Predicting the Pharmaceutical Needs of Hospitals

Amir Hossein Nabizadeh (✉ amir.nabizadeh@tecnico.ulisboa.pt)
Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman

Mohammed Mehdi Ghaemi
Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman

Daniel Goncalves
Instituto Superior Técnico, Universidade de Lisboa

Research Article

Keywords: Pharmaceutical Prediction, Random Forest, Neural Network, Confusion Matrix, Data Balancing

Posted Date: August 22nd, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3263123/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.
Predicting the Pharmaceutical Needs of Hospitals

Amir Hossein Nabizadeh\textsuperscript{1,2}\textsuperscript{*}, Mohammed Mehdi Ghaemi\textsuperscript{2†} and Daniel Goncalves\textsuperscript{2,3†}

\textsuperscript{1}Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, 7616913555, Iran.
\textsuperscript{2}INESC-ID, Rua Alves Redol, 9, Lisbon, 1000-029, Portugal.
\textsuperscript{3}Instituto Superior Tecnico, University of Lisbon, Lisbon, 1000-029, Portugal.

\textsuperscript{*}Corresponding author(s). E-mail(s): amir.nabizadeh@tecnico.ulisboa.pt; m.ghaemi@kmu.ac.ir; daniel.goncalves@inesc-id.pt;
\textsuperscript{†}These authors contributed equally to this work.

Abstract

\textbf{Purpose:} People’s lives are always threatened by various diseases. The role of health and medical services, in particular medicine, is undeniable in protecting their lives. Timely preparation and providing medicine for patients is vital since medicine shortage can endanger their lives while excessive accumulation of medicine can put them at expiration risk and waste health budgets. In this paper, we introduce a model for the prediction of commonly used medicine (type and amount) in hospitals.

\textbf{Methods:} We have used a dataset of Afzalipur Hospital in Kerman collected for three years consisting of 283 features, which included over 12293431 medicine and 9531 patients. Nine features were selected using experts’ feedback and were fed into the random forest and neural network algorithms. For the prediction task, medicine types and their amounts were predicted for each individual using different training sets. In addition, the right prediction time was also found which is when predictions have a promising accuracy while the executive team of a hospital has enough time to provide the right amounts of the most used medicine.

\textbf{Results:} The performance of algorithms was evaluated using a confusion matrix. Our results showed that the random forest had a promising performance in predicting the amounts of the most used medicine for a month using two years of data (accuracy 83.3\%) while its accuracy in predicting medicine was 35.9\%.

1
**Conclusion:** Although our results are promising, in particular, for predicting the amounts of medicine, these results could be enhanced and more reliable using other metadata like "patient underlying disease" and "medical tests results".

**Keywords:** Pharmaceutical Prediction, Random Forest, Neural Network, Confusion Matrix, Data Balancing.

1 Introduction

People’s health is completely dependent on medicine, and one of the most important criteria in the health system of any country is the amount of medicine consumed in that country. Medicine often poses an excessive cost to individuals and the healthcare system. Global budget spending on medicine in 2020 was around 1.3 trillion dollars while the United States alone spent around 350 billion dollars on it. These budgets are expected to rise at a rate of 3% to 6% annually worldwide. Funding medicine becomes more challenging when we know that the lack of resources in providing medicine and unequal access to it can negatively influence this rate [1].

Regarding the mentioned challenges, accurate management of supplying medicine is essential. The final goal of this management is the timely preparation and distribution of medicine among patients since the medicine shortage can endanger their lives, and on the other hand, excessive accumulation of medicine can put medicine at risk of expiration and waste health budgets.

In the past, traditional and simple methods have been used for estimating and predicting the required medicine. For instance, providing and storing a huge bulk of flu medicine in the autumn and winter seasons. Today, accurate and cost-effective methods have been introduced and used in developed countries for the prediction of required medicine that work based on evidence (e.g. dosage) and uncertainty variables [1–4]. These methods focused on various topics, such as medicine prediction [5–11], or medicine cost prediction [12–17].

Mahajan and Kumar implemented a time series model to predict the sales of pharmaceutical distribution companies [18] using various approaches, such as the ARIMA method [19], neural networks [20], advanced neural networks [21], and fuzzy neural networks [22]. Their results showed that the method could make an approximate prediction in real-time. Neural networks have also been used in [23]. In this study, Chang and his colleagues performed a weekly forecasting of pharmaceutical products using a deep neural network [24]. For the evaluation, a week-ahead sales prediction was done using three years of daily sales, which were grouped weekly. In another study, researchers used Quantum Neural Networks-QNN [25] to predict the exact generic medicine for a patient by considering disease factors [26]. For that, initially, the symptoms of disease (105 symptoms) were categorized by experts. To assess, the results of Support Vector Machines - SVM [27], Bayes [28], Random Forest - RF [29], and QNN were compared, which the QNN model had reached the accuracy of 95%.

Alves predicted the demand for hospital consumables [30] using neural networks - RNN [31], SVM, and RF with the help of parametric methods, such as Holt-Winters
model exponential smoothing [32], and the ARIMA method [19]. Results revealed that the machine learning algorithms were able to produce the lowest average prediction error. It was also concluded that the classification of items directly affects the performance of the prediction models. In a similar study, authors presented a model to use large data sets (50 variables collected from 101766 patients in 120 hospitals) for recommending new medicine to diabetic patients [33]. Ten variables were selected via feature selection methods and fed into the RF, Multi-Layer Perceptron - MLP [34], Logistic Regression - LR [35], J48 [36], and SVM algorithms, which RF and MLP had the best performance. RF, NN, and other methods like linear regression were also applied by Mbonynshuti et al. to predict the future trends in demanding the ten most used medicine in Rwanda [37]. To evaluate the results, predictions were compared with the actual values. Results showed that the RF had the best performance. In another study, the same authors used the RF model to predict the demand for essential medicine for Non-Communicable Diseases (NCD) based on past consumption data [38]. Their data included more than 500 medicine, and only 17 of the most used ones were selected for the prediction task. The evaluation showed that the RF model could accurately predict the medicine for a month.

Besides the mentioned studies that mainly used RF and NN algorithms, some studies used other methods for the prediction task. Permanasari et al. applied the short-term memory method - LSTM [39] to predict the need for medicine containing digestive enzymes [40]. Here, the autocorrelation function - ACF [41] and partial autocorrelation function - PACF [42] were used to identify the input of LSTM. Their results showed that the LSTM was highly accurate in predicting time series data. Time series methods were also used in [43] to predict future medicine consumption using 32 variables. Among these variables, medicine consumption, seasonal changes, distribution data, warehouse inventory, and purchase orders were the main ones. Quantitative evaluation conducted in five hospitals in India showed that the actual time required to receive the medicine varies between 15 and 30 days. Similarly, Papan and Fotiadis applied time series approaches to predict the consumption of RAPILYSIN LYPDINJ [44]. Autocorrelated Moving Average - ARMA [45] and Autocorrelated Moving Average Regression - ARIMA [19] models were the ones that applied for the prediction task. The evaluation was conducted using the consumption data of RAPILYSIN LYPDINJ collected for three years from a hospital in Singapore.

In [46], a model was introduced to assist patients in finding appropriate medicine for their diseases, which can be extended by hospitals to predict their needs. This model analyzed the history of a patient to check for any side effects of a suggested medicine. In addition, it checked the weather conditions and local maps (via Google’s API) where the patient was located, so that he/she could find nearby drugstores that have the medicine. In another study, researchers designed a preliminary prediction approach. It worked based on a dynamic meta-analysis to predict the temporary sale of a new generic medicine when there is a complete lack of information [47]. During the evaluation, a dynamic meta-analysis of the release parameters was able to predict the cycle of the medicine. Performance evaluation verified a high degree of accuracy between the previously observed values and its predicted average.
In [48], Natural Language Processing - NLP [49] combined with deep learning used on triage notes and other clinical data to predict the number of resources needed for a patient in the Emergency Department (ED) of a hospital. The evaluation data included the data of 144421 patients consisting of nursing triage notes and clinical variables collected at the triage time. Model training and validation were conducted using the Keras software library via TensorFlow 2.0 [50] through the Amazon Web Services Sagemaker [51]. To evaluate the model, the prediction values were compared with the estimations of two experienced nurses in the ED. The accuracy of predictions was almost similar to the accuracy of the nurses.

Despite the use of modern predicting methods by the aforementioned studies, these studies often ignored the time factor in their predictions. In addition, these methods were usually implemented and used in developed countries and have not been widespread in third-world countries. In the least developed countries, hospitals and medical centers still use traditional methods to estimate their required medicine. Traditional methods are not very complicated and do not have high accuracy [1]. The low accuracy of these methods can have various reasons, including not taking into account conditions, such as the amount of medicine produced by companies, medicine price variation, lack of raw materials, etc., which can have a high and direct impact on the amount of production and consumption of medicine. Machine Learning (ML) methods can be applied as a solution to prevent an unreasonable increase in medicine stock and, as a result, mutation of their price. To this end, in this study, we apply ML-based methods to predict the need (amounts and types of medicine) for the most used medicine in a hospital. For this purpose, we collected data on medicine used by patients in Afzalipour Hospital in Kerman. It was collected from 2018 to 2020 and included the data of 9351 different medicine prescribed to 121690 patients.

In brief, the contribution of our work is the following:

- Generating a hospital dataset collected for three years including 283 features.
- Proposing a model to predict the most used medicine of a hospital using ML techniques a month ahead.
- Proposing a model to predict the amounts of the most used medicine of a hospital using ML techniques a month ahead.
- Finding prediction time. It is when the predictions have a promising accuracy while management team of a hospital has enough time to provide the right amounts of the most used medicine.
- Finding a prediction model that constantly have a promising accuracy.
- Enhancing the prediction accuracy using a class balancing technique.

2 Research Methodology

In this section, we detail our prediction approach, describe what type of data is collected, how the most used medicine are identified, and what algorithms are used for the prediction task. In addition, we explain how we apply a feature selection technique to drop insignificant data attributes to improve prediction accuracy. Figure 1 depicts a general view of our approach and the steps it takes for the prediction.
2.1 Data Collection

Our data was collected from Afzalipur Hospital in Kerman from 2018 to 2020, including 283 clinical and demographical features. Table 1 details the data for each year. As presented in this table, the number of admission in 2020 was much less than in 2018 and 19. It was due to the reason that this hospital was mainly assigned for Covid-19 cases.

Table 1 Number of Admissions, Medicine, and total amounts of prescription.

<table>
<thead>
<tr>
<th>Year</th>
<th># Admission</th>
<th># Different Medicine</th>
<th># Prescribed Medicine</th>
<th># System Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>48012</td>
<td>4832</td>
<td>40635</td>
<td>2132075</td>
</tr>
<tr>
<td>2019</td>
<td>45171</td>
<td>4066</td>
<td>39946</td>
<td>2231849</td>
</tr>
<tr>
<td>2020</td>
<td>28507</td>
<td>3467</td>
<td>42352</td>
<td>2029816</td>
</tr>
<tr>
<td>Total</td>
<td>121690</td>
<td>9354</td>
<td>122934</td>
<td>6393740</td>
</tr>
</tbody>
</table>

Figure 2 shows the prescription amounts of each medicine. The most used medicine was prescribed less than 400K times in three years, and the least one was cisplatin/vial/0.5 mg/ml, 100 ml, which was prescribed only once. Among 9351 medicine, 1057 of them have been prescribed more than 1000 times. After an initial analysis, it was found that the medicine with the highest prescription was "plastic distilled water", which is not a medicine. So, these kind of medical items were ignored in our data.
2.2 Data Pre-process

Preprocessing played a pivotal role in our process and took a considerable amount of our time. Our preprocessing phase was conducted in three steps using R programming language. In the first step, we ignored the records having missing (Null) or invalid data and did not consider them in the prediction process.

In the second step, duplicated variables were removed from the data set. As mentioned earlier, the initial data consisted of 283 variables, some duplicated and stored with different names in different columns. Therefore, these types of variables were also ignored.

In the last step, it was found that several admissions had negative values for the amounts of prescribed medicine. It happened due to mistyping the amounts by operators, and also the improper design of the system in not allowing invalid amounts. After a consultation with an expert, these values were considered positive.

2.3 Identify Target Medicine

As stated before, we aim at predicting the type and amounts of the most used medicine in a hospital. For that, we initially needed to estimate the amounts of each medicine. Due to the negligence of operators and faulty design of the database, there were medicine stored having different IDs (Table 2). For instance, chlorsodium dextrose serum was stored having five various IDs. To solve this problem, we unified all IDs for a single medicine and accumulated their amounts. Due to the sensitivity of our field, this step was done manually. We then analyzed the amounts of all medicine and found that 20 medicine had a clear distance from the rest. Therefore, we only concentrated on them. Table 2 shows the top 20 medicine along with their ID and amounts.

2.4 Feature Selection

Feature selection is a technique to identify the relevant and informative data attributes while discarding the redundant and irrelevant ones [52–54]. It has several benefits. It avoids an overfitting problem [55, 56], facilitates the interpretability (easy to understand) of the prediction results [57–60], enhances the learning speed, reduces the storage volume, and decreases the noise caused by redundant and irrelevant attributes [61]. To this end, we used Boruta method that works based on an RF algorithm [62]. It iteratively compares the importance of original attributes with the shadow ones (created by shuffling the original attributes). The original attributes having lower importance than the shadow ones are dropped while the higher ones are kept as the main (confirmed) attributes. The shadow attributes are regenerated in each iteration. Boruta stops when no confirmed attributes are left or the stop criteria is met (i.e. max iteration).

By applying the Boruta only one attribute was detected as the confirmed one. After discussing the results with experts, we concluded that we do the feature selection manually using the experts’ feedback. Therefore, after discarding the duplicated features (40 features), constant features (100 features), and the irrelevant ones (134 features), we ended up with nine features including patients’ age, blood type, sex,
Table 2  Top 20 Medicine with their ID and amounts. Sorted by their amount.

<table>
<thead>
<tr>
<th>Medicine</th>
<th>Amount</th>
<th>ID01</th>
<th>ID02</th>
<th>ID03</th>
<th>ID04</th>
<th>ID05</th>
<th>ID06</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dextrose chlorsodium serum injection 0.3% 0.5 liters</td>
<td>178025</td>
<td>65165</td>
<td>53607</td>
<td>54922</td>
<td>65031</td>
<td>52363</td>
<td>-</td>
</tr>
<tr>
<td>Sodium chloride serum injection 0.9%/0.5 liter</td>
<td>177304</td>
<td>42870</td>
<td>52384</td>
<td>54345</td>
<td>46789</td>
<td>65004</td>
<td>56707</td>
</tr>
<tr>
<td>Sodium heparin/ampoule/5000 units/1 liter</td>
<td>128955</td>
<td>51857</td>
<td>55200</td>
<td>53788</td>
<td>36339</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clindamycin Phosphate/Ampoule/300mg/2ml</td>
<td>89676</td>
<td>51844</td>
<td>40009</td>
<td>55275</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vancomycin/vial/500 mg</td>
<td>58411</td>
<td>95643</td>
<td>48179</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dexamethasone/ampoule/8mg/2ml</td>
<td>57110</td>
<td>65019</td>
<td>51847</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pantazol vial / 40 mg</td>
<td>50976</td>
<td>47138</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clindamycin Phosphate/Ampoule/300mg/2ml</td>
<td>59264</td>
<td>45785</td>
<td>51847</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Calcium gluconate/ampoule/10%</td>
<td>31634</td>
<td>36200</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Potassium chloride/ampoule/10ml</td>
<td>29935</td>
<td>57479</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Naloxone/ampoule/0.4mg/1ml</td>
<td>29352</td>
<td>56354</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Azithromycin/capsule/250 mg</td>
<td>27688</td>
<td>66221</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Calcium carbonate/pill/500 mg</td>
<td>27028</td>
<td>56178</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Pantoprazole/pill/40 mg</td>
<td>26972</td>
<td>43079</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sodium chloride serum bottle/injection/0.4%/0.5 liter</td>
<td>26117</td>
<td>51921</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Famotidine/pill/40 mg</td>
<td>22600</td>
<td>64620</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Furosemide/ampoule/20 mg/2 ml</td>
<td>22187</td>
<td>46813</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Prednisolone/pill/5 mg</td>
<td>22026</td>
<td>47310</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ciprofloxacin/bag/200mg/100ml</td>
<td>21251</td>
<td>46524</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ranitidine/ampoule/50 mg</td>
<td>19657</td>
<td>46831</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

diagnosis code, disease code, physician code, admission month, prescribed medicine and their amounts for each patient.

2.5 Prediction Algorithms

After determining the informative features and discarding the irrelevant ones, we have applied the regular form of two predictive algorithms, Random Forest (RF) and Neural Network (NN) for the prediction task. Random Forest (RF) is a combination of decision tree predictors. Trees are depending on values of random vectors, which are sampled independently while having the same distribution [63–65]. RF is commonly used by researchers and has several advantages, such as high accuracy, modeling the complex interactions among predictor attributes, handling data with numerous features, reducing the over-fitting issue, parallel computation, and robustness to emissions data due to random sampling [66–68]. Neural Networks (NN) are made of node layers, containing one input layer, one or more hidden layers, and one output layer. Each node (neuron) is connected to another one and has an associated weight and threshold. If the output of a node is more than a specified threshold, that node is activated, sending data to the next layer. NN provide various advantages over traditional algorithms. They learn from data and can be applied to solve complex problems, and are also able to generalize (i.e. recognize patterns in data that traditional algorithms may not). Moreover, they are scalable and can handle large amounts of data quickly and accurately [69–73].
Our prediction was twofold. We initially predicted the amounts of medicine. For that, all amounts of the prescribed medicine for each patient were converted into 10-point intervals (e.g. all amounts from 1 to 10 were assigned to the first interval), and then these intervals were predicted. We then predicted the type of medicine.

In total, twenty-four models were built for the prediction task:

- Twelve predictive models (medicine and their amounts) for a month using RF and NN methods using one, two, and three years of data (four models for each year).
- Twelve predictive models (medicine and their amounts) for a season using RF and NN methods using one, two, and three years of data (four models for each year).

In our study, our training sets were the data of one, two, and three years while the test set was the data of a month/season. Different amounts of data were used for training to analyze how much data is needed for having a promising prediction accuracy, and also monitor which method consistently works better.

To find the optimal values for the hyper-parameters of the models, we used a 5-fold cross-validation technique along with the TuneLength method in the R programming language. TuneLength considered 5 default values for each of the parameters and selected a value that provided the highest accuracy for the prediction.

2.6 Evaluation

Prescribing the right medicine is vital since the wrong prescription can have the risk of death. Therefore, the high accuracy of our models is of special importance. To do so, we used the confusion matrix that estimates the effectiveness of the models. Based on this matrix, we calculated various measures, such as Accuracy, Recall, Precision, and F-1. The results for the last three measures (Recall, Precision, and F-1) are presented in A.3.

3 Results

As explained, our predictions were made for a month/season using different sets of training data:

- Prediction of medicine and their amounts for a month using data of one/two/three years.
- Prediction of medicine and their amounts for a season using data of one/two/three years.

We initially evaluated the prediction accuracy of NN and RF models. To generate the models, the data of one month/season was used as a test set, and one/two/three years of data was used as training sets (Figure 3).

Based on the results presented in Table 3, RF often outperformed the NN models. In this table, the accuracy using one year of training data is usually lower than the rest of the years. It could be due to not having enough data for training the models. On the other hand, the accuracy of results using three years of data is less than two years, which might be because of increasing the noise in the collected data that negatively
influences the accuracy. Considering the results in Table 3, both models had their best overall performance using two years of training data.

**Table 3** RF and NN Prediction Accuracy (Before Balance).

<table>
<thead>
<tr>
<th>Test Sets</th>
<th>Training Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>1 Year</td>
<td>24.22%</td>
<td>28.66%</td>
</tr>
<tr>
<td>RF</td>
<td>2 Years</td>
<td>31.89%</td>
<td>28.49%</td>
</tr>
<tr>
<td>RF</td>
<td>3 Years</td>
<td>27.15%</td>
<td>28.25%</td>
</tr>
<tr>
<td>NN</td>
<td>1 Year</td>
<td>8.11%</td>
<td>22.70%</td>
</tr>
<tr>
<td>NN</td>
<td>2 Years</td>
<td>28.79%</td>
<td>21.20%</td>
</tr>
<tr>
<td>NN</td>
<td>3 Years</td>
<td>9.23%</td>
<td>22.70%</td>
</tr>
</tbody>
</table>

By analyzing the prediction results, we realized that the accuracy of predicting medicine is significantly lower than the accuracy of their amounts. It could be due to the reason that all medicine were not prescribed in a balanced way, and the prescription of some medicine was far more than the others (Figure 4). Since some predictive models are very sensitive to imbalanced data, it is possible that this issue negatively influences their performance. For this purpose, after balancing the data of the prescribed medicine via SMOTE technique [74], we re-evaluated the accuracy of the models. As presented in Table 4, the accuracy of predicting medicine has been significantly improved.

**Table 4** RF and NN Prediction Accuracy (After Balance).

<table>
<thead>
<tr>
<th>Test Sets</th>
<th>Training Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>1 Year</td>
<td>83.18%</td>
<td>72.53%</td>
</tr>
<tr>
<td>RF</td>
<td>2 Years</td>
<td>35.95%</td>
<td>40.36%</td>
</tr>
<tr>
<td>RF</td>
<td>3 Years</td>
<td>39.53%</td>
<td>43.21%</td>
</tr>
<tr>
<td>NN</td>
<td>1 Year</td>
<td>8.09%</td>
<td>NA</td>
</tr>
<tr>
<td>NN</td>
<td>2 Years</td>
<td>28.79%</td>
<td>22.70%</td>
</tr>
<tr>
<td>NN</td>
<td>3 Years</td>
<td>29.33%</td>
<td>22.70%</td>
</tr>
</tbody>
</table>

Although the accuracy of medicine prediction has improved after balancing, the accuracy of their amounts, which was acceptable before balancing, has decreased. To
this end, we decided to use balancing only for the prediction of medicine and not use it for their amounts.

Due to the reason that RF often outperformed the NN models and normally the higher accuracies were obtained using two years of training data, our selection for the prediction is the RF model using two years of training data for the prediction of a month.

4 Discussion

The evaluation results showed both RF and NN models had almost the same accuracies (RF model had a slightly higher accuracy) while the time to generate RF models was far more than the NN models (results in A.2). The highest accuracy in predicting the amount of medicine belonged to the RF model, 83.3% (for a month) and 81.3% (for a season) using two years of data. Medicine prediction was far less accurate, and the best results using the RF model for a month and a season were 31.8% and 28.4%, respectively. Although the prediction accuracy for the amount of medicine was acceptable, it was low in medicine prediction. To enhance the accuracy, we balanced the prescribed medicine classes (reproduced medicines that were prescribed much less than the others). After balancing, the accuracies for medicine prediction raised to 35.9% (a month) and 40.3% (a season).

It is of interest that our results are compatible with the ones presented in (Mbonyinshuti et al., 2022). In this study, the authors used the RF model to predict the need for NCD medicine. Their results showed that the RF could accurately predict the medicine a month ahead. In addition, Iqbal et al. predicted future medicine consumption (Iqbal et al., 2017). They stated that the prediction should be made 15 to 30 days before the actual time of medicine requirement.
4.1 Limitations

Like any other study, our study had limitations that can be tackled in the future. These limitations are listed below:

- **Noisy Data**: Results showed that the accuracy did not enhance by increasing the amount of data. It could be because the noise increased with the increase of data, which negatively affects the accuracy.

- **Faulty Hospital System**: In the preprocessing phase, we found that the Afzalipur Hospital system has serious problems, including recording similar data in different columns, patient admission with no date, or even recording negative values for the volume of prescribed medicine. Although some of the issues are due to the inaccuracy of the system operator, the appropriate design of the system can diminish the mentioned issues to a large extent. It can result in having certified data, which might enhance the prediction accuracy and guarantee its reliability.

- **Metadata**: We used a series of patient, disease, and hospital variables for our prediction. There are still variables, such as "patient underlying disease" and "medical tests results" that can be explored to enhance the accuracy of predictions.

- **Results Uncertainty**: Despite predicting medicine and their amounts, results are not reliable and cannot be generalized since the data was insufficient, noisy, and was for a limited time frame (only three years).

- **Low Accuracy of Medicine Prediction**: The RF model could predict the amount of the most used medicine, but its accuracy for the prediction of medicine was not promising. Although by balancing the data accuracy of predicting medicine improved, these results are still low. Other models that consider time series data, like ARIMA, might help to raise the accuracy.

4.2 Future Directions

In this study, we predicted the most used medicine along with their amounts for all departments of a Hospital. Due to variations in the volume of the pharmaceutical needs of different departments, the possibility of accurately predicting these needs was trivial. Therefore, in the future, instead of focusing on all departments, which include a wide variety of prescribed medicine, it is recommended to concentrate only on one department. It can increase prediction accuracy due to the convergence of prescribed medicine.

Although predicting the need for commonly used medicine is of great importance, this prediction could be more beneficial considering other factors, such as medicine price, their criticality, and rarity.

Conclusions

In this study, we predicted the most used medicine (type and amount) in a hospital. For that, we used a dataset collected from a Hospital for three years. Nine features of the collected data, including patients’ age, blood type, sex, diagnosis code, disease code, physician code, admission month, prescribed medicine, and their amounts
were fed into the random forest and neural network algorithms. Finally, the performance of algorithms was evaluated using a confusion matrix. Our results showed that the random forest had a promising performance in predicting the amount of required medicine a month ahead using two years of data, but its accuracy in predicting medicine type was low (31.8%). We then balanced the dataset and could raise the accuracy of medicine type to 35.9%. Despite the promising accuracy of the random forest model, the required time to generate its model is far more than the NN model.

In the future, we intend to focus only on the requirements for one department of a hospital and use other factors, such as the rarity of medicine, for the prediction task.

Declarations

The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval

All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Kerman University of Medical Sciences, Kerman, Iran (IR.KMU.REC.1401.076).

Competing interests

The authors declare that they have no conflict of interest.

Authors’ contributions

The first author designed the prediction method, wrote the manuscript, and conducted the evaluation. The second author collected the dataset and helped in improving the initial idea. The last author revised the manuscript and assisted the first author in the evaluation task.

Funding

The authors did not receive support from any organization for the submitted work.

Availability of data and materials

Not applicable.

References


new medicines: searching for the balance between rising costs and limited budgets. Frontiers in Public Health 6, 328 (2018)


[49] Chowdhary, K., Chowdhary, K.: Natural language processing. Fundamentals of
artificial intelligence, 603–649 (2020)


919–926 (2016)


Appendix A

A.1 Accuracy of the generated models

In Table A1, higher accuracies are bolded. As shown in this table, RF has been superior to NN models in most cases. Sometimes the accuracy of the models is the same up to two decimal places, but considering more decimal places, their accuracy will be slightly different.

<table>
<thead>
<tr>
<th>Training Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>47.15%</td>
<td>36.93%</td>
</tr>
<tr>
<td></td>
<td>43.00%</td>
<td>34.00%</td>
</tr>
<tr>
<td></td>
<td>48.29%</td>
<td>37.13%</td>
</tr>
<tr>
<td>NN</td>
<td>37.76%</td>
<td>31.76%</td>
</tr>
<tr>
<td></td>
<td>30.40%</td>
<td>30.40%</td>
</tr>
<tr>
<td></td>
<td>39.73%</td>
<td>32.08%</td>
</tr>
</tbody>
</table>

A.2 Required time to generate the models

In Table A2, we presented the required time to build each model. It provides a better understanding of generating the models, and also when the accuracy of models is similar, it can be used as a selection criteria. The required time to generate NN models is significantly less than the RF models.

<table>
<thead>
<tr>
<th>Training Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td>RF</td>
<td>240.95</td>
<td>841.95</td>
</tr>
<tr>
<td></td>
<td>1484.9</td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>32.01</td>
<td>110.22</td>
</tr>
<tr>
<td></td>
<td>190.8</td>
<td></td>
</tr>
</tbody>
</table>

A.3 Precision, Recall and F1 of NN and RF models

The results presented in Tables A3 and A4 are compatible with the ones in Tables 3 and 4. In Tables A3 and A4, the Recall of the NN model is sometimes equal to 100%. It shows that the probability of not predicting the prescribed medicine is very low.
### Table A3  Precision, Recall and F1 of NN and RF Before Balancing.

<table>
<thead>
<tr>
<th>Test Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training Sets</strong></td>
<td>1 Year</td>
<td>2 Years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>33.33%</td>
<td>35.18%</td>
</tr>
<tr>
<td>NN</td>
<td>8.11%</td>
<td>28.79%</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>97.78%</td>
<td>76.54%</td>
</tr>
<tr>
<td>NN</td>
<td>15.01%</td>
<td>44.71%</td>
</tr>
<tr>
<td><strong>F1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>39.33%</td>
<td>43.78%</td>
</tr>
<tr>
<td>NN</td>
<td>8.09%</td>
<td>28.76%</td>
</tr>
</tbody>
</table>

### Table A4  Precision, Recall and F1 of NN and RF After Balancing.

<table>
<thead>
<tr>
<th>Test Sets</th>
<th>A Month</th>
<th>A Season</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training Sets</strong></td>
<td>1 Year</td>
<td>2 Years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>NN</td>
<td>8.09%</td>
<td>28.76%</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>98.81%</td>
<td>99.86%</td>
</tr>
<tr>
<td>NN</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td><strong>F1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RF</td>
<td>100.00%</td>
<td>78.46%</td>
</tr>
<tr>
<td>NN</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>