] Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul, Republic of Korea
[3] Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul, Republic of Korea

Abstract

Since mobile food delivery services have become one of the essential issues for the restaurant industry, predicting customer revisits is highlighted as one of the significant academic and research topics. Considering that the use of multimodal datasets has gained notable attention from several scholars to address multiple industrial issues in our society, we introduce CRNet, a multimodal deep convolutional neural network for predicting customer revisits. We evaluated our approach using two datasets [a customer repurchase dataset (CRD) and mobile food delivery revisit dataset (MFDRD)] and two state-of-the-art multimodal deep learning models. The results showed that CRNet obtained accuracies and Fi-Scores of 0.9575 (CRD) and 0.9436 (MFDRD) and 0.9730 (CRD) and 0.9509 (MFDRD), respectively, thus achieving higher performance levels than current state-of-the-art multimodal frameworks (accuracy: 0.7417–0.9012; F1-Score: 0.7461–0.9378). Future research should aim to address other resources that can enhance the proposed framework (e.g., metadata information).

Keywords: Customer revisit, CRNet, MFDRD, Customer repurchase

Introduction

With rapidly advanced information and communication technologies, mobile applications are being used with ease. Mobile applications provide easy access to information and utilitarian values of online services for customers [1]. Moreover, recent developments and the distribution of mobile devices and technologies have improved industries. For instance, several restaurants provide services to customers inside the restaurant and via door-to-door delivery. In addition, owing to the COVID-19 pandemic, food delivery services have rapidly improved and diffused in several nations [2].

To minimize the potential contagion of the SARS-CoV-2 virus [3], most of the nations imposed stay-at-home policies and firmly recommended the use of mobile food delivery services (MFD services). This has also been applied in South Korea—a nation with advanced delivery services [4].

This also relates to restaurants' survival, who need to respond to the huge demands of MFD services. Considering the findings of and definition provided by prior research on delivery services [5], MFD services are defined as "*food ordering/selecting/delivering systems, which connect specific restaurants, customers, and payment services through*

mobile applications" [5, 6]. During this pandemic, the global market for MFD services has increased significantly. For instance, the global online food delivery markets in 2021 are estimated at approximately 127 billion USD, with 10% annual compound growth rate of approximately 115 billion USD.[1]

In line with this, a number of scholars have investigated customer behavior in terms of both the motivations for and hindrances of employing online food delivery services. Yeo et al. [7] considered 224 customers to explore their motivations for using online food delivery services by considering hedonic factors. Furthermore, Suhartanto et al. [8] investigated customer loyalty to online food delivery service providers by considering the perceived quality of the services. However, they were limited to addressing customers' real service repurchases with several limitations–including the difficulties of data securement.

To efficiently address this issue, several scholars employed data mining or machine learning approaches for customer revisit behavior [9]. As one of the representative examples, Nilashi et al. [10] employed text mining, clustering and predictive regression techniques for customer decision making procedures via 4461 online review comments of 35 vegetarian friendly restaurants. However, the majority of prior examples employed single modality in addressing customer revisits (e.g. text [11]). Recent approaches indicated that employing multimodalities can contribute to the better understanding of customer-related tasks in several domains (e.g. healthcare [12], social media [13]).

To address this, we propose CRNet, a multimodal deep convolutional neural network for predicting customer revisits in MFD services. "Related work" section presents several cornerstones of data-driven approaches for customer revisits. "Our approach" section discusses the experiment conducted using the collected dataset. Both the results and implications are presented in "Experimental setup" and "Experimental results" sections, respectively.

Related work

Considering the rapidly improving restaurant market, a customer's intention to revisit particular restaurants have been consistently addressed. Because the intention to revisit is one of the most effective predictors of service success, marketing and management primarily focus on intention. For instance, Kumar et al. [14] explored customers' intention to revisit online food delivery applications through two theoretical frameworks: the stimulus-organism-response and pleasure arousal dominance theories. Based on 446 responses, the mediating roles of pleasure and arousal as well as the indirect determinant of aesthetic formality on revisit intention were confirmed with good fit indices [RMSEA (root mean square error of approximation) = 0.06, CFI (comparative fit index) = 0.92].

Rajput and Gahfoor [15] studied customers' intention to revisit fast food restaurants through the responses of 433 customers. Considering the conceptual research framework including food/service/environment quality and customer satisfaction, the

[1] https://www.prnewswire.com/news-releases/global-online-food-delivery-services-market-report-2021-market-is-expected-to-reach-192-16-billion-in-2025--from-126-91-billion-in-2021---long-term-forecast-to-2030--301285677.html.

structural results confirmed the direct motivations of customer satisfaction (0.528) and word of mouth (0.312), indirect factors, and the three quality dimensions.

Han et al. [16] investigated the relationship between revisit intention, perceived satisfaction, consumption emotions, and switching barriers through structural path analysis and qualitative approaches. With 406 validated samples in the United States, the moderating effect of customer satisfaction ($\beta = 0.71$) and the significant roles of comfort and annoyance on revisit intention were examined.

Meng and Choi [17] introduced a research model that analyzes customer' behavioral intention to revisit theme restaurants, based on the theory of planned behavior. Based on the results of an on-site survey with 357 customers, it was concluded their attitude and involvement play a mediating role in determining their intention to revisit.

Although several scholars have focused on addressing customer revisit behavior [18], there are several limitations. One such limitation is that there are a huge number of factors, which can have notable impacts on the academic generalization of customer revisits. In addition, it is too difficult to address customer 'actual revisits' with traditional data-driven approaches.

Due to the rapidly improved data-driven approaches and technologies for analyzing customer behavior, several recent scholars have addressed the "exact" and "direct" meanings of customer revisit. Furthermore, recent data analytics, as well as machine and deep learning approaches have allowed researchers to analyze customer revisits. For example, Hwang et al. [11] collected the responses of 133,872 airline service customers, extracted user experience features from customers' responses, and investigated whether each customer revisits the same airline. User experience dimensions are based on machine learning approaches. They presented that an 83.42% accuracy is achieved in predicting customer revisits in terms of airline services.

Kim and Lee [19] also proposed a systematic framework for predicting customer revisit intention for seven flagship stores. Based on three feature groups—upcoming events, group movements, and store accessibility with approximately 3.75 unique customers in the stores—67–80% accuracy was achieved in predicting customer revisits using the XGBoost classifier.

Kim et al. [20] also proposed a deep neural network framework to address customer repurchase behavior. Considering a 2-year survey including 119,923 (the first round) and 74,088 (the second round) respondents, and the framework integrated by the three sub-modules of long short-term memory (LSTM) (customer comments), convolutional neural network (CNN) (model images), and deep convolutional neural network layers (evaluation rating with brands), 95.13% recall, 94.18% F1-score, 93.25% precision (same brand), and 90.71% accuracy were achieved in predicting whether each customer purchases a smartphone from the same brand.

Our approach

We provide a brief explanation of CRNet, a multimodal deep CNN for customer revisit prediction in MFD services. An overview of CRNet is presented in Fig. 1. CRNet is organized into three modules: (1) a review comment module, (2) a review image module, and (3) an integrated module.

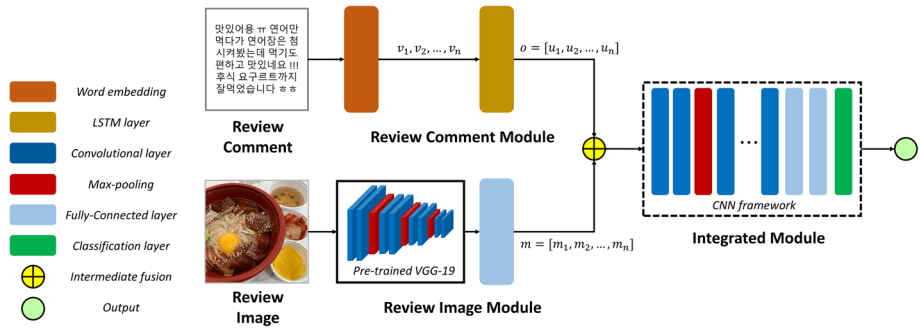

Fig. 1 Overview of CRNet

Review comment module

Each review comment was sequentially embedded with a size of 128. This means that each word in the comment is presented as a single vector, v_i. Thus, each comment is presented as $[v_1, v_2, \ldots v_n]$, where the length of the comment is denoted by n. Then, a LSTM network is employed to present the sequential relationships of each comment with 300 units. The word embedding sequence, v_i, and the outcome of the previous unit u_{i-1} from the LSTM units are integrated to present the outcome of word representation, u_i. With this procedure, the output of the LSTM network as the text feature, o, is presented and denoted as $o = [u_1, u_2, \ldots, u_n]$. This means that unique text feature presentations are used to show the inputs for the proposed model [21].

Review image module

Each input image is adjusted to 224×224 for the VGG-19 network, which is pretrained using ImageNet. The output of the VGG network is used as the image feature. Moreover, both 6×6 regions with 256 vector dimensions are employed for each image. Thus, m^* is presented and organized by $m^* = [m_1^*, m_2^*, \ldots, m_n^*]$. Then, the fully connected layer is added to convert each feature vector into a new vector with the same dimensions as the text feature presentations. Thus, $m = [m_1, m_2, \ldots, m_n]$ is presented as the image feature matrix [22].

Integrated module

We used intermediate fusion, which refers to "*the procedures of integrating learned feature representations from intermediate layers of neural networks with features from other modalities as input to a final model*" [23], because this fusion approach shows better performance than other fusion approaches, such as early and late (aggregation) fusions, in this task.

The feature representations extracted from review comments and images are employed as the input to the integrated module. The integrated module is organized into two CNN frameworks. We employed rectified linear units (ReLUs) for all layers without the classification layer. Softmax activation functions are used for the classification layer.

The first CNN framework is composed of 12 convolutional layers (3×3) with 16 filters and three fully connected layers. It has 128 neurons, and a 0.2 dropout rate. The

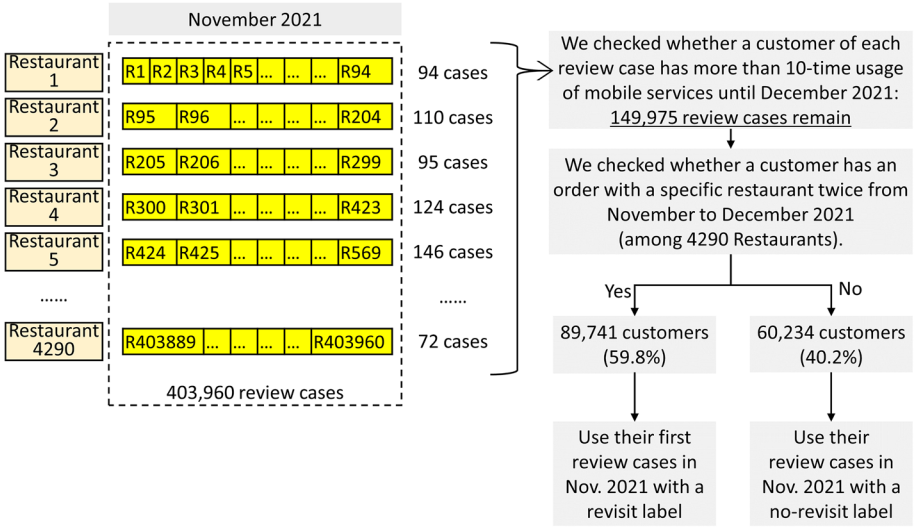

Fig. 2 Summary of the data collection and labelling procedures

max-pooling layer (4×4) is inserted in every two convolutional layers. The second CNN framework includes nine layers: six convolutional layers, two fully connected layers, and a softmax classification layer. The first and last three convolutional layers have 32 and 64 filters, respectively. We employed GridsearchCV to determine the most optimal parameters in 75 epochs, and the Adam optimizer was employed with a learning rate of 0.001.

Experimental setup

Datasets

Because only a limited number of scholars have provided datasets for customer revisits, we employed two approaches. First, well-known datasets—which not only address the concept of customer revisits but also consider both images and text features. Thus, the customer repurchase dataset, CRD [20], which includes 74,088 samples (58,175 for repurchases and 15,913 for non-repurchases), was considered. This dataset includes customer review comments and model images, which we employed as review comments and images, respectively.

Second, we aimed to collect a dataset from a popular MFD service in Korea, namely the MFD revisit dataset (MFDRD). Figure 2 shows the summary of the data collection and labelling procedures. We collected 403,960 review cases, which included both review comments with more than 10 words with an image regarding 4290 restaurants in November 2021. Then, in December 31, 2021, we checked each customer who wrote a review, to investigate whether the customer ordered from the same restaurant for more than 10 times.

Among 403,960 review cases, 149,975 review cases—customers who used a service more than 10 times—were employed. Among them, 89,741 (59.8%) customers revisited the same restaurants; while 60,234 customers (40.2%) did not have a previous order with the restaurant in November 2021. Figure 3 shows an example of the

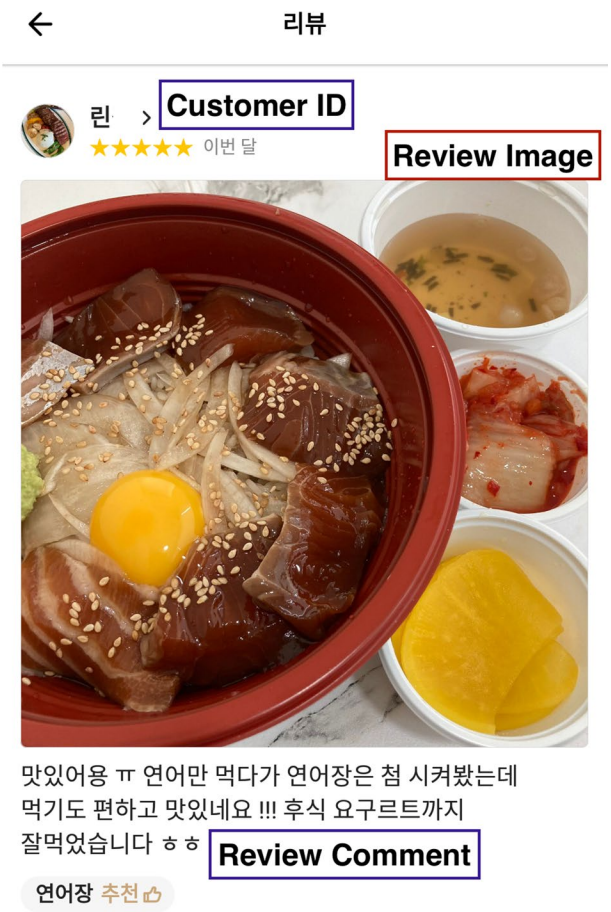

Fig. 3 Sample case from the mobile food delivery revisit dataset

customer review comments and images of MFDs. In the experiment, 60%, 20%, and 20% of the collected cases were employed for training, validation, and testing, respectively. The sample cases are presented in Fig. 3 and in Additional file 1.

Baseline models

Because our baseline models, as well as CRNet, should address both images and review comments at the same time, both the text-image fusion model [24] and multi-view attentional network model (MVAN) were employed as our baseline models [25].

- Text-image fusion model: an architecture identical to that of a CNN-based text and image fusion model proposed by Gallo et al. [24] was employed. To classify whether each sample is a revisit case, a softmax output layer was attached after the last fully connected layer. Moreover, GridsearchCV was used to find the most optimal parameters (learning rate: 0.01, epoch: 100, ADAM optimizer).
- MVAN: an architecture identical to the MVAN model proposed by Yang et al. [25] was employed, which is organized in the order of feature mapping, interactive learning, and feature fusing modules. To examine a binary revisit classification, the softmax output layer was revised. GridsearchCV was also employed to find the most effective parameters (learning rate: 0.005, epoch: 70, ADAM optimizer).

Table 1 Summary of the results with two datasets

Models	Class	Precision	Recall	F1-Score	Accuracy
Dataset: CRD					
Text-image fusion model	Revisit	0.8753	0.7825	0.8263	0.7417
	Not	0.4266	0.5921	0.4959	
MVAN	Revisit	0.9280	0.9478	0.9378	0.9012
	Not	0.7926	0.7308	0.7605	
CRNet	Revisit	**0.9700**	**0.9760**	**0.9730**	**0.9575**
	Not	0.9102	0.8896	0.8998	
Dataset: MFDRD					
Text-image fusion model	Revisit	0.7725	0.7215	0.7461	0.7065
	Not	0.6231	0.6843	0.6522	
MVAN	Revisit	0.9364	0.9106	0.9098	0.8951
	Not	0.8415	0.8846	0.8747	
CRNet	Revisit	**0.9915**	0.9136	**0.9509**	**0.9436**
	Not	0.8850	**0.9883**	0.9338	

The bold values mean the greatest performance levels

Evaluation metrics

We employed four evaluation metrics to test both baseline models and CRNet: *accuracy*, *precision*, *recall*, and *F1-Score*. The following equations were employed to compute *precision*, *recall*, *F1-Score*, and *accuracy*:

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{False Negative (FN)}}$$

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

$$\text{Accuracy} = \frac{\text{TP} + \text{True Negative (TN)}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

Experimental results

Table 1 summarizes the experimental results. On the first dataset, CRD, the proposed model, CRNet achieved an accuracy of 95.75%, which is significantly higher than those of the text-image fusion model (74.17%) and MVAN (90.12%). The F1-score showed a similar trend considering the customer revisit class. Considering the characteristics of the proposed research question, the F1-score of CRNet in the revisit class (0.9730) was higher than those of MVAN (0.9378) and the text-image fusion model (0.7825).

Similar results were also obtained with MFDRD. CRNet (94.36%) achieved higher accuracy than MVAN (89.51%) and the text-image fusion model (70.65%). Moreover, all other metrics of CRNet were higher than those of the other models.

Discussion and conclusion

In this study, we examined CRNet, a multimodal deep learning framework for predicting customer revisits. Two datasets that address customer revisits were employed in our experiment. In addition, two state-of-the-art multimodal deep learning models for binary classification tasks were employed to evaluate the proposed framework. After developing the text and image modules, the integrated module was attached to determine whether there was a customer revisit.

The experimental results on two datasets, CRD and MFDRD, indicate that the proposed model, CRNet, can effectively examine customer revisit behavior using synchronized multimodal cases—including textual and image contexts. Given three modules and two datasets, we achieved better accuracy and higher scores compared with other baseline models.

From a managerial perspective, considering that the implementation of CRNet is effective in the experiment, service providers, who have multimodal datasets, can easily employ CRNet to explore their customers' revisit behavior. In addition, consistent with the findings of prior research [21, 26], we confirmed that employing multimodal datasets can present the greater performance than using single-modality datasets in specific tasks.

Although several implications have been provided, notable concerns remain. First, the suggested framework, CRNet, is only effective with synchronized datasets—which are organized by both image and text information. When a partial dataset is provided, the suggested framework cannot be effectively applied. Second, although the employed services in this study do not allow multiple accounts in Korea, there may be different accounts, which can use the actual delivery services. Third, the proposed framework was examined using two datasets. This means that comparable results cannot be obtained on datasets that address customer revisits for other services. Moreover, future research should aim to address other resources for customer revisit information, which can enhance the proposed framework (e.g., metadata information).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40537-022-00674-4.

Additional file 1. Samples of the collected responses.

Acknowledgements
Not applicable.

Author contributions
Park fully contributed to the design, implementation, and analysis of the research with the examination of the manuscript. The author read and approved the final manuscript.

Funding
This research was supported by National Research Foundation (NRF) of Korea Grant funded by the Korean Government (MSIT) (No. 2021R1A4A3022102). This work was also supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00358, AI·Big data based Cyber Security Orchestration and Automated Response Technology Development).

Availability of data and materials
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The author declares that he has no competing interests.

Received: 27 February 2022 Accepted: 29 November 2022
Published online: 03 January 2023

References

1. Park E, Ohm J. Factors influencing users employment of mobile map services. Telemat Inform. 2014;31(2):253–65.
2. Mehrolia S, Alagarsamy S, Solaikutty VM. Customers response to online food delivery services during COVID-19 outbreak using binary logistic regression. Int J Consum Stud. 2021;45(3):396–408.
3. Kim J, Aum J, Lee S, Jang Y, Park E, Choi D. FibVID: comprehensive fake news diffusion dataset during the COVID-19 period. Telemat Inform. 2021;64: 101688.
4. Kim JJ, Kim I, Hwang J. A change of perceived innovativeness for contactless food delivery services using drones after the outbreak of COVID-19. Int J Hosp Manag. 2021;93: 102758.
5. Hong C, Choi HH, Choi E-KC, Joung H-WD. Factors affecting customer intention to use online food delivery services before and during the COVID-19 pandemic. J Hosp Tour Manag. 2021;48:509–18.
6. Ray A, Dhir A, Bala PK, Kaur P. Why do people use food delivery apps (FDA)? A uses and gratification theory perspective. J Retail Consum Serv. 2019;51:221–30.
7. Yeo VCS, Goh S-K, Rezaei S. Consumer experiences, attitude and behavioral intention toward online food delivery (OFD) services. J Retail Consum Serv. 2017;35:150–62.
8. Suhartanto D, Helmi Ali M, Tan KH, Sjahroeddin F, Kusdibyo L. Loyalty toward online food delivery service: the role of e-service quality and food quality. J Foodserv Bus Res. 2019;22(1):81–97.
9. Park E, Kang J, Choi D, Han J. Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews. Curr Issues Tour. 2020;23(5):605–11.
10. Nilashi M, Ahmadi H, Arji G, Alsalem KO, Samad S, Ghabban F, Alzahrani AO, Ahani A, Alarood AA. Big social data and customer decision making in vegetarian restaurants: a combined machine learning method. J Retail Consum Serv. 2021;62: 102630.
11. Hwang S, Kim J, Park E, Kwon SJ. Who will be your next customer: a machine learning approach to customer return visits in airline services. J Bus Res. 2020;121:121–6.
12. Rahman MA, Hossain MS, Alrajeh NA, Gupta B. A multimodal, multimedia point-of-care deep learning framework for COVID-19 diagnosis. ACM Trans Multimed Comput Commun Appl. 2021;17(1s):1–24.
13. Chandrasekaran G, Nguyen TN, Hemanth DJ. Multimodal sentimental analysis for social media applications: a comprehensive review. Wiley Interdiscip Rev Data Min Knowl Discov. 2021;11(5):1415.
14. Kumar S, Jain A, Hsieh J-K. Impact of apps aesthetics on revisit intentions of food delivery apps: the mediating role of pleasure and arousal. J Retail Consum Serv. 2021;63: 102686.
15. Rajput A, Gahfoor RZ. Satisfaction and revisit intentions at fast food restaurants. Future Bus J. 2020;6:1–12.
16. Han H, Back K-J, Barrett B. Influencing factors on restaurant customers' revisit intention: the roles of emotions and switching barriers. Int J Hosp Manag. 2009;28(4):563–72.
17. Meng B, Choi K. An investigation on customer revisit intention to theme restaurants: the role of servicescape and authentic perception. Int J Contemp Hosp Manag. 2018;30(3):1646–62.
18. Park E. Motivations for customer revisit behavior in online review comments: analyzing the role of user experience using big data approaches. J Retail Consum Serv. 2019;51:14–8.
19. Kim S, Lee J-G. A systematic framework of predicting customer revisit with in-store sensors. Knowl Inf Syst. 2020;62(3):1005–35.
20. Kim J, Ji H, Oh S, Hwang S, Park E, del Pobil AP. A deep hybrid learning model for customer repurchase behavior. J Retail Consum Serv. 2021;59: 102381.
21. Oh S, Ji H, Kim J, Park E, del Pobil AP. Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. Inf Technol Tour. 2022;24(1):109–26.
22. Kim D, Choi J, Ahn S, Park E. A smart home dental care system: integration of deep learning, image sensors, and mobile controller. J Ambient Intell Humaniz Comput. 2021. https://doi.org/10.1007/s12652-021-03366-8.
23. Huang S-C, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3(1):1–9.
24. Gallo I, Calefati A, Nawaz S, Janjua MK. Image and encoded text fusion for multi-modal classification. In: 2018 digital image computing: techniques and applications (DICTA). IEEE; 2018. p. 1–7.
25. Yang X, Feng S, Wang D, Zhang Y. Image-text multimodal emotion classification via multi-view attentional network. IEEE Trans Multimed. 2020;23:4014–26.

26. Gu F, Chung M-H, Chignell M, Valaee S, Zhou B, Liu X. A survey on deep learning for human activity recognition. ACM Comput Surv. 2021;54(8):1–34.

Publisher's Note