eWound-PRIOR: An Ensemble Framework for Cases Prioritization After Orthopedic Surgeries

Felipe Neves[1], Morgan Jennings[2], Miriam Capretz[3], Dianne Bryant[2], Fernanda Campos[1], and Victor Ströele[1(✉)]

[1] Department of Computer Science, Federal University of Juiz de Fora, Juiz de Fora, Brazil
{felipe.neves.braz,victor.stroele}@ice.ufjf.br,
fernanda.campos@ufjf.edu.br

[2] Department of Physical Therapy, The University of Western Ontario, London, Canada
{mjenni3,dianne.bryant}@uwo.ca

[3] Department of Electrical and Computer Engineering, The University of Western Ontario, London, Canada
mcapretz@uwo.ca

Abstract. Patient follow-up appointments are an imperative part of the healthcare model to ensure safe patient recovery and proper course of treatment. The use of mobile devices can help patient monitoring and predictive approaches can provide computational support to identify deteriorating cases. Aiming to aggregate the data produced by those devices with the power of predictive approaches, this paper proposes the eWound-PRIOR framework to provide a remote assessment of postoperative orthopedic wounds. Our approach uses Artificial Intelligence (AI) techniques to process patients' data related to postoperative wound healing and makes predictions as to whether the patient requires an in-person assessment or not. The experiment results showed that the predictions are promising and adherent to the application context, even if the on-line questionnaire had impaired the training model and the performance.

1 Introduction

Patients are becoming great consumers of health information through the use of the Internet. It can favor an innovative healthcare model, providing the opportunity to support services like on-line consultation, patient and physician's education, appointment booking, patient's assessment, and monitoring. In large hospitals or clinics, the amount of daily information is usually expressive as different types of medical data are available for treatments and diagnosis. These data help physicians in their daily work, allowing them to understand the factors related to the patients' health.

Based on the hospital dataset, predictions can be made aiming to help physicians to understand the factors of a specific disease. Several works adopted predictive approaches in diagnosis and, forms of treatment [5], detection and risk of chronic diseases [6, 11] and, the use of medical notes to make predictions in health centers [12]. In this context, this

research deals with postoperative patients of orthopedic surgeries, identifying worsening in treatments, and informing a case prioritization. Patient follow-up appointments are an imperative part of the healthcare model to ensure safe patient recovery and proper course of treatment. The current standard of care following elected orthopedic surgery includes routine in-person follow up. However, technology now exists to enable remote patient assessment, so the patient may not need to come into the clinic physically. Marsh et al. [1] compared a web-based follow up for patients at least 12 months following total hip/knee arthroplasty to the standard in-person assessment. The web-based follow-up was feasible, cost-effective alternative and as successful as in-person assessment at identifying adverse events (i.e., no adverse event was missed).

The incidence of early adverse events (within three months) following elected orthopedic surgery are rare. Early minor complications may include infection, thromboembolic events, stiffness, and instability and are more common than major complications or death. Surgical site infection rates for common elected orthopedic surgeries such as hip/knee replacement, high tibial osteotomy, and anterior cruciate ligament reconstruction (ACLR), are1.4%, 5% and, 0.48% respectively [2, 3]. Thus, follow-up appointments within the first three months are often unremarkable with no change in clinical management [4]. An online follow-up assessment may allow surgeons to maximize their efficacy while maintaining the effectiveness of identifying adverse events, such as surgical site infection, to manage patient treatment. While in-person follow-up involves many patient questions and assessments, a surgical site assessment is one component that if it is done remotely, may ease patient concerns during recovery and assist with prioritization of patients in a clinical setting.

Previous literature has examined predictive approaches in patient diagnosis and treatment [5], and the detection and risk of chronic diseases [6]. Common concerns about using predictive approaches focus on patient's care. A predictive approach requires special attention to patient monitoring and computational support of changes in patient health to ensure patient safety and that no worsening be missed.

One promising approach is the adoption of Artificial Intelligence predictive models in healthcare environments, which favors the patients' classification by context. These models, based on Machine Learning Models (ML), aim to understand the factors around a context and convert a scenario into a class (classification models) or a number (regression models [7]). Machine Learning models can predict, for example, if a patient presents a risk of worsening case and learn from this prediction. However, depending on the application context and the data used, the accuracy can be impaired. Problems such as noise and missing values are factors that directly influence the performance of a model. For instance, some approaches try to minimize possible error and over-fitting of the model by using ensemble learning models [8]. Ensemble learning is a Machine Learning approach that solves a problem by training multiple models. Unlike ordinary approaches in which a single hypothesis is learned from training data, ensemble approaches attempt to build a set of hypotheses and combine them to build a new hypothesis [9]. These approaches combine various learning algorithms to produce a result that is more adherent and precise [9]. Previous studies show that an ensemble model often performs better than the individual learners, as base learners [10].

The primary research question of the study is as follows: Is it possible to use an ensemble Machine Learning strategy to identify postoperative patients of orthopedic surgeries that need to go to the hospital/clinic to receive in-person care, after evaluating their reported data from the on-line questionnaire? The proposed solution is a prediction system to provide a remote assessment of postoperative orthopedic wounds. It uses Artificial Intelligent models to process patients' data related to postoperative wound healing and make predictions if the patient requires an in-person assessment. The eWound-PRIOR combines multiple models through an ensemble approach to achieve the most accurate prediction. The prediction results allow us to identify patients who need prioritized attention and notify both patients and physicians of the decision. The eWound-PRIOR framework was implemented through a mobile application to collect and transmit data. The use of such approach can bring benefits to patients and physicians through early diagnosis of disease and identification of at-risk patients, thereby reducing the flow of unnecessary patient to clinics or hospitals.

The main goals of this paper are to: develop the eWound-PRIOR that uses Machine Learning models combined in an ensemble approach to classify and predict postoperative cases requiring follow-up, and evaluate the solution with real patients' data from the orthopedic sector of a Hospital. This paper is organized as follows: Section 2 introduces the foundation of technologies used in this research. Section 3 presents the related work applied to ensemble Machine Learning models for healthcare. Section 4 describes the proposed framework and how we combined the ML models. Section 5 describes the evaluation process: planning, execution, results and, discussion analysis. Finally, Section 6 presents the conclusion and future research directions.

2 Background

This section presents an overview of the main concepts and definitions of the ensemble learning approach used in this research, as the Machine Learning algorithms, the autoencoder, and the metrics used to compare the accuracy of the ensemble classifications with individual learners. The use of Machine Learning ensemble approaches is widely explored in specific applications. Confidence estimation, feature selection, addressing missing features, incremental learning from sequential data, and imbalanced data problems, are examples of the applicability of ensemble systems. The original goal of using ensemble-based decision systems is to improve the assertiveness of a decision. By weighing various opinions and combining them through some smart process to reach a final decision can reduce the variance of results [19].

The eWound-PRIOR framework is composed of the most adherent set of models for the classification problem. We identified the classic models from the literature and the empirical tests. The ones we will use in this research are: (i) *Decision Tree (DT)*: is an abstract structure that is characterized by a tree, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent classes or class distributions [16]. (ii) *K-Nearest Neighbors (KNN)*: is a classical and lazy learner model, based on learning by analogy, which measures the distance between a given test tuple with training tuples and compares if they are similar [16]. (iii) *Random Forest (RF)*: is a classifier consisting of a collection of decision trees that classify

sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting [17]. (i) *Multi-Layer Perceptron (MLP)*: is represented by a neural network that learns the pattern of each classification through the data by calibrating the weights in each layer. The model can learn a nonlinear function approximator for either classification or regression [18].

An autoencoder algorithm [15] is part of a special family of dimensionality reduction methods, implemented using artificial neural networks. It aims to learn a compressed representation for input through minimizing its reconstruction error [8]. During the training step, the input dataset is compressed by the encoder and then, the decoder reconstructs the data by minimizing the reconstruction error. The ability to learn from the "intrinsic data structure" is useful when the available data have noise and/or too many missing values. In this study, the ReLU was used as the autoencoder.

The evaluation metrics analyze the accuracy and highlight the sensitivity or true positive rate (TPR) and the specificity or true negative rate (TNR). TPR measures a model capacity to identify cases that need in-person attention. The TNR measures a model capacity to identify cases that do not need in-person attention. The sensitivity and specificity were evaluated using Eqs. (1) and (2), respectively. Where true positive (TP) is the number of cases that need attention that is correctly identified as in-person consultation is necessary. True negative (TN) is the number of cases that do not need in-person care that is correctly identified as it does not need special care. P is the total number of positive instances and N is the total number of negative instances.

$$Sensitivity = \frac{TP}{P} \tag{1}$$

$$Specificity = \frac{TN}{N} \tag{2}$$

3   Related Work

This paper highlights the use of predictive models, based on ensemble Machine Learning techniques in the context of healthcare. In the literature, some researchers applied those predictive models in recommender systems aiming to detect chronic diseases [11], detect the possibility of heart disease [6, 12, 13] and estimate the likelihood that an adverse event was present in postoperative cases [4, 14]. To the best of our knowledge, no previous study has explored ensemble approaches to predict the likelihood of postoperative wound infection to notify patients and physicians whether an in-person assessment is required, using a mobile device.

Mustaqeem et. al. [6] proposed a hybrid prediction model for a recommender system, which detects and classifies subtypes of heart disease. From these classifications it makes recommendations. The work performs a probabilistic analysis and generates the recommendation from the case confirmation. Rallapalli and Gondkar [11] proposed the usage of an ensemble model, which includes a deep learning model to predict the exact attributes required to assess the patient's risk of having a chronic illness. For the authors, the ensemble approach results are superior to individual learners. Another study that

addresses the prediction of heart disease is presented in [12]. The proposed work tries to maximize the accuracy of the predictions through the application of an ensemble model. Their approach combines three different learning algorithms thought majority voting and presents the outperformance of the ensemble approach for their dataset. Tuli et al. [13] proposed the HealthFog framework for integrating ensemble deep learning in edge computing devices and for a real-life application of automatic heart disease analysis. The framework delivers healthcare as a fog service using IoT devices for efficiently managing the data of heart patients. The proposed architecture can provide recommendations based on the extracted data. The results pointed out that the use of deep learning in the continuous flow of data and combined in an ensemble form gets a significant improvement of prediction problems. Predictions in postoperative cases are addressed by some works. We highlight Zhang et al. [14] who focused on predicting complications in postoperative pediatric cataract patients by using data mining. The relationship between attributes that can contribute to complications was identified. The results point out that complications can be predicted and except for age and gender other attributes such as position, density, and area of cataracts are related to complications. Also, Jeffery [4] developed a patient-reported e-Visit questionnaire, in two- and six-week cases following orthopedic surgery. The author used the data collected to build a statistical model using logistic regression to estimate the likelihood that an adverse event was present. A notification is made as to whether the patient should be seen by the surgeon in-person. In the study, among the two weeks patients, only 24.3% of patients needed the appointment. For patients who returned for an in-person follow-up six-weeks postoperative, only 31.6% of patients needed the appointment. This questionnaire was the base for our mobile app.

A comparison is presented to identify the similarities and differences between our work and the selected ones. The criteria are studies about healthcare and prediction (C1); the use of ensemble predictive models or Machine Learning models in e-health (C2); evaluation with real patients' data (C3); and predicting the likelihood of an adverse event in postoperative cases (C4). Table 1 presents the comparative analysis. Our proposal meets the comparative criteria as it focuses on early diagnosis of worsening cases in postoperative orthopedic recovery and predicts and notifies patients to receive prioritized care. The framework is based on ensemble Machine Learning models and our case study is with patients from the orthopedic sector of a Hospital.

Table 1. Comparative analysis.

Criteria	[6]	[4]	[14]	[12]	[11]	[13]	eWound-PRIOR
				Related Works			
C1	●	◑	●	●	●	●	●
C2	◑	○	○	●	◑	●	●
C3	●	●	●	●	●	●	●
C4	◑	●	◑	○	○	○	●

Labels
●Fully meets   ○Does not meet   ◑Partially meets

4   eWound-PRIOR Framework

The main goal of this research is to develop an ensemble framework to predict the
likelihood of suffering an adverse event, of wound healing, following elected orthopedic
surgery on the knee, and identify patients who require an in-person assessment. The
application aims to reduce the flow of patients in hospitals and clinics requiring wound
checks and to reassure patients throughout their wound healing. The eWound-PRIOR is
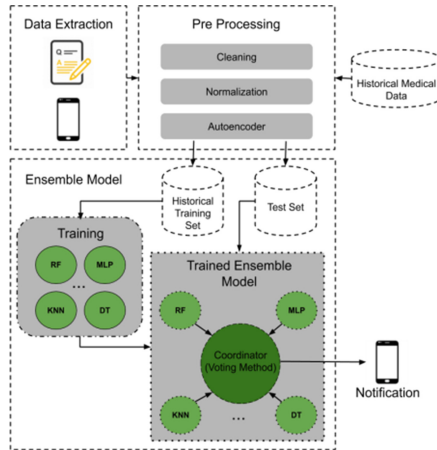illustrated in Fig. 1 and the main steps are described next.

Fig. 1. eWound-PRIOR framework.

Data Extraction - The first step refers to the data extraction process. It considers all
available data and extracts the information considered important for the predictions, e.g.
symptoms and medication. The information can be accessed and captured by medical
notes and electronic forms. This process extracts the historical medical information
(HMD) provided by the hospital database for training purposes and collects the real-time
patient data from a mobile application for prediction.

Pre-processing - This step pre-processes the patients' data and historical medical data.
The process removes noise values and replaces the missing ones. It also normalizes the
data to prepare the dataset for training the ensemble model. To synthesize the amount of
information, an autoencoder is trained with the normalized data to make data dimension-
ality reduction. The autoencoder is tuned until the Mean Square Error (MSE) reaches
the minimum.

Ensemble Model - The ensemble model represents the eWound-PRIOR core by exe-
cuting the pre-processed data captured in the first step. It is responsible for predicting
whether the patient needs to go to the hospital/clinic for an in-person appointment. The
eWound-PRIOR framework model was designed to be composed of ensemble Machine
Learning models. To synchronize and combine each model with the Voting Method [19],
we use a coordinator service to manage each prediction result. The coordinator com-
bines several classic ML models, implemented as software autonomous services, with

reactivity, intelligence, and social features. As autonomous services, each model is able of handling with requests and predicting cases requiring follow-up. In this research, we use the KNN, DT, MLP, and RF ML models. The models are supervised because it is a classification problem, which consists of indicating if a patient had or not a worsening in his/her treatment. Dealing with postoperative patients eWound-PRIOR ensemble seeks accurate results with a higher certainty for the predictions.

The Ensemble Strategy is based on the voting method [19] and we use the weighted average of the models' results (probability for a certain classification) to define the final classification for a patient's context. We set the weight of the "Wound healing well" classification as one and the "Wound requires care" classification as two. This allows focusing on patients that need special care/attention. The return value is an object with the final classification and its intensity. In the case of classification as "Wound requires care", the patient is notified to go to the hospital/clinic where the doctor can assist him/her. The hospital/clinic administration is also notified to follow-up with this patient via a patient dashboard. The strategy is described in Algorithm 1.

Data: An array A of objects (Classification, intensity);
Result: An object (Classification, intensity)
$NegSum \leftarrow 0$
$PosSum \leftarrow 0$
$WeightSum \leftarrow 0$
for $i \leftarrow 0$ **to** $A.length$ **do**
    $Prob \leftarrow A[i].intensity$
    $Class \leftarrow A[i].classication$
    if $Class == attention$ **then**
        $NegSum \leftarrow NegSum + (1 - Prob)$
        $PosSum \leftarrow PosSum + 2 * Prob$
    else
        $PosSum \leftarrow PosSum + 2 * (1 - Prob)$
        $NegSum \leftarrow NegSum + Prob$
    end
    $WeightSum \leftarrow WeightSum + 1$
end
$MediaPos \leftarrow \frac{PosSum}{2*WeightSum}$
$MediaNeg \leftarrow \frac{NegSum}{WeightSum}$
if $MediaNeg \leq MediaPos$ **then**
    return (attention, MediaPos);
else
    return (not attention, MediaNeg);
end

Algorithm 1: Ensemble combination

Notification - The last step is responsible for notifying the patients about their health. If the patient is deemed a priority case requiring attention, the patient and the hospital/clinic administration are notified to arrange an in-person follow-up appointment. This allows medical facilities to get better management of people by prioritizing emergency cases.

5   Evaluation

To evaluate the eWound-PRIOR framework, we had a partnership with Western University, which allowed access to the study proposed by Jeffery et al. [4] and part of the patients' dataset. In this context, the research was evaluated using real-world data. Considering our context, the case study main steps are: planning, execution, and results.

Planning - We first designed the eWound-PRIOR workflow with the main activities to illustrate the research (Fig. 2). Jeffery et al. [4] study proposed the e-Visit questionnaire,

which must be answered on-line by the patient two and six weeks after the surgery. The questions are about the patient's health after the surgery, the prescribed medication, and the symptoms in their daily life. It has a Demographic Information and eWound forms to be filled by the patient, one Risk Form, and the Surgeon's Data Form.
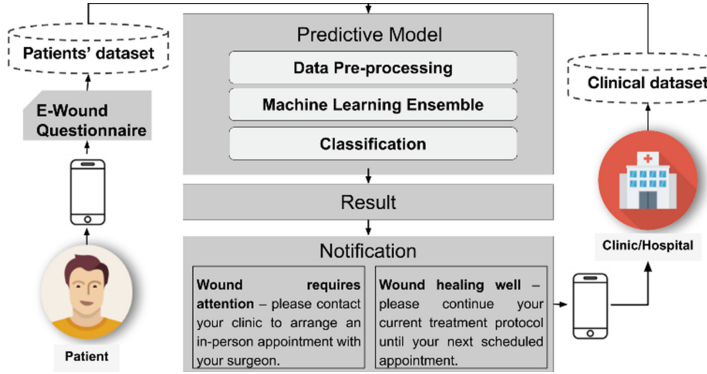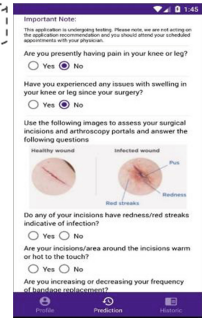

Fig. 2. eWound-PRIOR Workflow.

Fig. 3. eWound APP.

Based on the eWound questionnaire, we developed a mobile app (Fig. 3). After the patient completes the questionnaire, the answers are stored in a database, which is available to the eWound-PRIOR framework process.

Execution - The eWound questionnaire captures data about the patient's clinical signs, and symptoms, pain, prescribed medications and how they self-assess their wound healing at two- and six-week postoperative. The dataset had a total of 352 patients who completed the questionnaire. The answers related to wound healing were captured. The dataset is divided into two subsets of data, one related to patients that did the surgery in the past two and six weeks. This shows that two different trained ensemble models are needed specifically for each period time patients have an appointment.

After the extraction process, the data were pre-processed by removing the duplicated answers and inconsistent cases. The missing values were filled by the average feature values. Preprocessing the datasets is important because they have too many missing and noise values. Figure 4 shows the pre-processing main activities. As the e-Visit questionnaire has a skip logic, we treated the sub-questions by setting their values to -1, where the main question has the no as value. The questionnaire skip logic implies that if the sub-questions were answered that usually means the patients answered yes to the main question. Whereas if they did not answer all of them, it is because they answered no to the main question. In this case, -1 indicates that the sub-questions were not answered, meaning that the patient did not feel anything related to the main question.

We normalized the cleaned data by using the Normalizer provided by the Sklearn framework, and we trained an autoencoder structured with 10 hidden layers. ReLU was used as the activation function for all layers. After the autoencoder training, the Mean Square Error found was 0.003 for both datasets. This value may be considered a good

result for an autoencoder [20]. We performed empirical tests to set the resulting number of columns (questions) of the dataset using the autoencoder, ranging from 10 to 46. In our experiments, 20 columns showed a better performance for the available datasets. Therefore, the cleaned data were encoded, reducing the number of columns from 46 to 20. The classifications remained untouched after the process, with 1 and 0 values for "Wound requires care" and "Wound healing well" classification. All questions were used in this process, which was transformed from 46 columns in the CSV to 20 columns. By encoding the original columns to extract the intrinsic information in the dataset, it was possible to improve the accuracy of the classifications.

Aiming to obtain the average of predictions, instead of selecting the best set of training instances, we applied the SMOTEENN, a combination of Edited Nearest Neighbors (ENN) with Synthetic Minority Over-sampling Technique (SMOTE), an over-sample technique to multiply the minority instances to balance the classes for training. We randomized the selection of cases from the preprocessed datasets, splitting the cases into train and test sets. We randomly selected 80% of each dataset for training and 20% for testing. The process was executed 100 times.

We selected the most adherent parameters based on the preprocessed dataset using the Grid Search CV function from the Sklearn framework [21]. The parameterization process was carried out by setting cross-validation with 10 folds based on the (Stratified) KFold [21]. We also used the Sklearn framework as a provider for each model. Table 2 shows the model's parameterization. After the parametrization and pre-processing process, the two sets of models were sent to the coordinator service. As in the previous study, the results were combined with the voting ensemble method. The experiments were done in a server with Ubuntu 18.4 LTS; 94 GB RAM; Intel Xeon CPU E5–2630. The results are discussed next.
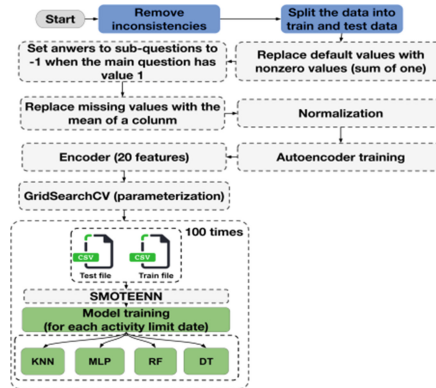

Fig. 4. Pre-processing step

Experimental Results and Discussion - We applied the same process for training and testing the two- and six-week datasets, which generated two sets of models. Each set was used for predicting the necessity of care for patients according to the number of days that had passed since surgery. Table 3 shows the results for the sensitivity (sens.)

Table 2. Model's parametrization

Model	Parametrization	
	2 weeks dataset	6 weeks dataset
KNN	Algorithm: auto Metric: *Euclidian* K: 2	Algorithm: auto Metric: *Euclidian* K: 3
Decision tree	Criterion: *entropy* Max features: 6 Splitter: *best* Max depth: 8	Criterion: *gini* Max features: 10 Splitter: *random* Max depth: *none*
MLP	# Hidden layers: 50 Activation func: *relu* # Input: 20 # Output: 1 Batch size: 29 Learning rate: *constant* Warm start: true Solver: *lbfgs*	# Hidden layers: 50 Activation func: *identity* # Input: 20 # Output: 1 Batch size: 12 Learning rate: *adaptative* Warm start: true Solver: *lbfgs*
Random forest	Criterion: *gini* # Estimators: 11 Warm start: true Max features: 6	Criterion: *entropy* # Estimators: 76 Warm start: true Max features: 9

and specificity (spec.) metrics for each model for the two- and six-week datasets. To understand how precise the results were from the outliers, we have calculated the standard deviation (STD) for each metric.

Sensitivity and specificity exist in a state of equilibrium [23]. The ability to correctly identify people who need special attention (sensitivity) usually causes a reduction in specificity (meaning more false positives). Likewise, high specificity generally implies a lower sensitivity (more false negatives). Still, high sensitivity is clearly important where the test is used to identify a severe but treatable disease. Although the ensemble presents the lowest value in the standard deviation compared to the individual models, there is a significant variation of the results for each model. As the datasets were provided with patients' answers, many missing values were present in the answers of each case. The models presented difficulty to correctly classify the cases once the data treatment was carried out by replacing the missing answers. The missing values replacement did not differentiate very well each case because various cases got no value in the sub-answers, and the value for replacement was unique for all.

Several classifications were only 60% certain in the predictions for all models, even using the autoencoder, which impairs the results of the ensemble. As we used different weights for each classification, the sensitivity got a good result for both datasets. However, the specificity was impaired by the several uncertain classifications of each model. For the missed cases requiring care, incorrectly classified as "Wound healing well", we

Table 3. Experiment results

Model	Two-week dataset		Six-week dataset	
	Sens. (STD)	Spec. (STD)	Sens. (STD)	Spec. (STD)
eWound-PRIOR	0.730 (0.12)	0.428 (0.08)	0.728 (0.11)	0.339 (0.08)
KNN	0.497 (0.14)	0.650 (0.08)	0.561 (0.14)	0.465 (0.12)
DT	0.529 (0.17)	0.582 (0.09)	0.524 (0.13)	0.527 (0.11)
MLP	0.741 (0.20)	0.344 (0.23)	0.666 (0.14)	0.507 (0.15)
RF	0.526 (0.16)	0.591 (0.09)	0.521 (0.13)	0.533 (0.09)

found that most patients answered the questions related to the body temperature and the wound as not presenting any problem. This causes several columns (sub-answers) in the dataset to have a replacement, which generates unique values in some columns of the encoded dataset after using the autoencoder.

Analysis of each set of questions revealed problems in the model classification. For example, cases with "red streaks" or an incision that is "hot to the touch" showed signs that the patient may be at risk of infection and need physician care. However, our model incorrectly classified these patients as "Wound healing well" due to the noise generated from responses to other questions. Although we have preprocessed the data, this study is limited by the noise and missing data in the datasets. As we can see from the results, having too many replacements on missing and implicit values in the dataset is not ideal, and may generate an incorrect classification. Without the skip logic we would capture more information related to the patients' health and reduce the number of replacements, consequently reducing the preprocessing step to prepare the data for training.

From this case study, we must highlight two genuine lessons learned. The first deals with a health perspective. The skip logic questionnaire is easier for patients as it is not necessary to answer the questions that do not apply. Originally, the questionnaire was created with no obligation to answer all the sub-questions. In this way, patients often skip a question that could be important for the classifications. Machine Learning models consider all columns in a dataset to pursue predictions. In our case, as many cases got negative classification (wound healing well - do not need to be seen), the pre-processing step had to replace the implicit and missing values, impairing the models' training and performance. The second lesson comes from our previous experience in using ensemble strategies in healthcare and other domains. In general, they are not affected using different Machine Learning models, if individually they show good accuracy. The final ensemble classifications are, in general, better than those provided by individual models [22]. eWound-PRIOR framework, with multiple classifications as a predictive basis, together with the human expertise and other resources added to the questionnaire such as images will guarantee greater autonomy and assertiveness.

6 Conclusion and Future Work

Our research proposes the eWound-PRIOR framework for prioritization of postoperative cases, an ensemble model as the predictive core. The framework applies autonomous services and Machine Leaning models capable of cooperating and aggregating results to maximize the accuracy and certainty of the predictions. We developed a mobile application to collect postoperative patient's health via an on-line questionnaire. We can answer our research question: it is possible to identify postoperative patients of orthopedic surgeries that need to go to the hospital/clinic to receive in-person care, after evaluating their reported data from an on-line questionnaire using an ensemble Machine Learning strategy. eWound-PRIOR ensemble made correct predictions for 50% (general accuracy) of cases for the two- and six-week datasets. It demonstrates good sensitivity and poor specificity. We compared the results of each Machine Learning model separately and together in an ensemble strategy. The prediction was impaired by the number of missing values in the dataset. The constrains came from the questionnaire skip questions, which gave the patients the option of not answering most of them. Future research needs to improve the models' parameters with new cases and a larger dataset and to improve its effectiveness to correctly identify patients requiring in-person follow-up. Furthermore, the eWound application interface should be evaluated to ensure the patient's satisfaction. We also recommend the application be tested using a diagnostic validity study design to compare its predictions to the current gold standard of in-person patient assessment.

References

1. Marsh, J., Hoch, J.S., Bryant, D., MacDonald, S.J., Naudie, D., McCalden, R., Howard, J., Bourne, R., McAuley, J.: Economic evaluation of web-based compared with in-person follow-up after total joint arthroplasty. JBJS **96**(22), 1910–1916 (2014)
2. Salvati, E., Robinson, R., Zeno, S., Koslin, B., Brause, B., Wilson, J.P.: Infection rates after 3175 total hip and total knee replacements performed with and without a horizontal unidirectional filtered air-flow system. J. Bone Joint Surg. Am. **64**(4), 525–535 (1982)
3. Wildner, M., Peters, A., Hellich, J., Reichelt, A.: Complications of high tibial osteotomy and internal fixation with staples. Arch. Orthop. Trauma Surg. **111**(4), 210–212 (1992)
4. Jeffery, W.G.: e-visits for early post-operative visits following orthopaedic surgery can they add efficiency without sacrificing effectiveness. Electronic Thesis and Dissertation Repository, vol. 5053 (2017). https://ir.lib.uwo.ca/etd/5053
5. Ali, F., Islam, S.R., Kwak, D., Khan, P., Ullah, N., Yoo, S.J., Kwak, K.: Type-2 fuzzy ontology–aided recommendation systems for IoT–based healthcare. Comput. Commun. **119**, 138–155 (2018)
6. Mustaqeem, A., Anwar, S.M., Khan, A.R., Majid, M.: A statistical analysis based recommender model for heart disease patients. Int. J. Med. Inf. **108**, 134–145 (2017)
7. Dreiseitl, S., Ohno-Machado, L.: Logistic regression and artificial neural network classification models: a methodology review. J. Biomed. Inform. **35**(5–6), 352–359 (2002)

8. Araya, D.B., Grolinger, K., ElYamany, H.F., Capretz, M.A., Bitsuamlak, G.: An ensemble learning framework for anomaly detection in building energy consumption. Energy Build. **144**, 191–206 (2017)

9. Zhou, Z.H.: Ensemble Learning, pp. 411–416. Boston. Springer, Heidelberg (2015)

10. Opitz, D., Maclin, R.: Popular ensemble methods: an empirical study. J. Artif. Intell. Res. **11**, 169–198 (1999)

11. Rallapalli, S., Gondkar, R.: Big data ensemble clinical prediction for healthcare data by using deep learning model. Int. J. Big Data Intell. **5**(4), 258–269 (2018)

12. Kurian, R.A., Lakshmi, K.: An ensemble classifier for the prediction of heart disease. Int. J. Sci. Res. Comput. Sci. **3**(6), 25–31 (2018)

13. Tuli, S., Basumatary, N., Gill, S.S., Kahani, M., Arya, R.C., Wander, G.S., Buyya, R.: "Health-fog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated IoT and fog computing environments. Fut. Gener. Comput. Syst. **104**, 187–200 (2020)

14. Zhang, K., Liu, X., Jiang, J., Li, W., Wang, S., Liu, L., Zhou, X., Wang, L.: Prediction of postoperative complications of pediatric cataract patients using data mining. J. Transl. Med. **17**(1), 2 (2019)

15. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science, Technical Report (1985)

16. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier, Amsterdam (2011)

17. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)

18. McClelland, J.L., Rumelhart, D.E., Group, P.R., et al.: Parallel Distributed Processing, vol. 2. MIT press, Cambridge (1987)

19. Polikar, R.: Ensemble Learning, pp. 1–34. Springer, Heidelberg (2012)

20. Tan, C.C., Eswaran, C.: Using autoencoders for mammogram compression. J. Med. Syst. **35**(1), 49–58 (2011)

21. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**(Oct), 2825–2830 (2011)

22. Braz, F., Campos, F., Stroele, V., Dantas, M.: An early warning model for school dropout: a case study in e-learning class. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), vol. 30, no. 1, p. 1441 (2019)

23. Lalkhen, A.G., McCluskey, A.: Clinical tests: sensitivity and specificity. Contin. Educ. Anaesth. Crit. Care Pain **8**(6), 221–223 (2008). https://doi.org/10.1093/bjaceaccp/mkn041