Original Paper

A Smartphone-Based Decision Support Tool for Predicting Patients at Risk of Chemotherapy-Induced Nausea and Vomiting: Retrospective Study on App Development Using Decision Tree Induction

Abu Saleh Mohammad Mosa^{1,2,3,4}, MS, PhD; Md Kamruz Zaman Rana^{1,2}, MS; Humayera Islam^{2,4,5}, MS; A K M Mosharraf Hossain^{6,7}, MD; Illhoi Yoo^{1†}, PhD

⁴Center for Biomedical Informatics, University of Missouri School of Medicine, Columbia, MO, United States

⁷Department of Hematology and Medical Oncology, BayCare Health System, South Florida Baptist Hospital, Plant City, FL, United States [†]deceased

Corresponding Author:

Abu Saleh Mohammad Mosa, MS, PhD Health Management and Informatics University of Missouri School of Medicine 1 Hospital Drive Columbia, MO, 65212 United States Phone: 1 573 882 8349 Email: mosaa@health.missouri.edu

Abstract

Background: Chemotherapy-induced nausea and vomiting (CINV) are the two most frightful and unpleasant side effects of chemotherapy. CINV is accountable for poor treatment outcomes, treatment failure, or even death. It can affect patients' overall quality of life, leading to many social, economic, and clinical consequences.

Objective: This study compared the performances of different data mining models for predicting the risk of CINV among the patients and developed a smartphone app for clinical decision support to recommend the risk of CINV at the point of care.

Methods: Data were collected by retrospective record review from the electronic medical records used at the University of Missouri Ellis Fischel Cancer Center. Patients who received chemotherapy and standard antiemetics at the oncology outpatient service from June 1, 2010, to July 31, 2012, were included in the study. There were six independent data sets of patients based on emetogenicity (low, moderate, and high) and two phases of CINV (acute and delayed). A total of 14 risk factors of CINV were chosen for data mining. For our study, we used five popular data mining algorithms: (1) naive Bayes algorithm, (2) logistic regression classifier, (3) neural network, (4) support vector machine (using sequential minimal optimization), and (5) decision tree. Performance measures, such as accuracy, sensitivity, and specificity with 10-fold cross-validation, were used for model comparisons. A smartphone app called CINV Risk Prediction Application was developed using the ResearchKit in iOS utilizing the decision tree algorithm, which conforms to the criteria of explainable, usable, and actionable artificial intelligence. The app was created using both the bulk questionnaire approach and the adaptive approach.

Results: The decision tree performed well in both phases of high emetogenic chemotherapies, with a significant margin compared to the other algorithms. The accuracy measure for the six patient groups ranged from 79.3% to 94.8%. The app was developed using the results from the decision tree because of its consistent performance and simple, explainable nature. The bulk questionnaire approach asks 14 questions in the smartphone app, while the adaptive approach can determine questions based on the previous questions' answers. The adaptive approach saves time and can be beneficial when used at the point of care.

¹Health Management and Informatics, University of Missouri School of Medicine, Columbia, MO, United States

²Institute for Data Science and Informatics, University of Missouri, Columbia, MO, United States

³Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, United States

⁵Institute of Statistical Research and Training, University of Dhaka, Dhaka, Bangladesh

⁶Ellis Fischel Cancer Center, University of Missouri School of Medicine, Columbia, MO, United States

Conclusions: This study solved a real clinical problem, and the solution can be used for personalized and precise evidence-based CINV management, leading to a better life quality for patients and reduced health care costs.

(JMIR Mhealth Uhealth 2021;9(12):e27024) doi: 10.2196/27024

KEYWORDS

chemotherapy; CINV risk factors; data mining; prediction; decision trees; clinical decision support; smartphone app

Introduction

Background

Chemotherapy is a drug treatment commonly used to treat nearly every type of cancer [1]. As estimated, each year, as many as 1 million Americans receive some type of chemotherapy [2]. Cancer cells multiply at an unusually faster rate compared to healthy cells, and chemotherapy is used to kill those fast-growing cells in the body. However, chemotherapy can lead to many side effects, such as nausea, vomiting, appetite changes, anemia, hair loss, constipation, and diarrhea, among others [3-11]. Chemotherapy-induced nausea and vomiting (CINV) are the two most frightful and unpleasant side effects of chemotherapy [3,4,12-15].

CINV can lead to consequences that affect both patients and the health care system as a whole. First, CINV engenders other side effects, such as nutritional deficits, dehydration, and electrolyte imbalance, which diminishes the quality of life in cancer patients [16-20]. Second, the various side effects of CINV lead to a low-quality social life [19,21]. Third, CINV can also lead to loss of workdays, which in return increases the economic burden [19,22-24]. Fourth, CINV surges health care costs arising from CINV-related outpatient visits, hospitalization, and the cost of drugs [18,19,22-27]. Fifth, intolerance of cancer patients toward CINV can lead to discontinuation of cancer treatment, leading to poor treatment outcomes, treatment failure, or even death [12,28-30].

The management of CINV is a complex process due to two factors. The first level of complexity arises from the different impacts of the different emetogenicity levels of the chemotherapeutic agents. The emetogenicity of chemotherapy is fractionated into four emetic risk categories based on the percentage of patients who suffer from CINV without antiemetics: (1) minimal (<10%), (2) low-emetogenic chemotherapy (LEC: 10%-30%), (3) moderate-emetogenic chemotherapy (MEC: 30%-90%), and (4) high-emetogenic chemotherapy (HEC: >90%). CINV has two different pathophysiological phases (acute and delayed) that can lead to different consequences, adding a second level of complexity. The acute phase of CINV occurs within the first 24 hours of chemotherapy. Chemotherapy triggers the release of serotonin in the peripheral pathway (gastrointestinal tract), which binds to the 5-hydroxytryptamine (5-HT₃) receptors and sends a signal to the vomiting center in the medulla [31,32]. The central pathway is associated with the delayed phase of CINV that occurs after the first 24 hours of chemotherapy administration and may persist up to 1 week. This pathway is located in the brain, where chemotherapy triggers a neuropeptide release

named substance P, which binds to the neurokinin-1 (NK-1) receptor in the vomiting center, causing CINV [31,32].

There are several antiemetic guidelines for the management of CINV, such as the American Society of Clinical Oncology (ASCO) guideline [33,34], the National Comprehensive Cancer Network (NCCN) guideline [35], and the guideline from the Multinational Association of Supportive Care in Cancer (MASCC) in cooperation with the European Society of Medical Oncology (ESMO) [36]. Despite the improvements in CINV management, many recent studies have reported various percentages of patients experiencing CINV with the use of antiemetics: 28% [37], 38%-52% [38], 56.1% [39], 61.2% [19], and 62% [20]. The guideline-recommended standard antiemetic prophylaxis takes only the chemotherapeutic emetogenicity into consideration for CINV management.

However, several patient-related risk factors can potentially worsen the risk of CINV, but none of the guidelines considers those factors [40]. Since physicians cannot entirely rely on the guidelines, they use their own experiences to manage CINV. Consequently, CINV management is inconsistent among physicians, since their decisions are subjective to their experiences in managing CINV [41].

The use of risk prediction algorithms for clinical decision making at the point of care would require completing and processing massive patient panels, which can be time consuming and can lead to inaccurate results [42]. In recent years, smartphones have become popular among physicians for accessing health care information at the point of care [43]. The advent of open-source frameworks, such as Apple ResearchKit, Apple CareKit, and Android frameworks (eg, PhoneGap), has opened up tremendous opportunities to capture patient-related data and deliver patient-specific clinical decision support information through smartphones. Data mining techniques are beneficial in predictive analytics on medical data [44]. Various machine learning (ML) algorithms have the potential to help build robust clinical decision support systems using clinical data. Smartphone apps integrated with robust clinical decision support developed from rigorously validated ML models and artificial intelligence (AI) can be immensely useful for clinicians and can significantly improve overall health care delivery.

Objective

The objective of this study was to develop a smartphone app for clinical decision support to predict patients' risk of CINV using patient-related risk factors. ML algorithms, such as the decision tree, naive Bayes algorithm, logistic regression classifier, neural network, and support vector machine, were applied to determine the best-performing algorithm for CINV risk prediction based on electronic medical records (EMRs). Standard performance metrics, such as accuracy, sensitivity,

XSL•FO RenderX

and specificity, were used to compare the performance among the algorithms. This paper also illustrates the use of the ML model to develop a smartphone app and demonstrates its usage from the users' perspective. The developed app aims to help clinicians identify high-CINV-risk patients and can be integrated with antiemetic guidelines for better CINV management.

Methods

Data Sources and Population Selection

This was a retrospective study, and data were collected from the EMRs from a single center called the University of Missouri Ellis Fischel Cancer Center. The study was approved by the MU Health Sciences institutional review board. Our study included only patients who received chemotherapy and standard antiemetic prophylaxis (based on national antiemetic guidelines) at the oncology outpatient service from June 1, 2010, to July 31, 2012. However, we excluded patients with missing information and those who underwent concurrent radiotherapy or surgical procedures.

We planned to collect two independent data sets for each stage of CINV. Since acute and delayed CINV follows two different pathophysiologies, we planned to discover the patient-related risk factors for causing CINV during both phases independently. In each data set, there were three groups based on the emetogenicity level of the chemotherapy regimens. Of the four emetic risk categories, the minimal risk category of chemotherapy for causing CINV is not clinically crucial, since only less than 10% of those patients suffer from CINV. Thus, we collected data in three separate groups corresponding to three clinically meaningful categories: low, moderate, and high.

Our significant interest classes included both CINV and non-CINV cases. However, LEC led to CINV in less than 30% of patients, and the use of standard antiemetic treatment further reduced this percentage. Thus, the data set had few CINV cases compared to non-CINV cases. In addition, the number of CINV cases was higher than the non-CINV cases in the HEC group. Hence, class balancing in each data group (LEC, MEC, and HEC) was considered necessary. We addressed the class imbalance issue by making the data set's size in each class for each group approximately equal.

Variable Selection

In a previous study, we completed a systematic review by following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline to identify potential patient-related variables that cause CINV [45]. Our previous study used MEDLINE to identify articles that demonstrated patient-related risk factors of CINV through clinical studies. A total of 26 patient-related risk factors were documented in that study from reviewing 49 articles [46]. For this study, we included 14 independent variables and 1 dependent variable (CINV outcome) [46]. We chose the risk factors based on the recommendations from chemotherapy experts in the MU Elis Fischel Cancer Center and our literature review. The selected variables were also easy to collect through clinical encounters, which can facilitate the usability of the prediction model at the point of care before chemotherapy.

Data Mining

Data mining or knowledge discovery in databases (KDD) can discover hidden patterns, previously unknown, and potentially useful information from data. In general, data mining algorithms are categorized into two groups: descriptive or unsupervised learning and predictive or supervised learning. In supervised learning, the class labels of the observations or tuples are known, whereas in unsupervised learning, those class labels are unknown. For this study, we developed a prediction model that falls into the supervised learning or classification category.

Classification is a supervised learning method for building classification models based on a data set (called training data) and the values in classifying attributes (called a class label). The classification model is used to predict the categorical class label. Classification is a two-step process in which the model is constructed in the first step and the accuracy of the model is determined using a data set (called test data set) in the second step. The accuracy of the classification model is the percentage of test data set tuples that are correctly classified by the model. To overcome the overfitting problem, the test data set must be independent of the training data set. In general, the classification model consists of IF-THEN rules or mathematical formulas. For our study, we used five popular data mining algorithms: (1) naive Bayes [47], (2) logistic regression classifier [48], (3) neural network (voted perceptron) [49], (4) support vector machine (using sequential minimal optimization) [50], and (e) decision tree [51-53]. There are several tools available for data mining. We used the most widely used tools, called WEKA [54]. Performance measures, such as accuracy, sensitivity, and specificity, were used for model comparisons. In addition, 10-fold cross-validation were used for model validation [55].

Smartphone App Development

ResearchKit is an open-source framework based on iOS that makes it easy to create mobile apps. It allows researchers and drug developers to tailor it to their own particular needs, whether for collecting clinical research data, recruiting patients, or obtaining informed consent. The framework allows for collecting information through electronic data capture, creating a small task to gather any specific information required for the study, and then storing the data as part of a sandbox, thereby protecting patient information. We developed our smartphone app using some modules, including a survey engine, visual consent flow, and active tasks from this framework. As the users of this app will be care providers, and no identifiable data will be stored, we did not use the visual consent flow. The smartphone app was built using the algorithm that had the most consistent performance among the ML algorithms and is also explainable, usable, and actionable AI for clinical decision support.

Results

Data Summary

In total, 6124 records were extracted based on inclusion and exclusion criteria. The number of records was 3053 and 3071 for the acute-phase and the delayed-phase data set, respectively. Table 1 presents the frequency distribution of both data sets for

XSL•FO RenderX

combinations of three chemotherapy categories and two treatment outcomes.

Table 1.	Data	summary
----------	------	---------

CINV ^a treatment group	Records, n	CINV, n (%)	No CINV, n (%)
Acute phase			
HECb ^b	1026	504 (49.12)	522 (50.88)
MEC ^c	1012	506 (50.00)	506 (50.00)
LEC ^d	1015	506 (50.15)	509 (49.85)
Total	3053	1519 (49.75)	1534 (50.25)
Delayed phase			
HEC	1166	586 (50.26)	580 (49.74)
MEC	891	447 (50.17)	444 (49.83)
LEC	1014	519 (51.18)	495 (48.82)
Total	3071	1552 (50.54)	1519 (49.46)

^aCINV: chemotherapy-induced nausea and vomiting.

^bHEC: high-emetogenic chemotherapy.

^cMEC: moderate-emetogenic chemotherapy.

^dLEC: low-emetogenic chemotherapy.

Data Mining Model Performance Comparison

The models' performances for all the emetogenicity levels and CINV phases (accuracy, sensitivity, specificity) are compared in Figure 1. The differences between performances of the different models were not consistent in each data set's model. The naive Bayes algorithm showed the best performance in the acute phase for LEC (the accuracy was 96.6%, sensitivity was

96.3%, and specificity was 96.8%), the acute phase for MEC (the accuracy was 90.8%, sensitivity was 89.3%, and specificity was 92.3%), and the delayed phase for MEC (the accuracy was 81.5%, sensitivity was 81.7%, and specificity was 81.3%). For the delayed phase for LEC, the support vector machine gave the best performance (the accuracy was 89.5%, sensitivity was 87.8%, and specificity was 91.3%).



Figure 1. Accuracy, sensitivity, and specificity of different ML algorithms used to predict CINV status among patients. CINV: chemotherapy-induced nausea and vomiting; HEC: high-emetogenic chemotherapy; LEC: low-emetogenic chemotherapy; MEC: moderate-emetogenic chemotherapy; ML, machine learning.



Neural Network 📕 Logistic Regression 📕 Support Vector Machines 📕 Naive Bayes 📃 Decision Tree

The decision tree gave the most consistent performance in both phases of HEC, with a significant margin compared to the other algorithms. Although different algorithms gave the best performance for different stages, we selected the decision tree model to develop the app for its consistent performance across measures and its simple, explainable nature. Moreover, clinical decision support integrated with explainable, usable, and actionable AI is more convenient for oncologists to understand, and thus, it can help them understand the app's background functioning.

Decision Tree Models

The six decision tree models for predicting CINV in both acute and delayed phases for each type of emetogenicity resulted in six flowcharts (Figures 2-7). Table 2 shows the description of the abbreviated form of each patient-related risk factor shown in the decision trees. We optimized the confidence factor for tree size and used the same confidence factor for all the decision trees. A threshold of >0 was used as the cutoff point. The accuracy of the six models was 94.8%, 88.5%, 90.2%, 79.3%, 88.7%, and 81%, respectively. In addition, sensitivity (correct prediction for the positive outcome of CINV) measures were 96.3%, 88.2%, 90.3%, 76.3%, 90.3%, and 85.2%, respectively,

```
https://mhealth.jmir.org/2021/12/e27024
```

while specificity (correct prediction for the negative outcome and 76.7%, respectively. of CINV) measures were 93.3%, 88.7%, 90.1%, 82.4%, 87.2%,

Figure 2. Decision tree. Phase: acute; emetogenicity: low. CINV: chemotherapy-induced nausea and vomiting.



Figure 3. Decision tree. Phase: delayed; emetogenicity: low. BMI: body mass index; CINV: chemotherapy-induced nausea and vomiting.



Figure 4. Decision tree. Phase: acute; emetogenicity: moderate. CINV: chemotherapy-induced nausea and vomiting.



XSL•FO RenderX

Figure 5. Decision tree. Phase: delayed; emetogenicity: moderate. CINV: chemotherapy-induced nausea and vomiting.



Figure 6. Decision tree. Phase: acute; emetogenicity: high. BMI: body mass index; CINV: chemotherapy-induced nausea and vomiting.



Figure 7. Decision tree. Phase: delayed; emetogenicity: high. CINV: chemotherapy-induced nausea and vomiting.



Table 2. Patient-related risk factors and their abbreviations used in the decision trees.

Risk factor abbreviation	Description
smoker	Is the patient a current smoker?
race	Race of the patient
age	Age of the patient in years
bmi	Body mass index during chemotherapy
anxiety	Did the patient have anxiety during chemotherapy?
prior_cinv	History of previous CINV ^a
n_prior_chemo	Number of prior chemotherapy regimen
n_comorbidities	Number of comorbidities
sex	Sex of the patient
alcohol	Alcohol consumption
stage	Stage of cancer
type	Type of cancer
dehydration	Did the patient have dehydration during chemotherapy?

^aCINV: chemotherapy-induced nausea and vomiting.

Clinical Decision Support Smartphone App

The clinical decision support smartphone app for CINV was developed using the output of the decision tree models obtained from the above analyses. The app was built on iOS and developed considering space usage and the possible variation of its users' technological skills. We created active tasks, depending on the flowcharts. In addition, the survey engine helped us to easily implement the questionnaire survey.

The app was created using two different approaches: (1) the bulk questionnaire approach and (2) the adaptive questionnaire

approach. In the bulk questionnaire approach, all 14 questions regarding CINV risk factors were asked one by one. After receiving the responses of the patients on all the questions, the predictive analyzer predicted the recommendations on both phases, depending on the six flowcharts obtained by applying the decision tree algorithm. In Figure 8, the flow for the bulk approach is shown. For a better experience, the clinician has the freedom to go back and change the input and recalculate the answer. An example of a set of answers is given in Figure 9. Depending on all the answers and using the six flowcharts' logic, the system selects the result for both the acute and delayed phases and displays it.

Figure 8. Flow diagram of CINV risk prediction smartphone app using the bulk questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.





Figure 9. Application GUI for the bulk questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.



The main limitation of the bulk questionnaire approach is that the physician at the point of care has to answer all the 14 questions to get to the final recommendation, even though not all questions are required for decision making for that patient. ResearchKit allows us to customize the questionnaire by adding features such as skipping questions or creating multiple paths, depending on the answer of the parent node of the decision trees. However, in this study, risk factors did not form a consistent hierarchy across the flowcharts, and thus skipping questions from a fixed questionnaire did not help. Moreover, some of the flowcharts had the same child under the parent node regardless of the answer, following different paths afterward. For instance, in Figure 5, the parent node is dehydration but the child node is anxiety regardless of whether dehydration is true or false. This motivated us to build a more time-energy-efficient approach called the adaptive questionnaire approach.

In an adaptive approach, the rule-based system first chooses a flowchart for the acute phase, depending on the emetogenicity level. A flowchart can have different paths, depending on the answers to the question as they come in the hierarchy of the decision tree. This approach follows a single path from the flowchart to generate a questionnaire for the clinician and saves all the answers in a database. Upon recommending the acute phase, the rule-based system chooses another flowchart for the delayed phase. This time, not all the questions in that flowchart are asked; instead, the app asks only the unanswered questions. There is a step generator feature at play for both acute and delayed phase prediction. The step generator determines the question paths for the patient, generates a new step if the question is unanswered, and use the answer from the saved answers for the already answered questions to generate the recommendation. In this approach, only the minimum questions needed to give a recommendation are included in the questionnaire, making the app more effective, faster, and user friendly. In Figure 10, the flow for the adaptive survey approach is shown.



Mosa et al

Figure 10. Flow diagram of CINV risk prediction smartphone app using the adaptive questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface; HEC: high-emetogenic chemotherapy; LEC: low-emetogenic chemotherapy; MEC: moderate-emetogenic chemotherapy.



For the adaptive approach, the app's questionnaire comes in dynamic format. The flow of the adaptive approach for a single path is illustrated in Figure 11. In this scenario, the user selected MEC as the emetogenicity of chemotherapy for the acute phase, and the model chose the decision tree for acute MEC shown in Figure 4. According to this flowchart, the first question was "whether the patient had anxiety during the chemotherapy," for which the user selected "no" as an answer. Following this answer, the next question was "the history of previous

chemotherapy." The user selected "yes," which led to the next question about "dehydration." Since the answer was "yes" for dehydration, the next question was about "smoking status." Only by asking these four questions, the system identified that the patient is at high risk of CINV. Although there are 14 risk factors, our dynamic approach only asks the questions that are necessary, choosing one pathway from the flowchart, which depends on the answers to the previous questions.



Figure 11. Application GUI for the adaptive questionnaire approach. CINV: chemotherapy-induced nausea and vomiting; GUI: graphical user interface.



The user started back again and now selected the delayed phase in the app for determining CINV risk. The system selected the flowchart from Figure 5 this time. One advantage of this adaptive approach is that it will not ask questions that have already been answered. For example, although in the delayed-phase flowchart, the first question was about dehydration, this was not asked, since this was already answered in the acute-phase mode. The question of anxiety was also skipped for the same reason. The third question in the delayed-phase flowchart was about "the number of prior chemotherapy regimens." Since this question was never asked, the system picked this question next and the user selected 3 as an answer. Thus, using the answers to these questions, the app generated the recommendation that the patient has a high risk of CINV in the delayed phase.

https://mhealth.jmir.org/2021/12/e27024

RenderX

Discussion

Principal Findings

CINV is a major side effect of chemotherapy among cancer patients. Appropriate examination of patient-specific risk factors before selecting premedications for CINV is critical in cancer care [56]. Better control of CINV has both short- and long-term effects in cancer care, leading to improved therapy tolerability, less anxiety, higher patient satisfaction, and avoidance of immediate discontinuation of the treatment [28,57-59]. Our previous study on finding risk factors through a systematic literature review shed light on the prevalent risk factors of CINV, as seen in the existing literature [46]. Patient-specific factors, such as smoking and alcohol status, sex, age, and the body mass index (BMI), can play a vital role in determining

their effect on CINV. This study used data mining to discover significant relationships among the patient-related risk factors that influence the occurrence of CINV. Six independent data sets (three chemotherapy groups and two phases of CINV in each chemotherapy group) were individually analyzed to build the best-possible prediction models for CINV prediction. The risk factors used for building the models can be easily collected at the point of care or are available in the hospital EMRs. Among the popular data mining algorithms used for our study, the decision tree model performed consistently across the measures for both CINV phases.

A rule-based app can be considered an appropriate choice for its simplicity in explaining the model to a clinician and implementing it in a software application. Thus, we developed a CINV smartphone app using the results from the decision tree model because of its consistent performance and simplicity. We implemented two approaches, bulk and adaptive, to develop the CINV risk prediction app using ResearchKit. If the questions could be generated from multiple flowcharts, designing a fixed-order questionnaire might not help build an efficient app. The question hierarchy was not consistent across different tree models. Instead of asking input to all variables, we developed an adaptive approach to present a minimal number of questions for computing the prediction. The fixed (bulk)-order approach will ask 14 questions for any of those 115 decision paths, but for the adaptive approach, the maximum questions asked will be equal to the depth of that flowchart (up to 9 questions). This makes the app both time and energy efficient for the user and can reduce the physicians' time at the point of care.

The developed smartphone app for recommending patients at risk of CINV can help improve the prevention of CINV among cancer patients. The target users (ie, clinicians) can use this app at the point of care during the prescription of antiemetics. This app will help identify patients at risk of CINV based on patient-related risk factors. Having this knowledge of the patients before the prescription of antiemetics can help design a better treatment plan, leading to better CINV management. Furthermore, the app took significantly less space and was developed considering the possible variation in users' technological skills. It does not require any permission, which will help users use it more effortlessly. The oncologist will have complete access to the risk calculation algorithm in their

smartphone, which will drastically reduce the amount of time required to help a large group of people and will have the flexibility to provide personalized care to every patient, improving their quality of life.

Limitations

In this study, the data were collected by retrospective record review. Prospective validation is needed to confirm the usefulness of the model in a real clinical setting. The research also shows that female patients with pregnancy-related nausea and vomiting have a higher risk of CINV. However, this information was missing from our data set. This information could considerably enhance the prediction results. The data have a lower representation of Asians and Hispanics. A multicenter or multinational study, including various populations, is needed to overcome this shortcoming. In addition, if we use EMR data to integrate with the app, there is no difference between the bulk and adaptive approaches. However, if the app is used as a prediction tool at the point of care, the adaptive approach is more time and energy efficient, thus decreasing the chances of wrong input answers. In addition, for hospitals without any EMR system, this app can be extremely beneficial for cancer patients.

Future Work

In the future, our plan is to deploy this app in point-of-care settings by integrating it into EMRs to predict the risk of CINV. We can also perform a clinical study for estimating outcomes and improvement. Currently, this app is developed only for the iOS platform, which can be expanded to Android in the future.

Conclusions

This study aimed to solve a real clinical problem, and the solution can reduce the gap between clinical practice and evidence-based guidelines for CINV management. Our study will promote the notion of precision medicine by integrating patient-related risk factors and antiemetic treatment recommendations. Hence, our efforts can lead to increased quality of the patients' life and reduced health care cost. An effort to reduce the care provider's time has high importance at the point of care. A less time-consuming decision support tool to predict patients at risk will help care providers provide better care in general.

Conflicts of Interest

None declared.

References

- 1. DeSantis CE, Lin CC, Mariotto AB, Siegel RL, Stein KD, Kramer JL, et al. Cancer treatment and survivorship statistics, 2014. CA Cancer J Clin 2014;64(4):252-271 [FREE Full text] [doi: 10.3322/caac.21235] [Medline: 24890451]
- 2. Hawkins R, Grunberg S. Chemotherapy-induced nausea and vomiting: challenges and opportunities for improved patient outcomes. Clin J Oncol Nurs 2009 Feb;13(1):54-64. [doi: <u>10.1188/09.CJON.54-64</u>] [Medline: <u>19193549</u>]
- Sun CC, Bodurka DC, Weaver CB, Rasu R, Wolf JK, Bevers MW, et al. Rankings and symptom assessments of side effects from chemotherapy: insights from experienced patients with ovarian cancer. Support Care Cancer 2005 Apr;13(4):219-227. [doi: 10.1007/s00520-004-0710-6] [Medline: 15538640]
- 4. Feyer P, Jordan K. Update and new trends in antiemetic therapy: the continuing need for novel therapies. Ann Oncol 2011 Jan;22(1):30-38 [FREE Full text] [doi: 10.1093/annonc/mdq600] [Medline: 20947707]

- 5. Heidary N, Naik H, Burgin S. Chemotherapeutic agents and the skin: an update. J Am Acad Dermatol 2008 Apr;58(4):545-570. [doi: 10.1016/j.jaad.2008.01.001] [Medline: 18342708]
- Davila M, Bresalier RS. Gastrointestinal complications of oncologic therapy. Nat Clin Pract Gastroenterol Hepatol 2008 Dec;5(12):682-696. [doi: <u>10.1038/ncpgasthep1277</u>] [Medline: <u>18941434</u>]
- Stein A, Voigt W, Jordan K. Chemotherapy-induced diarrhea: pathophysiology, frequency and guideline-based management. Ther Adv Med Oncol 2010 Jan;2(1):51-63 [FREE Full text] [doi: 10.1177/1758834009355164] [Medline: 21789126]
- 8. Haslam IS, Pitre A, Schuetz JD, Paus R. Protection against chemotherapy-induced alopecia: targeting ATP-binding cassette transporters in the hair follicle? Trends Pharmacol Sci 2013 Nov;34(11):599-604. [doi: 10.1016/j.tips.2013.09.001] [Medline: 24100054]
- 9. Paus R, Haslam IS, Sharov AA, Botchkarev VA. Pathobiology of chemotherapy-induced hair loss. Lancet Oncol 2013 Feb;14(2):e50-e59. [doi: 10.1016/S1470-2045(12)70553-3] [Medline: 23369683]
- Balagula Y, Lacouture ME, Cotliar JA. Dermatologic toxicities of targeted anticancer therapies. J Support Oncol 2010;8(4):149-161. [Medline: <u>20822032</u>]
- Hu C, Chan W, Delclos GP, Du XL. Adjuvant chemotherapy and risk of gastrointestinal, hematologic, and cardiac toxicities in elderly patients with stage III colon cancer. Am J Clin Oncol 2012 Jun;35(3):228-236. [doi: <u>10.1097/COC.0b013e318210f812</u>] [Medline: <u>21383605</u>]
- Coates A, Abraham S, Kaye SB, Sowerbutts T, Frewin C, Fox RM, et al. On the receiving end: patient perception of the side-effects of cancer chemotherapy. Eur J Cancer Clin Oncol 1983 Feb;19(2):203-208. [doi: <u>10.1016/0277-5379(83)90418-2</u>] [Medline: <u>6681766</u>]
- 13. Cooper S, Georgiou V. The impact of cytotoxic chemotherapy: perspectives from patients, specialists and nurses. Eur J Cancer 1992 Jan;28:S36-S38. [doi: 10.1016/0959-8049(92)90635-f] [Medline: 1627406]
- 14. Durant JR. The problem of nausea and vomiting in modern cancer chemotherapy. CA Cancer J Clin 1984;34(1):2-6 [FREE Full text] [doi: 10.3322/canjclin.34.1.2] [Medline: 6420016]
- 15. Morrow GR. Chemotherapy-related nausea and vomiting: etiology and management. CA Cancer J Clin 1989;39(2):89-104 [FREE Full text] [doi: 10.3322/canjclin.39.2.89] [Medline: 2495160]
- 16. Bloechl-Daum B, Deuson RR, Mavros P, Hansen M, Herrstedt J. Delayed nausea and vomiting continue to reduce patients' quality of life after highly and moderately emetogenic chemotherapy despite antiemetic treatment. J Clin Oncol 2006 Sep 20;24(27):4472-4478. [doi: 10.1200/JCO.2006.05.6382] [Medline: 16983116]
- Hilarius DL, Kloeg PH, van der Wall E, van den Heuvel JJG, Gundy CM, Aaronson NK. Chemotherapy-induced nausea and vomiting in daily clinical practice: a community hospital-based study. Support Care Cancer 2012 Jan;20(1):107-117 [FREE Full text] [doi: 10.1007/s00520-010-1073-9] [Medline: 21258948]
- Lachaine J, Yelle L, Kaizer L, Dufour A, Hopkins S, Deuson R. Chemotherapy-induced emesis: quality of life and economic impact in the context of current practice in Canada. Support Cancer Ther 2005 Apr 01;2(3):181-187. [doi: 10.3816/SCT.2005.n.011] [Medline: 18628170]
- Haiderali A, Menditto L, Good M, Teitelbaum A, Wegner J. Impact on daily functioning and indirect/direct costs associated with chemotherapy-induced nausea and vomiting (CINV) in a U.S. population. Support Care Cancer 2011 Jun;19(6):843-851. [doi: 10.1007/s00520-010-0915-9] [Medline: 20532923]
- Cohen L, de Moor CA, Eisenberg P, Ming EE, Hu H. Chemotherapy-induced nausea and vomiting: incidence and impact on patient quality of life at community oncology settings. Support Care Cancer 2007 May;15(5):497-503. [doi: 10.1007/s00520-006-0173-z] [Medline: 17103197]
- 21. Pirri C, Katris P, Trotter J, Bayliss E, Bennett R, Drummond P. Risk factors at pretreatment predicting treatment-induced nausea and vomiting in Australian cancer patients: a prospective, longitudinal, observational study. Support Care Cancer 2011 Oct;19(10):1549-1563. [doi: 10.1007/s00520-010-0982-y] [Medline: 20811914]
- Shih YT, Xu Y, Elting LS. Costs of uncontrolled chemotherapy-induced nausea and vomiting among working-age cancer patients receiving highly or moderately emetogenic chemotherapy. Cancer 2007 Aug 01;110(3):678-685 [FREE Full text] [doi: 10.1002/cncr.22823] [Medline: 17567835]
- 23. Ballatori E, Roila F, Ruggeri B, Porrozzi S, Iannopollo M, Soru G, et al. The cost of chemotherapy-induced nausea and vomiting in Italy. Support Care Cancer 2007 Jan;15(1):31-38. [doi: <u>10.1007/s00520-006-0094-x</u>] [Medline: <u>16788840</u>]
- 24. Ihbe-Heffinger A, Ehlken B, Bernard R, Berger K, Peschel C, Eichler H, et al. The impact of delayed chemotherapy-induced nausea and vomiting on patients, health resource utilization and costs in German cancer centers. Ann Oncol 2004 Mar;15(3):526-536 [FREE Full text] [doi: 10.1093/annonc/mdh110] [Medline: 14998860]
- 25. Viale PH, Grande C, Moore S. Efficacy and cost: avoiding undertreatment of chemotherapy-induced nausea and vomiting. Clin J Oncol Nurs 2012 Aug;16(4):E133-E141. [doi: <u>10.1188/12.CJON.E133-E141</u>] [Medline: <u>22842698</u>]
- 26. Yeh Y, McDonnell A, Klinger E, Fowler B, Matta L, Voit D, et al. Comparison of healthcare resource use between patients receiving ondansetron or palonosetron as prophylaxis for chemotherapy-induced nausea and vomiting. J Oncol Pharm Pract 2011 Sep;17(3):179-185. [doi: 10.1177/1078155210366491] [Medline: 20452991]
- Craver C, Gayle J, Balu S, Buchner D. Clinical and economic burden of chemotherapy-induced nausea and vomiting among patients with cancer in a hospital outpatient setting in the United States. J Med Econ 2011;14(1):87-98. [doi: 10.3111/13696998.2010.547237] [Medline: 21241160]

- Nevidjon B, Chaudhary R. Controlling emesis: evolving challenges, novel strategies. J Support Oncol 2010;8(4 Suppl 2):1-10. [Medline: 20865953]
- Schwartzberg LS. Chemotherapy-induced nausea and vomiting: clinician and patient perspectives. J Support Oncol 2007 Feb;5(2 Suppl 1):5-12. [Medline: <u>17366928</u>]
- 30. Richardson JL, Marks G, Levine A. The influence of symptoms of disease and side effects of treatment on compliance with cancer therapy. J Clin Oncol 1988 Nov;6(11):1746-1752. [doi: 10.1200/JCO.1988.6.11.1746] [Medline: 3183704]
- Hesketh PJ, Van Belle S, Aapro M, Tattersall FD, Naylor RJ, Hargreaves R, et al. Differential involvement of neurotransmitters through the time course of cisplatin-induced emesis as revealed by therapy with specific receptor antagonists. Eur J Cancer 2003 May;39(8):1074-1080. [doi: 10.1016/s0959-8049(02)00674-3] [Medline: 12736106]
- 32. Thompson N. Optimizing treatment outcomes in patients at risk for chemotherapy-induced nausea and vomiting. Clin J Oncol Nurs 2012 Jun 01;16(3):309-313. [doi: 10.1188/12.CJON.309-313] [Medline: 22641323]
- Basch E, Prestrud AA, Hesketh PJ, Kris MG, Feyer PC, Somerfield MR, American Society of Clinical Oncology. Antiemetics: American Society of Clinical Oncology clinical practice guideline update. J Clin Oncol 2011 Nov 01;29(31):4189-4198 [FREE Full text] [doi: 10.1200/JCO.2010.34.4614] [Medline: 21947834]
- Basch E, Hesketh PJ, Kris MG, Prestrud AA, Temin S, Lyman GH. Antiemetics: American Society of Clinical Oncology clinical practice guideline update. J Oncol Pract 2011 Nov;7(6):395-398 [FREE Full text] [doi: <u>10.1200/JOP.2011.000397</u>] [Medline: <u>22379425</u>]
- 35. Ettinger DS, Armstrong DK, Barbour S, Berger MJ, Bierman PJ, Bradbury B, National Comprehensive Cancer Network. Antiemesis. J Natl Compr Canc Netw 2012 Apr;10(4):456-485. [doi: <u>10.6004/jnccn.2012.0047</u>] [Medline: <u>22491046</u>]
- Roila F, Herrstedt J, Aapro M, Gralla RJ, Einhorn LH, Ballatori E, ESMO/MASCC Guidelines Working Group. Guideline update for MASCC and ESMO in the prevention of chemotherapy- and radiotherapy-induced nausea and vomiting: results of the Perugia consensus conference. Ann Oncol 2010 May;21 Suppl 5:v232-v243 [FREE Full text] [doi: 10.1093/annonc/mdq194] [Medline: 20555089]
- Jones JM, Qin R, Bardia A, Linquist B, Wolf S, Loprinzi CL. Antiemetics for chemotherapy-induced nausea and vomiting occurring despite prophylactic antiemetic therapy. J Palliat Med 2011 Jul;14(7):810-814 [FREE Full text] [doi: 10.1089/jpm.2011.0058] [Medline: 21554125]
- Glaus A, Knipping C, Morant R, Böhme C, Lebert B, Beldermann F, et al. Chemotherapy-induced nausea and vomiting in routine practice: a European perspective. Support Care Cancer 2004 Oct;12(10):708-715. [doi: <u>10.1007/s00520-004-0662-x</u>] [Medline: <u>15278682</u>]
- Molassiotis A, Saunders MP, Valle J, Wilson G, Lorigan P, Wardley A, et al. A prospective observational study of chemotherapy-related nausea and vomiting in routine practice in a UK cancer centre. Support Care Cancer 2008 Feb;16(2):201-208. [doi: 10.1007/s00520-007-0343-7] [Medline: 17926070]
- 40. Jordan K, Gralla R, Jahn F, Molassiotis A. International antiemetic guidelines on chemotherapy induced nausea and vomiting (CINV): content and implementation in daily routine practice. Eur J Pharmacol 2014 Jan 05;722:197-202. [doi: 10.1016/j.ejphar.2013.09.073] [Medline: 24157984]
- 41. Molassiotis A, Brearley SG, Stamataki Z. Use of antiemetics in the management of chemotherapy-related nausea and vomiting in current UK practice. Support Care Cancer 2011 Jul;19(7):949-956. [doi: 10.1007/s00520-010-0909-7] [Medline: 20574666]
- 42. Lee VC. Mobile Devices and Apps for Health Care Professionals: Uses and Benefits. USA: P T Internet MediMedia USA; 2014.
- 43. Mosa ASM, Yoo I, Sheets L. A systematic review of healthcare applications for smartphones. BMC Med Inform Decis Mak 2012 Jul 10;12(1):67 [FREE Full text] [doi: 10.1186/1472-6947-12-67] [Medline: 22781312]
- 44. Marinov M, Mosa ASM, Yoo I, Boren SA. Data-mining technologies for diabetes: a systematic review. J Diabetes Sci Technol 2011 Nov 01;5(6):1549-1556 [FREE Full text] [doi: 10.1177/193229681100500631] [Medline: 22226277]
- 45. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
- 46. Mosa A, Hossain AM, Lavoie B, Yoo I. Patient-related risk factors for chemotherapy-induced nausea and vomiting: a systematic review. Front Pharmacol 2020;11:329 [FREE Full text] [doi: 10.3389/fphar.2020.00329] [Medline: 32296333]
- 47. John G, Langley P. Estimating continuous distributions in Bayesian classifiers. arXiv. Preprint posted online February 20, 2013. [FREE Full text]
- 48. Cessie SL, Houwelingen JCV. Ridge estimators in logistic regression. Appl Stat 1992;41(1):191. [doi: 10.2307/2347628]
- 49. Freund Y, Schapire R. Large margin classification using the perceptron algorithm. Mach Learn Springer 1999;37(3):296. [doi: <u>10.1145/279943.279985</u>]
- 50. Platt J. Fast training of support vector machines using sequential minimal optimization. In: Advances in Kernel Methods: Support Vector Machine. New York, NY: ACM; 1998:41-65.
- Wang XZ, Buontempo FV, Young A, Osborn D. Induction of decision trees using genetic programming for modelling ecotoxicity data: adaptive discretization of real-valued endpoints. SAR QSAR Environ Res 2006 Oct;17(5):451-471. [doi: 10.1080/10629360600933723] [Medline: 17050186]

- 52. Quinlan JR. C4. 5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann; 1993.
- 53. Quinlan JR. Data Mining Tools See5 and C5.0. 2004. URL: <u>http://www.rulequest.com/see5-info.html</u> [accessed 2021-11-11]
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I. The WEKA data mining software. SIGKDD Explor Newsl 2009 Nov 16;11(1):10-18. [doi: 10.1145/1656274.1656278]
- Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, et al. Machine learning methods to predict diabetes complications. J Diabetes Sci Technol 2018 Mar 12;12(2):295-302 [FREE Full text] [doi: 10.1177/1932296817706375] [Medline: 28494618]
- 56. Berger MJ, Ettinger DS, Aston J, Barbour S, Bergsbaken J, Bierman PJ, et al. NCCN guidelines insights: antiemesis, version 2.2017. J Natl Compr Canc Netw 2017 Jul;15(7):883-893. [doi: <u>10.6004/jnccn.2017.0117</u>] [Medline: <u>28687576</u>]
- Roscoe JA, Morrow GR, Colagiuri B, Heckler CE, Pudlo BD, Colman L, et al. Insight in the prediction of chemotherapy-induced nausea. Support Care Cancer 2010 Jul;18(7):869-876 [FREE Full text] [doi: 10.1007/s00520-009-0723-2] [Medline: 19701781]
- 58. Osoba D, Zee B, Pater J, Warr D, Latreille J, Kaizer L. Determinants of postchemotherapy nausea and vomiting in patients with cancer. Quality of Life and Symptom Control Committees of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 1997 Jan;15(1):116-123. [doi: 10.1200/JCO.1997.15.1.116] [Medline: 8996132]
- 59. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: <u>10.1093/jnci/85.5.365</u>] [Medline: <u>8433390</u>]

Abbreviations

AI: artificial intelligence
ASCO: American Society of Clinical Oncology
BMI: body mass index
CINV: chemotherapy-induced nausea and vomiting
EMR: electronic medical record
ESMO: European Society of Medical Oncology
GUI: graphical user interface
HEC: high-emetogenic chemotherapy
KDD: knowledge discovery in databases
LEC: low-emetogenic chemotherapy
MASCC: Multinational Association of Supportive Care in Cancer
MEC: moderate-emetogenic chemotherapy
ML: machine learning
NCCN: National Comprehensive Cancer Network
NK-1: neurokinin-1

Edited by R Kukafka, G Eysenbach; submitted 20.01.21; peer-reviewed by YC Chen, A Khayat; comments to author 25.02.21; revised version received 11.03.21; accepted 04.08.21; published 02.12.21

Please cite as:

Mosa ASM, Rana MKZ, Islam H, Hossain AKMM, Yoo I A Smartphone-Based Decision Support Tool for Predicting Patients at Risk of Chemotherapy-Induced Nausea and Vomiting: Retrospective Study on App Development Using Decision Tree Induction JMIR Mhealth Uhealth 2021;9(12):e27024 URL: https://mhealth.jmir.org/2021/12/e27024 doi: 10.2196/27024 PMID:

©Abu Saleh Mohammad Mosa, Md Kamruz Zaman Rana, Humayera Islam, A K M Mosharraf Hossain, Illhoi Yoo. Originally published in JMIR mHealth and uHealth (https://mhealth.jmir.org), 02.12.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR mHealth and uHealth, is properly cited. The complete bibliographic information, a link to the original publication on https://mhealth.jmir.org/, as well as this copyright and license information must be included.