Journal of Biomedical Informatics 137 (2023) 104268

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

Deep-learning-based personalized prediction of absolute neutrophil count recovery and comparison with clinicians for validation

Hyunwoo Choo [a,b], Su Young Yoo [a], Suhyeon Moon [c], Minsu Park [d], Jiwon Lee [e], Ki Woong Sung [e], Won Chul Cha [f], Soo-Yong Shin [a,b,c,*], Meong Hi Son [e,*]

[a] Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea
[b] Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Seoul, Republic of Korea
[c] Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea
[d] Department of Information and Statistics, Chungnam National University, Korea 99 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea
[e] Department of Pediatrics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[f] Department of Emergency Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
Neutropenia
Deep learning model
Absolute neutrophil count recovery
Artificial intelligence

ABSTRACT

Neutropenia and its complications are major adverse effects of cytotoxic chemotherapy. The time to recovery from neutropenia varies from patient to patient, and cannot be easily predicted even by experts. Therefore, we trained a deep learning model using data from 525 pediatric patients with solid tumors to predict the day when patients recover from severe neutropenia after high-dose chemotherapy. We validated the model with data from 99 patients and compared its performance to those of clinicians. The accuracy of the model at predicting the recovery day, with a 1-day error, was 76%; its performance was better than those of the specialist group (58.59%) and the resident group (32.33%). In addition, 80% of clinicians changed their initial predictions at least once after the model's prediction was conveyed to them. In total, 86 prediction changes (90.53%) improved the recovery day estimate.

1. Introduction

Chemotherapy-induced neutropenia is the primary dose-limiting toxicity in patients with cancer treated with chemotherapy [1]. According to the widely used Common Toxicity Criteria from the National Cancer Institute (NCI), neutropenia has four grades of severity; when it is < 500 /μL it is defined as grade four toxicity [2]. Because neutrophils respond to bacteria, viruses, and fungi, patients with neutropenia are more susceptible to infections. In the case of neutropenic fever, empirical antibiotic treatment is required to prevent patients from progressing to sepsis [3]. Patients treated with high-dose chemotherapy who are expected to have severe neutropenia are isolated in the transplant unit (if the facility is available) until they recover from severe neutropenia. Therefore, it is important to approximately predict the day of the recovery to prepare for the individual's risk of infection and for the planning of the transplant units.

Previously, predicting the recovery from myelosuppression mainly relied on pharmacokinetic-pharmacodynamic (PK-PD) models and mathematical modeling. Friberg et al. introduced models to predict chemotherapy-induced myelosuppression using both drug-specific parameters and common-to-all-drug parameters [4]. As they included the cell supply and feedback mechanism, their model is considered the gold standard for myelosuppression modeling. However, it requires some empirically derived methods, such as a constant transit time between transit compartments or the mean maturation time from proliferation to the circulation of stem cells, which limits its application in most clinical cases [5]. Several PK-PD models also improved Friberg's model, focusing on specific drugs (6-mercaptopurine and sunitinib) [6] or specific protocols (NOPHO ALL-2008) [7]. Currently, multiple drugs are used for different protocols [8], and the doses and concentrations of

Abbreviations: NCI, National Cancer Institute; PK-PD, pharmacokinetic-pharmacodynamic; ANC, absolute neutrophil count; EHR, electronic health record; CTCAE, Common Terminology Criteria for Adverse Events; MI-CLAIM, minimum information about clinical artificial intelligence modelling; CONSORT, consolidated standards of reporting trials–artificial intelligence; CDW, clinical data warehouse; G-CSF, granulocyte colony-stimulating factor; TFT, temporal fusion transformer; sMAPE, symmetric mean absolute percentage error; TAM, technology acceptance model; IQR, interquartile range.

* Corresponding authors at: Department of Digital Health, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea (S.-Y. Shin).
E-mail addresses: sy.shin@skku.edu (S.-Y. Shin), meonghi.son@samsung.com (M.H. Son).

https://doi.org/10.1016/j.jbi.2022.104268
Received 10 June 2022; Received in revised form 27 November 2022; Accepted 7 December 2022
Available online 10 December 2022

each drug are determined by various factors, such as the body surface area and the patient's renal function [9]. Therefore, it is difficult to obtain generalized predictions using the PK-PD model.

Several studies have used machine learning to predict absolute neutrophil count (ANC) values. Netterberg et al. used support vector machines to predict future ANC values [10]. Cuplov et al. modeled the effect of ifosfamide on neutrophils and platelets using a gradient boosting regression method that included an explicit interpolation of the data [11]. Both studies showed the potential for data-driven models to predict hematologic recovery; however, due to the nature of the drug-focused model, various variables affecting the recovery from myelosuppression were not considered as factors [12,13]. The electronic health record (EHR) data contains most of the possible factors affecting myelosuppression but has the distinctive character of irregularity in terms of input interval since they are derived from the patient's irregular visits to the hospital [14]. Understanding the known limitation of the EHR data, we aimed to find the best possible model for ANC recovery after high-dose chemotherapy in pediatric patients with solid tumors

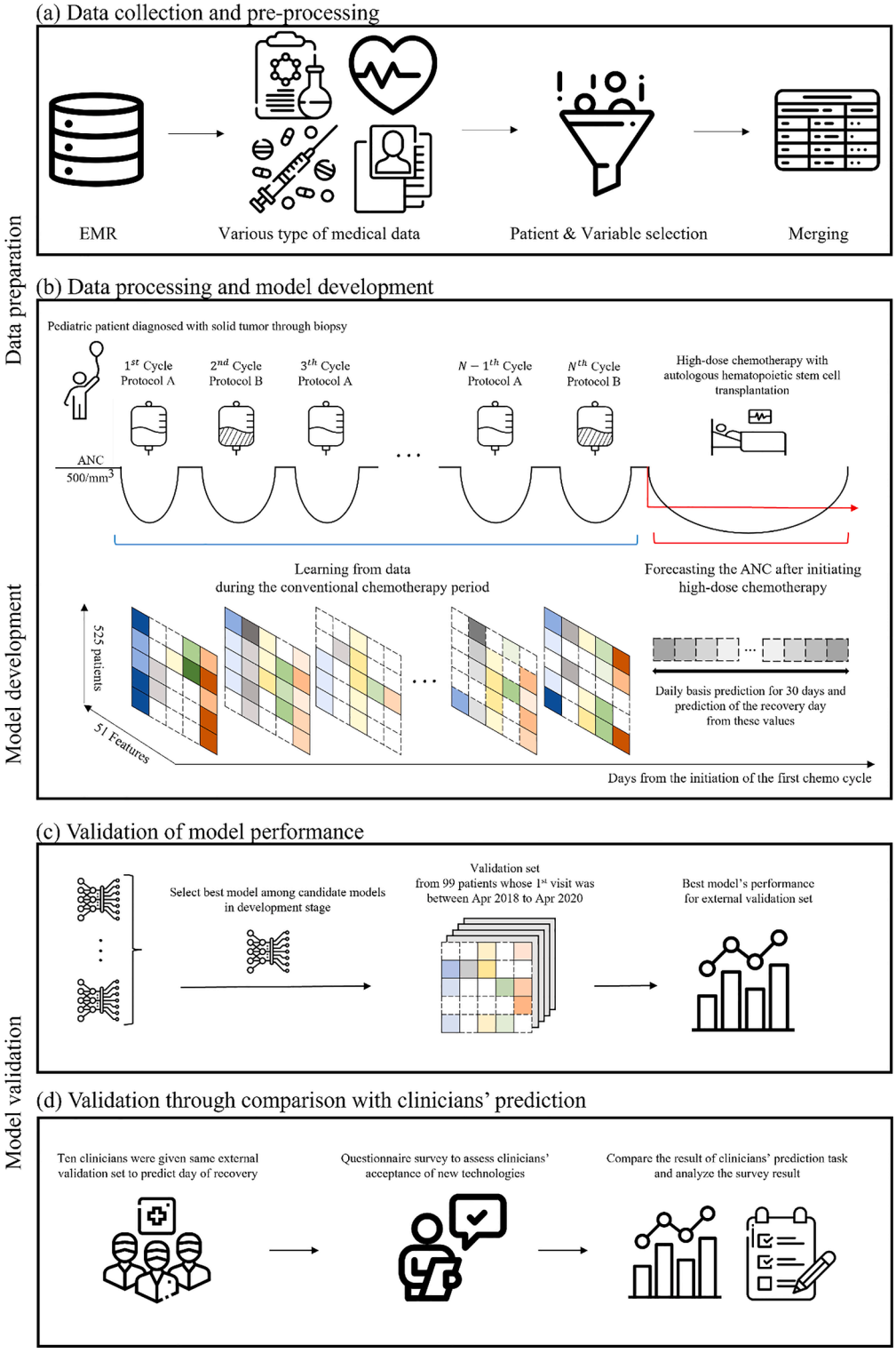

Fig. 1. Study overview (a) Data collected from the Samsung Medical Center and filtered according to the inclusion and exclusion criteria. (b) According to the patient's chemotherapy schedule, all data entries are annotated regarding the day of chemotherapy and the inclusion of that day's data regarding neutropenia$_{500}$. For each cycle, data were labeled by subtracting the conventional chemotherapy start day from the recovery day. The deep learning model predicted the ANC change for 30 days from the high-dose chemotherapy start day and calculated the recovery day from the prediction value. (c) Candidate models were validated using test datasets and the best-performing models were selected. (d) The performance of the model was compared with that of the clinicians. Additionally, a survey was conducted.

overcoming irregularly sampled and limited size of data.

Herein, we trained some state-of-the-art deep learning models and selected the best performing model for predicting personalized ANC recovery after high-dose chemotherapy in pediatric patients with solid tumors and compared the predictive ability of the model with that of clinicians.

2. Material and methods

2.1. Definition

The criterion of severe neutropenia is an ANC value < 500 / μL, which is a grade four toxicity by Common Terminology Criteria for Adverse Events (CTCAE v5.0) [2] from NCI, and this period is denoted by neutropenia$_{500}$. Recovery day, which is the predictive goal of the deep learning model, is the first day when the ANC value becomes ≥ 500 /μL following at least three consecutive days of ANC values < 500/μL after the initiation of high-dose chemotherapy.

2.2. Study design

The overall process of the study is summarized in Fig. 1. We followed the Minimum information about clinical artificial intelligence modeling (MI-CLAIM) checklist and consolidated standards of reporting trials–artificial intelligence (CONSORT-AI) checklist to report the results of the clinical artificial intelligence model [15,16].

To predict the recovery day from neutropenia after administration of high-dose chemotherapy in pediatric patients with solid tumors, some candidate deep learning models were trained with the training dataset. In the training process, data for 10 % of the cases were saved for internal validation. Tests were performed with the test dataset from cases of the same condition; the best performing model was then selected. To evaluate the performance of the model, the enrolled clinicians were asked to predict the recovery day from neutropenia for each case included in the test dataset. Participants were then asked again if they would change the decision after the model predicted value was given. The clinicians who participated in the comparison also responded to the questionnaire designed and developed for this study.

The institutional review board of the Samsung Medical Center approved the study and waived the need for informed consent (IRB No: 2019-11-205 & 2021-02-059).

2.3. Data collection

We retrospectively collected data from the clinical data warehouse (CDW) at the Samsung Medical Center. The training dataset was collected from pediatric patients with solid tumors between January 2000 and April 2018. The test dataset was collected from the patients between April 2018 and April 2020. The detailed inclusion and exclusion criteria are described in Fig. 2.

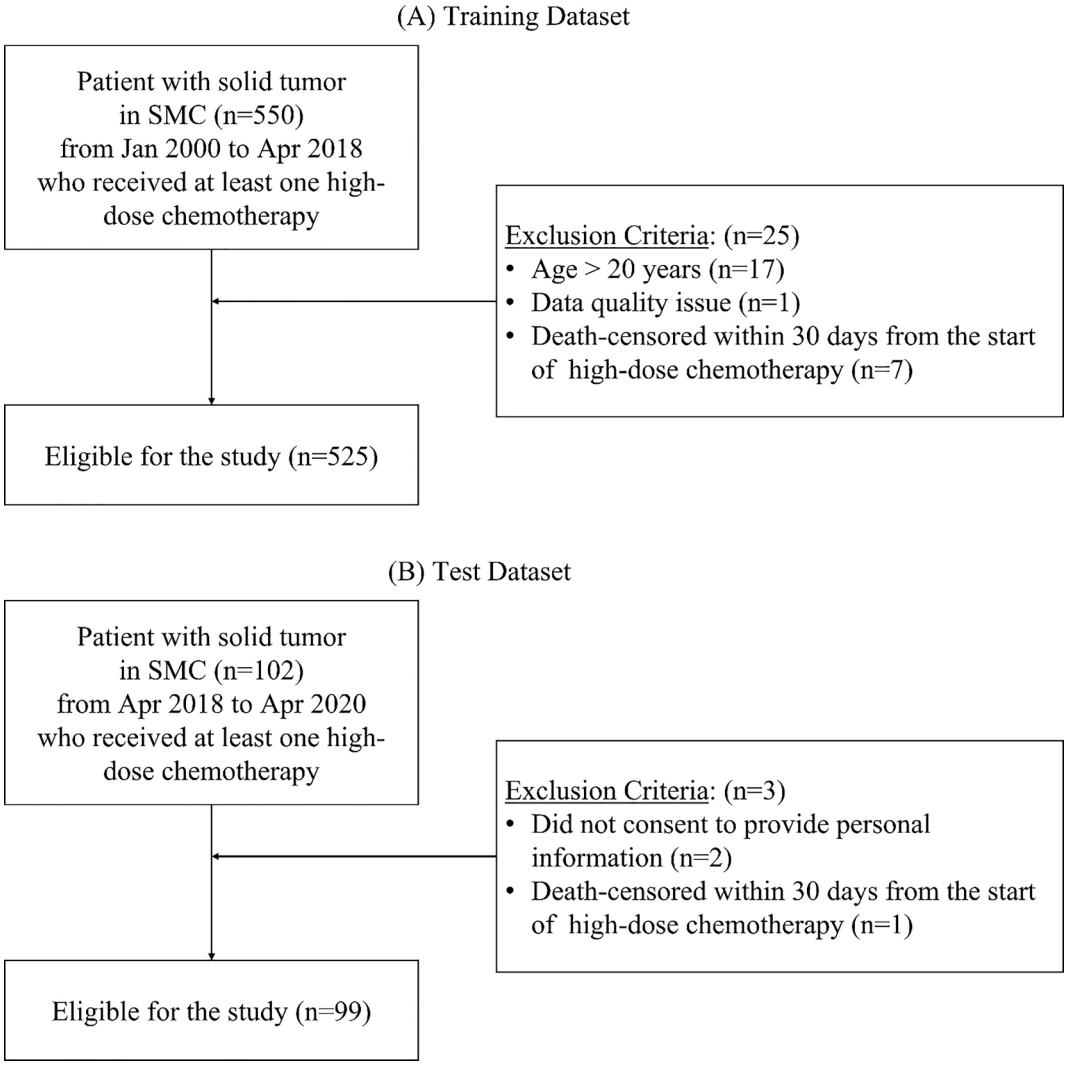

Fig. 2. Inclusion and exclusion criteria (A) Training dataset (B) Test dataset.

Data from both the conventional chemotherapy period (before high-dose chemotherapy) and the high-dose chemotherapy period were collected. During those periods, we collected laboratory test results, records of chemotherapy cycles with the initiation date, the actual dose of administered chemotherapeutic agents, prescription of granulocyte colony-stimulating factor (G-CSF), records of platelet transfusion, and transplanted CD34 + cell count for autologous stem cell transplantation after high-dose chemotherapy. The body weight and height were also measured to calculate the dose of each chemotherapeutic agent in the protocols.

2.4. Data preprocessing

For all the data preprocessing processes, please refer to the Data preprocessing section of the supplementary material. Here, we describe only the unique process of dealing with the various chemotherapeutic agents included in our model.

Patients received heterogeneous chemotherapy regimens with variable reduction rates for both conventional and high-dose chemotherapy. Therefore, we created a new variable called "intensity" by multiplying the ratio of the actual dose to the planned dose for each chemotherapeutic agent. There is no consistent reference to exactly how many times more toxic high-dose chemotherapy is compared to conventional chemotherapy. Thus, the value was determined by the experiment. We repeatedly experimented with a range of numbers to find the value for the best performance. Consequently, we found that the model has maximum performance when the value is 3. We compared the performances with and without this preprocessing method.

2.5. Model selection and description

Among the time series forecasting algorithms, we chose Google's temporal fusion transformer (TFT) [17] to predict the change in the ANC value within 30 days of the high-dose chemotherapy cycle. This model took the entire history including all the variables, which we mentioned in Section 2.3., of conventional chemotherapy for each patient as input. Additionally, the planned high-dose chemotherapy schedule and the chemotherapeutic agents were provided as input. After the model provided a 30-day prediction, we calculated when the neutropenia$_{500}$ ended from these values. We randomly selected 50 out of a total of 525 patients and placed them aside as holdout data, and used this as the internal validation dataset. The remaining data from 475 patients were used as the training dataset. We used the symmetric mean absolute percentage error (sMAPE) loss function for ANC forecasting.

$$sMAPE = \sum \frac{2^{*}|y_{true} - y_{pred}|}{|y_{true}| + |y_{pred}|}$$

A detailed explanation of the TFT and the comparison with other candidate algorithms are shown in the Model development section of the supplementary material.

2.6. Evaluation metric for model performance

To evaluate the accuracy of the model for the prediction of the day of neutrophil count recovery, the proportion that exactly matched the actual value was calculated. Additionally, proportions that matched the actual values within one day and two days were calculated, respectively. We visually represented the difference between the actual end date of neutropenia$_{500}$ and the model predicted end date of neutropenia$_{500}$, in days. We assumed that showing prediction errors in days would be more intuitive for the clinicians.

2.7. Validation of the model by comparison with the clinicians' performance

We enrolled 10 clinicians willing to participate in the comparison study through an intranet announcement at Samsung Medical Center. All participants were anonymized as R*1* to R*10* and grouped into resident and specialist groups (pediatricians undergoing hematology-oncology fellowship or practicing as a pediatric hematology-oncologist, with at least two to five sample sizes per group).

For each case from the test dataset, each participant reviewed the data (patient conditions) and predicted the patient's recovery day from neutropenia. After predicting each case, the model-predicted recovery day was revealed to the participants; they were again asked if they wanted to change the estimated day of recovery, considering the model's prediction. The evaluation was performed considering the fatigue and concentration of the participants. We only included the results of participants who evaluated all the cases.

The results of the clinicians' prediction task were statistically analyzed for differences between the resident group, the specialist group, and the model with the best performance. The results of change in prediction after the prediction of the model was revealed to clinicians were also statistically analyzed for any trends in the number or magnitude of changes.

2.8. Development and survey of the questionnaire

We developed a new questionnaire based on the technology acceptance model (TAM) [18], and the details are described in the Development of the questionnaire section of the supplementary material. All 10 participating clinicians answered the questionnaire developed herein to evaluate the difference between individuals in terms of personal characteristics and their beliefs regarding the data and the model.

2.9. Statistical analysis

We compared the predicted values of the model and human experts. The pairwise proportion test was used to compare groups, and Levene's test [19] was used to check groups for the assumption of homogeneity of variance of the residuals. Bonferroni's correction was used for multiple comparisons in analyses between pairwise groups. Wilcoxon's rank sum test was used to compare the degree of change in prediction between the resident and specialist groups. Cronbach's alpha test was used to ensure that the questionnaire items were appropriately correlated. Items with Cronbach's alpha ≤ 0.7 were considered as having a poor correlation between factors.

In all analyses, the significance level was set to $p < 0.05$. All continuous variables are represented by median (interquartile range [IQR]), and all nominal variables are represented by n (%). All statistical analyses were carried out using R package, version 4.0.3 (The R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org).

3. Results

3.1. Data statistics

For the two study periods, data from 525 patients (2000–2018 period) for the training dataset and 99 patients (2019–2020 period) for the test dataset were collected. Overall, there were 51 variables, including blood cell count with differential and chemistry panels, name of the chemotherapeutic agents, dose of the administered chemotherapeutic agents, platelet transfusion records, and G-CSF injection records. Detailed information about each variable is described in supplementary table 1.

The median number of time points per patient was 102 (range 18–244) for the training data set and 101 (range 28–237) for the test

data set. These time points were collected from the EHR. Each chemotherapy cycle contained a median of 15 time points; 31.3 % of the total time points were distributed between days 0 and 5 of each chemotherapy cycle. Of the time points, 41.4 % were distributed between days 11 and 20.

Patient demographics are summarized in Table 1. There was no statistically significant difference between the training dataset and the test dataset, except for the tumor type. There were more patients diagnosed with brain tumors in the test data set compared to the training dataset (34.5 % vs 63 %, $p < 0.001$).

3.2. Model performance

Table 2 shows the performance of the best model as the difference in the actual end date of neutropenia$_{500}$ from the model-inferred end date of neutropenia$_{500}$ in days. For the test dataset, the model exactly predicted the end day of neutropenia$_{500}$ for 24 of 99 test data (24.24 %). The errors within 1 day and 2 days were 76.76 % (76/99) and 94.94 % (94/99), respectively.

Fig. 3 shows the performance of the model with the test dataset using one chemotherapeutic agent as a single variable, without using our proposed method to process data of multiple chemotherapeutic agents. All other conditions were the same in both models, except that the number of variables was increased to 81 in the model that treated each chemotherapeutic agent as a different variable. The model that processed chemotherapeutic agents as a single variable performed better.

In both subplots, the x-axis is the difference between the actual recovery day and the model-inferred recovery day, in days. (a) result of the model when data of various chemotherapeutic agents were integrated as a single value. (b) result of the model when different chemotherapeutic agents were treated as different variables.

3.3. Validation of the model through comparison with clinicians' prediction result

Table 2 shows the result of comparing the predictive performance of the specialist and resident groups. The specialist group showed a better prediction performance than the resident group when the prediction error was allowed for up to 1 day (58.59 % vs 32.33 %, $p = 0.001$) and 2 days (82.33 % vs 50.51 %, $p < 0.001$).

Table 1
Demographic information.

Variable	Training dataset 2000–2018 (n = 525)	Test dataset 2019–2020 (n = 99)	*p*-value (method)
Sex			
Female	**306 (58 %)**	**48 (48 %)**	**0.2551**
Male	**219 (41 %)**	**51 (52 %)**	
			(chi-squared test)
Age, years			
Median	**4**	**4**	**0.2101**
Range	**0–20**	**0–20**	
			(Mann–Whitney test)
Tumor type			
Brain tumor	**181 (34.5 %)**	**63 (63 %)**	**< 0.001**
Neuroblastoma	**226 (43 %)**	**27 (27 %)**	
Other	**118 (22.5 %)**	**12 (12 %)**	**(chi-squared test)**
Follow-up period (days)*			
Median	**225**	**234**	**0.1014**
Range	**27–472**	**52–953**	
			(Mann–Whitney test)
Conventional chemotherapy (the number of cycles)			
Median	**6**	**6**	**0.6917**
Range	**1–16**	**2–15**	
			(Mann–Whitney test)

* From the first day of the initial conventional chemotherapy to the last day of the final conventional chemotherapy.

Table 2
Prediction performance of our model and clinicians with inter-group comparison.

	Model	Overall (n = 10)	Resident group (n = 6)	Specialist group (n = 4)	Resident vs Specialist group
Allowed error	**Accuracy (%)**	**Median (IQR)**	**Median (IQR)**	**Median (IQR)**	***P*-value**
Exact day (No error allowed)	**24.24**	**16. (12.75, 21.5)**	**8.59 (1.01, 16.41)**	**20.20 (15.66, 23.99)**	**0.097**
Within one day of error	**76.7**	**55.5 (49.5, 59.0)**	**32.33 (8.08, 55.81)**	**58.59 (55.31, 68.69)**	**0.001**
Within two days of error	**94.94**	**79.5 (75.75, 82.75)**	**50.51 (17.42, 80.05)**	**82.33 (79.55, 87.38)**	**<0.001**

Data are median (IQR) or n (%). Performance is measured using accuracy with different levels of error.

Fig. 4 demonstrates the comparison results between clinicians and the proposed model. When an error was allowed up to 1 or 2 days, the model showed a better predictive ability than the specialist group. The resident group showed a lower predictive ability than the specialist group and the models for all error tolerances.

In Fig. 4b, the resident group has a much wider dispersion of errors than the model and specialist groups. Similarly, errors in each group for each case are shown in Fig. 4c, and the specialist group and the model show similar oscillation trends, but the resident group oscillates over a much wider range.

Statistical comparison of the percentage of correct answers according to error tolerance between each group. The three groups were compared using the pairwise proportion test. In the analysis between pairs, Bonferroni corrections were made for multiple comparisons. $p \geq 0{\cdot}05$; $^{*}p < 0{\cdot}05$; $^{**}p < 0{\cdot}01$; $^{***}p < 0{\cdot}001$ (b) The difference between the predicted value and the correct answer as a violin plot, in each group. Levene's test was used to statistically test the variance difference of the group. $p \geq 0{\cdot}05$; $^{*}p < 0{\cdot}05$; $^{**}p < 0{\cdot}01$; $^{***}p < 0{\cdot}001$ (c) The difference between the predicted value and the ground truth for each case, in each group. The size of the dot is increased by the number of clinicians who made the same prediction. The dots in light gray indicate within 1 day, and the dots in dark gray indicate within 2 days.

3.4. Clinicians' change in prediction after looking at the model's prediction

In total, 80 % of the clinicians changed their predictions at least once after the model's prediction was revealed, regardless of whether they belonged to the specialist or resident groups. There were a total of 106 prediction changes. Most of the change in prediction (95/106) was made in the direction of getting closer to the model's prediction, but some changes (11/106) were made in the direction away from the model's prediction. Among the changes made in the direction of the model's prediction, the prediction error decreased in 86/95 cases, but the error increased in 9/95 cases. All 11 changes in the prediction made further away from the model's prediction resulted in further deviation from the correct answer.

3.5. Questionnaire survey result analysis

Fig. 5 summarizes each question in the questionnaire and the scores of the responses on the 5-point Likert scale for the resident and specialist groups (1 = not at all agree, 5 = totally agree). The interpretation of the low and high Likert scores differed depending on the question. High

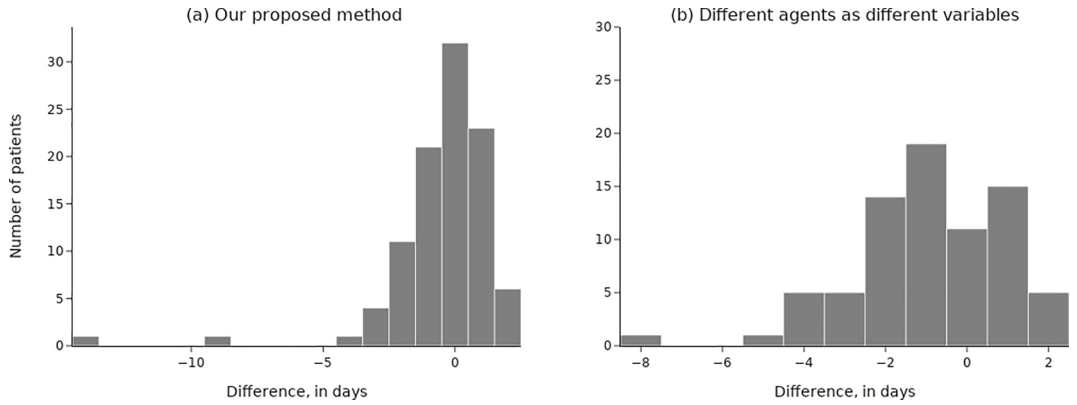

Fig. 3. Effect of the proposed chemotherapeutic agent data handling method.

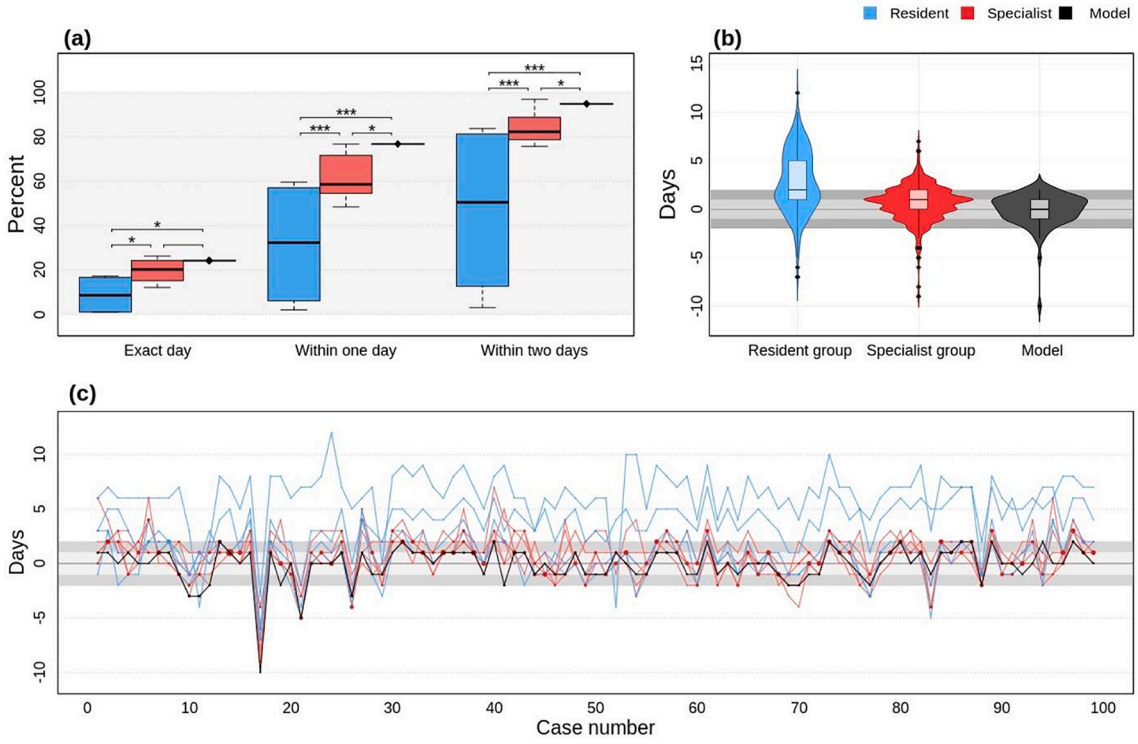

Fig. 4. Comparison of the predicted values and answers for the model and human expert.

average Likert scale scores were seen on questions such as whether to use a predictive model in a clinical setting, place more importance on subjective thinking at work, or change predictions when the model's performance is better, indicating an above-average level of agreement. Conversely, questions such as whether the introduction of the deep learning model would undermine clinicians' social values, undermine autonomy, or threaten patients' privacy showed low average Likert scores, which means that clinicians disagree with these questions. In comparing the results between the specialist and resident groups, there were slight differences in predictions on privacy infringement, but there was no significant statistical difference overall.

4. Discussion

This work is a novel attempt to use deep learning techniques to predict a patient's neutrophil recovery day by utilizing the patient's daily monitoring data from EHR. To the best of our knowledge, this is the first study to establish a model that can predict the personalized recovery day for each patient. There have been several machine learning approach studies concerned with mortality prediction [20], prediction of the onset of leucopenia [21,22], or modeling myelosuppression for a specific single chemotherapeutic agent [11,23]; however, there are no studies that directly compared the model with clinicians. In this study, we tried to show the efficacy of the model as compared to clinicians, rather than just showing the performance of the model. The model performed similarly or slightly better than the specialist group and showed much higher accuracy than the resident group, which showed large variability in prediction. Although the performance of our model may seem low compared to that observed in previous studies, including comparisons with clinicians such as reading fundus photographs [24], chest radiographs [25], or acute kidney injury prediction [26], it should be considered that most of these studies were predictions for binary outcomes with non-time series data. In contrast, our model entailed predictions for non-binary outcomes with time-series data, which are much more difficult. Further, our model was designed to respond when various chemotherapeutic agents were used simultaneously, and it was

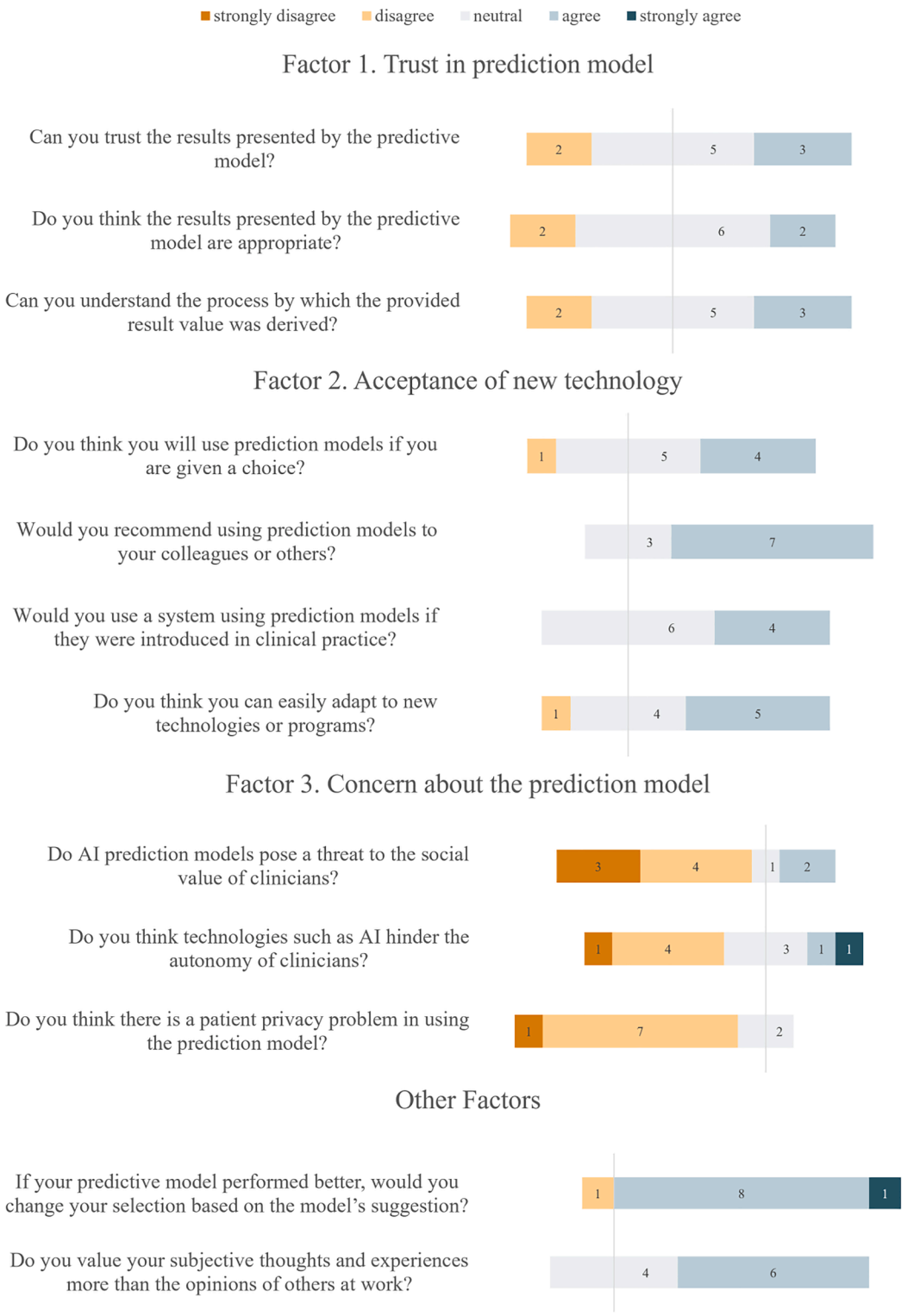

Fig. 5. The 5-point Likert scale responses of groups of specialists and residents for each factor in the questionnaire (1 = not at all agree, 5 = totally agree). The number in each bar is the number of respondents.

also confirmed that our proposed method showed better performance than the conventional method.

Another point of interest was whether clinicians would accept the model's prediction. There are several previous studies [27,28] on this topic, and two claim that clinicians accepted the predictions of the deep learning model. As in a previous study [29], we developed a new questionnaire based on TAM and analyzed the results of the questionnaire and the results of changes in the predictions of clinicians complementary to each other. In our study, 80 % of clinicians changed their prediction at least once, and 40 % changed their prediction more than 5 % of the total selection. This implies that the deep learning model can influence the decision-making ability of clinicians when the task of prediction is difficult. When comparing the specialist and resident groups, the specialist group changed their choices less frequently, and

even if they did, the difference from the original value was smaller.

Further, we observed an interesting attitude toward accepting the model's prediction. There were 11 cases where some participants changed their predictions away from the model's prediction, resulting in the worsening of the original predictions. These were found in two participants, and they tended to respond negatively to "Trust in the prediction models" from the survey designed based on TAM. Several studies [30,31] have reported cases where a model's incorrect prediction was followed, and inaccurate results were obtained; in our study, we further identified cases where the model's correct prediction was completely ignored, and the opposite action resulted in inaccurate outputs.

The limitations of this study are as follows. First, this was a single-center retrospective study. Second, unlike adults, the incidence of pediatric solid tumors is low, and the indication for high-dose chemotherapy is limited to high-risk or relapsed patients with solid tumors. Third, we have confirmed the possibility of accepting the model's predictions but not the acceptance in real clinical settings. Finally, since a survey was used rather than an interview, the subjectivity of the respondents is reflected a lot, and an in-depth analysis of the reasons for changing predictions was impossible. Despite the above limitations, our model has similar predictive power to that of the best-predicting clinicians in our study; hence, our model is considered acceptable.

5. Conclusions

In this study, we used deep learning techniques to predict the time to recovery from neutropenia at the individual patient level and compared the predictive ability of the model with that of human experts. Additionally, we demonstrated that clinicians accept the results of the prediction model well and that it positively affects the clinicians' opinion. As a result, we found that the deep learning model shows a better predictive ability than human experts and can help reduce the prediction errors of clinicians.

6. Statement of significance

Problem or Issue	Can deep learning models predict when patients will recover from severe neutropenia after high-dose chemotherapy, and can the model's prediction change clinicians' predictions?
What is Already Known	Several attempts have been made to predict an individual's ANC recovery; however, they have often been limited for certain drugs and lack generalized predictions [1]
What this Paper Adds	A deep learning model showed similar or better predictive performance than clinicians. Eighty percent of clinicians changed their initial predictions at least once after the model's prediction. Deep learning models can predict when patients will recover from severe neutropenia, and clinicians showed high acceptability of the model

CRediT authorship contribution statement

Hyunwoo Choo: Methodology, Software, Validation, Writing – original draft, Visualization. **Su Young Yoo:** Methodology, Validation, Formal analysis. **Suhyeon Moon:** Formal analysis. **Minsu Park:** Formal analysis. **Jiwon Lee:** Resources. **Ki Woong Sung:** Resources. **Won Chul Cha:** Resources. **Soo-Yong Shin:** Supervision, Writing – review & editing. **Meong Hi Son:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: MHS, SYS, and HC have a pending patent application on some of the material reported in this manuscript. The remaining authors declare no competing interests.

Acknowledgments

We would like to thank MKM and PMJ for their assistance with the data collection and review.

Role of the funding source

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 1711114031) and by Samsung Medical Center Grant #SMO1210431.

Authorship contributions

Contribution: HC performed experiments, drafted the manuscript, and created the figures; HC and MHS analyzed the results; JL and KWS contributed to the clinical data and implementation of the results; SM and MP contributed to statistical analyses and interpretation of the data and results; SYY and WCC developed the questionnaire, conducted surveys, and analyzed the results; SYS and MHS designed the research and reviewed the paper. All authors discussed the results and commented on the manuscript.

Data availability

To protect the privacy of patients, the institutional review board of the Samsung Medical Center did not approve data publishing. The dataset will be available upon request after approval from the institutional review board and the data review board of the Samsung Medical Center.

Code availability

We used python 3.7, pandas 1.2.0 for data preparation, pytorch 1.7, pytorch lightning 1.1.5, and pytorch forecasting 0.8.3 for model building; optuna 2.4 for optimizing the model parameters; and plotly 4.14 for visualization. The code is available at https://github.com/bmi_skku_edu/nadir_prediction.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2022.104268.

References

[1] J. Crawford, D.C. Dale, G.H. Lyman, Chemotherapy-induced neutropenia: risks, consequences, and new directions for its management, Cancer 100 (2004) 228–237, https://doi.org/10.1002/cncr.11882.
[2] Common terminology criteria for adverse events (CTCAE), Cancer.gov. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/docs/CTCAE_v5_Quick_Reference_8.5x11.pdf, 2021 (accessed 7 May 2021).
[3] T. Lehrnbecher, Treatment of fever in neutropenia in pediatric oncology patients, Curr. Opin. Pediatr. 31 (2019) 35–40, https://doi.org/10.1097/MOP.0000000000000708.
[4] L.E. Friberg, A. Henningsson, H. Maas, L. Nguyen, M.O. Karlsson, Model of chemotherapy-induced myelosuppression with parameter consistency across drugs, J. Clin. Oncol. 20 (2002) 4713–4721, https://doi.org/10.1200/JCO.2002.02.140.
[5] M. Craig, A.R. Humphries, M.C. Mackey, A Mathematical model of granulopoiesis incorporating the negative feedback dynamics and kinetics of G-CSF/neutrophil binding and internalization, Bull. Math. Biol. 78 (2016) 2304–2357, https://doi.org/10.1007/s11538-016-0179-8.
[6] S. Karppinen, O. Lohi, M. Vihola, Prediction of leukocyte counts during paediatric acute lymphoblastic leukemia maintenance therapy, Sci. Rep. 9 (2019) 18076, https://doi.org/10.1038/s41598-019-54492-5.
[7] R. Khosravan, S.G. DuBois, K. Janeway, E. Wang, Extrapolation of pharmacokinetics and pharmacodynamics of sunitinib in children with gastrointestinal stromal tumors, Cancer Chemother. Pharmacol. 87 (2021) 621–634, https://doi.org/10.1007/s00280-020-04221-x.
[8] R.B. Mokhtari, T.S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das, H. Yeger, Combination therapy in combating cancer, Oncotarget. 8 (23) (2017) 38022–38043, https://doi.org/10.18632/oncotarget.16723.

[9] H. Gurney, How to calculate the dose of chemotherapy, Br. J. Cancer. 86 (2002) 1297–1302, https://doi.org/10.1038/sj.bjc.6600139.
[10] I. Netterberg, E.I. Nielsen, L.E. Friberg, M.O. Karlsson, Model-based prediction of myelosuppression and recovery based on frequent neutrophil monitoring, Cancer Chemother. Pharmacol. 80 (2017) 343–353, https://doi.org/10.1007/s00280-017-3366-x.
[11] V. Cuplov, N. André, Machine learning approach to forecast chemotherapy-induced haematological toxicities in patients with rhabdomyosarcoma, Cancers (Basel). 12 (2020) 1944, https://doi.org/10.3390/cancers12071944.
[12] P. Nieboer, E.G.E. de Vries, E. Vellenga, W.T.A. van der Graaf, N.H. Mulder, W. J. Sluiter, J.T.M. de Wolf, Factors influencing haematological recovery following high-dose chemotherapy and peripheral stem-cell transplantation for haematological malignancies; 1-year analysis, Eur. J. Cancer. 40 (8) (2004) 1199–1207, https://doi.org/10.1016/j.ejca.2004.01.029.
[13] M.H. Son, D.H. Kim, S.H. Lee, K.H. Yoo, K.W. Sung, H.H. Koo, J.Y. Kim, E.J. Cho, E. S. Kang, D.W. Kim, Hematologic recovery after tandem high-dose chemotherapy and autologous stem cell transplantation in children with high-risk solid tumors, J. Korean Med. Sci. 28 (2) (2013), 220, https://doi.org/10.3346/jkms.2013.28.2.220.
[14] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.
[15] B. Norgeot, G. Quer, B.K. Beaulieu-Jones, A. Torkamani, R. Dias, M. Gianfrancesco, R. Arnaout, I.S. Kohane, S. Saria, E. Topol, Z. Obermeyer, B. Yu, A.J. Butte, Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist, Nat. Med. 26 (9) (2020) 1320–1324, https://doi.org/10.1038/s41591-020-1041-y.
[16] X. Liu, S. Cruz Rivera, D. Moher, M.J. Calvert, A.K. Denniston, A.-W. Chan, A. Darzi, C. Holmes, C. Yau, H. Ashrafian, J.J. Deeks, L. Ferrante di Ruffano, L. Faes, P.A. Keane, S.J. Vollmer, A.Y. Lee, A. Jonas, A. Esteva, A.L. Beam, A.-W. Chan, M.B. Panico, C.S. Lee, C. Haug, C.J. Kelly, C. Yau, C. Mulrow, C. Espinoza, J. Fletcher, D. Paltoo, E. Manna, G. Price, G.S. Collins, H. Harvey, J. Matcham, J. Monteiro, M.K. ElZarrad, L. Ferrante di Ruffano, L. Oakden-Rayner, M. McCradden, P.A. Keane, R. Savage, R. Golub, R. Sarkar, S. Rowley, Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension, Nat. Med. 26 (9) (2020) 1364–1374, https://doi.org/10.1038/s41591-020-1034-x.
[17] B. Lim, S.Ö. Arık, N. Loeff, T. Pfister, Temporal fusion transformers for interpretable multi-horizon time series forecasting, Int. J. Forecast. 37 (2021) 1748–1764, https://doi.org/10.1016/j.ijforecast.2021.03.012.
[18] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Q. 13 (1989) 319, https://doi.org/10.2307/249008.
[19] H. Levene, Robust Tests for equality of variances in contributions to probability and statistics, in: I. Olkins (Ed.), Stanford, Stanford University Press, CA, 1960, pp. 278–292.
[20] B.J. Cho, K.M. Kim, S.E. Bilegsaikhan, Y.J. Suh, Machine learning improves the prediction of febrile neutropenia in Korean inpatients undergoing chemotherapy for breast cancer, Sci. Rep. 10 (2020) 14803, https://doi.org/10.1038/s41598-020-71927-6.
[21] C. Zhu, S.H. Lin, X. Jiang, Y. Xiang, Z. Belal, G. Jun, R. Mohan, A novel deep learning model using dosimetric and clinical information for grade 4 radiotherapy-induced lymphopenia prediction, Phys. Med. Biol. 65 (3) (2020) 035014, https://doi.org/10.1088/1361-6560/ab63b6.
[22] S.M. Naushad, P. Dorababu, Y. Rupasree, A. Pavani, D. Raghunadharao, T. Hussain, S.A. Alrokayan, V.K. Kutala, Classification and regression tree-based prediction of 6-mercaptopurine-induced leucopenia grades in children with acute lymphoblastic leukemia, Cancer Chemother. Pharmacol. 83 (5) (2019) 875–880, https://doi.org/10.1007/s00280-019-03803-8.
[23] T. Shibahara, S. Ikuta, Y. Muragaki, Machine-learning approach for modeling myelosuppression attributed to nimustine hydrochloride, JCO Clin. Cancer Inform. 2 (2018) 1–21, https://doi.org/10.1200/CCI.17.00022.
[24] D. Lin, et al., Application of comprehensive artificial intelligence retinal expert (CARE) system: A national real-world evidence study, Lancet Digit. Health. 3 (2021) e486–e495, https://doi.org/10.1016/S2589-7500(21)00086-8.
[25] J.T. Wu, K.C.L. Wong, Y. Gur, N. Ansari, A. Karargyris, A. Sharma, M. Morris, B. Saboury, H. Ahmad, O. Boyko, A. Syed, A. Jadhav, H. Wang, A. Pillai, S. Kashyap, M. Moradi, T. Syeda-Mahmood, Comparison of Chest Radiograph Interpretations by Artificial Intelligence Algorithm vs Radiology Residents, JAMA Netw. Open. 3 (10) (2020), e2022779, https://doi.org/10.1001/jamanetworkopen.2020.22779.
[26] N. Rank, B. Pfahringer, J. Kempfert, C. Stamm, T. Kühne, F. Schoenrath, V. Falk, C. Eickhoff, A. Meyer, Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance, NPJ Digit. Med. 3 (1) (2020), https://doi.org/10.1038/s41746-020-00346-8.
[27] S. Gaube, H. Suresh, M. Raue, A. Merritt, S.J. Berkowitz, E. Lermer, J.F. Coughlin, J.V. Guttag, E. Colak, M. Ghassemi, Do as AI say: susceptibility in deployment of clinical decision-aids, NPJ Digit. Med. 4 (1) (2021), https://doi.org/10.1038/s41746-021-00385-9.
[28] INFANT Collaborative Group, Computerised interpretation of fetal heart rate during labor (INFANT): a randomised controlled trial, Lancet. 389 (2017) 1719–1729, https://doi.org/10.1016/S0140-6736(17)30568-8.
[29] S. Jauk, D. Kramer, A. Avian, A. Berghold, W. Leodolter, S. Schulz, Technology acceptance of a machine learning algorithm predicting delirium in a clinical setting: a mixed-methods study, J. Med. Syst. 45 (4) (2021), https://doi.org/10.1007/s10916-021-01727-6.
[30] M. Jacobs, et al., How machine-learning recommendations influence clinician treatment selections: the example of the antidepressant selection, Transl. Psychiatry. 11 (2021) 108, https://doi.org/10.1038/s41398-021-01224-x.
[31] C.J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, D. King, Key challenges for delivering clinical impact with artificial intelligence, BMC Med. 17 (2019) 195, https://doi.org/10.1186/s12916-019-1426-2.