Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 239 (2024) 1532–1538

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS – International Conference on ENTERprise Information Systems / ProjMAN – International Conference on Project MANagement / HCist – International Conference on Health and Social Care Information Systems and Technologies 2023

Proof of concept: a practical approach to machine learning in video consultations' acceptance in higher education institutions.

Jorge Tavares[a]*, Filipe Viana Pereira[a]

[a]*NOVA IMS, Universidade Nova de Lisboa, Portugal*

Abstract

Video consultations and telemedicine play a key role in healthcare. It is crucial to assist institutions in identifying which patients are more likely to use them, increasing the adoption rate. Even though Artificial Intelligence models are important, certain cases require a human centric approach to inform patients and doctors, ensuring trust in the tools provided. This can be achieved through interpretable Machine Learning models and Statistical Analysis. In this paper we show that even though the accuracy of the Machine Learning models used (Decision Trees, Logistic Regression, and Random Forest) was between 84%-89% on the training set and 69%-76% on the test set the models lacked good ability to identify the users. Afterwards, additional statistical analysis evidenced that the main driver to use video consultations in a sample of participations with a higher-than-average education is the presence of one or more chronic diseases.

Keywords: Video consultations; Technology adoption; Covid-19; Machine Learning; Artificial Intelligence

* Corresponding author. Tel.: +41 76 608 64 45
E-mail address: tavares.jorge@gmail.com

10.1016/j.procs.2024.06.328

1. Introduction

Video consultations play a significant role for the future of healthcare by solving some of the upcoming healthcare challenges, as pointed by the European Commission in Europe and the National Academy of Medicine in the United States of America [1-2]. This technology can improve patient quality and enhance their access to healthcare [1]. During the COVID-19 pandemic, there was an increase in the use of telemedicine for urgent and non-urgent care visits [3]. During the Covid-19 pandemic there were significant restrictions to access overwhelmed healthcare systems, due to that, both patients and healthcare providers acknowledged significant benefits of using video consultations [4]. This increase in Video Consultations during the Covid-19 pandemic was sustained even after it's end, indicating the significant role they are currently playing [4-6].

Artificial Intelligence (AI) is a field of knowledge, which combines computer science and advanced statistics to support problem-solving [7]. AI can be divided in two sub-fields, machine learning (ML) and deep learning [7].

The concept of ML is based on employing computer algorithms capable of recognizing patterns and effectively training the machine for tasks such as prediction, recommendation, or finding patterns in data. [8].

Deep learning is a newer and more complex approach to AI, that uses deep neural networks. The neural network begins with an input layer that then progresses to a variable number of hidden layers [7]. As the algorithm of deep neural networks studies a problem using multiple layers, it can successively refine itself without the need for explicitly programmed directions [7].

It is a fact that by using deep learning the models usually achieve higher accuracy compared with ML [8]. However, when using ML it is possible to better understand which are the input variables that have a higher influence on the output variable [8]. This makes it more useful when we need to understand the reasons why something is happening and increases the transparency of the model. Therefore, with relatively small datasets and more exploratory studies, a machine learning approach can still help understand the reason behind a certain behaviour [7]. It is important to note that some socio-demographic factors affecting patients' adoption of mobile health tools have been identified [9]. The key barriers were lack of technical skills (digital literacy), lower income, lower education, and older age [9].

As the use of ML applications in healthcare continues to grow, there is a pressing need for the development of interpretable models to increase trust among doctors and patients, leading to an increase in their acceptance [10]. However, achieving interpretability is a complex task that must be tailored to the specific application [10]. Considering the specific data structure related to our study and the healthcare's interpretability needs, a ML approach best fulfils the goals of our study compared with deep learning. Predicting behaviors is equally important to understanding the underlying reasons behind it, in order to achieve informed decision-making and subsequent actions.

This study aims to understand the socio-demographic characteristics of video consultations users in Portugal during the COVID-19 pandemic by using a machine learning approach.

2. Methodology

2.1. Data Collection

The survey was distributed via e-mail in two institutes of higher education in Lisbon (students from all university degrees, teachers, and administrative personnel). The survey was answered on-line and asked socio-demographic questions about the user (age, education, gender, existence of chronic health condition). Additionally, it was also inquired if the person was a user or not of video-consultations over the past year. Before the participants could answer the survey, an introduction about the research purpose was provided. Anonymity and confidentiality assurance was stated to the participants. The participants had to state their agreement to participate in the survey. In total, 346 valid answers were obtained. It was possible to answer the survey in Qualtrics through mobile phones and computers. The survey was available between the 27th April 2021 and 2nd May 2021. The first pilot stage was collecting and reviewing 10 replies to ensure the survey was successful. These were excluded from this analysis.

2.2. Data Analysis

Google Colaboratory, or Colab for short was used to write python code for ML for this research study. Python version 3.10.11 and scikit-learn version 1.2.2 were used to train our data with the following ML Algorithms: Decision Tree (DecisionTreeClassifier), Logistic Regression (LogisticRegression), and Random Forest (RandomForestClassifier). Additional complementary statistical analysis was done with SciPy version 1.10.1. For data processing and visualization, the following packages were included: Pandas, NumPy, matplotlib, graphviz and seaborn.

3. Results

3.1. Descriptive results

Table 1. Sample characteristics

Characteristics	Sample (number)	Sample (percentage)
Age (mean = 34.8)		
[18-34]	224	64.74
[35-49]	22	6.36
[50-64]	77	22.25
>64	23	6.65
Gender		
Male	136	39.31
Female	210	60.69
Education		
High School completed	23	6.64
Bachelor's Degree (on-going or completed)	176	50.88
Master's Degree (on-going or completed)	145	41.90
Chronic disease		
Has one or more Chronic Disease	61	17.63
Does not have Chronic Disease	285	82.37
Video consultation		
User	68	19.65
Non-User	278	80.35

From the participants that answered our survey, 19.7% used video consultations. Literature concerning Portugal identifies the use of video consultations before Covid as unusual [11]. Guideline recommendations in Portugal advocate that implementation strategies for the use of video consultations should not aim older populations (>65 years old) and aim for people with higher digital literacy [11]. By doing this we can provide a useful service to this target group and make the healthcare system more available to the older people and with less digital literacy skills [11]. This was the reason for our sampling strategy concerning the target population for this study.

3.2. Multivariate results – Machine learning approach

We used a supervised learning (classifier approach) to identify the users and non-users of video consultations and the relevant parameters that are associated to the class classification. A total valid 346 questionnaires were collected and processed. Age will be analysed as a continuous variable, whereas all other variables were classified as nominal and dummy variables created. Due to the low number of people with PhD (n=2), this class was dropped from the analysis. For the ML analysis, we split 30% of our data for the test set. The following algorithms were implemented from scikit-learn in the following sequential order: Decision Tree (DecisionTreeClassifier), Logistic Regression (LogisticRegression) and Random Forest (RandomForestClassifier). Before running the Logistic Regression algorithm we implemented from sklearn.preprocessing the MinMaxScaler [8]. Ensembles are methods that combine multiple machine learning models to create more powerful and accurate models [8]. One of the most effective models according to the literature that belongs to this category is Random Forest [8]. Our analysis strategy compromises two traditional ML algorithms plus a more sophisticated ensemble model, from which we compare the results.

We initially applied the Decision Tree classier using a max tree depth =4, and kept as default the standard parameters from sklearn.tree – DecisionTreeClassifier. The accuracy of the training data set was 84.3% and the test data set = 72.1%. If the overall accuracy results seem to be reasonable [8], a simple analysis of the decision tree (figure 1), seems to point-out to a classification issue with the user's class. A good model should have a good precision as well as high recall. The F1 score combines both these aspects in one single metric [8]. Particularly more relevant is the F1 score when there is an imbalance between positive and negative cases, like with our research where we only have 1/5 of the sample as users [8]. The F1 score for the test data set for non-users was 0.84 and for users was only 0.06 (see table 2). There is clearly an issue to classify the users.

To attempt to improve the results of our Decision Tree classifier from sklearn.model_selection we used GridSearchCV, to finetune our model with the potential best parameters [8]. The best parameters from our grid search were: {'criterion': 'gini', 'max_depth': 2, 'min_impurity_decrease': 0.5, 'min_samples_leaf': 1}. The accuracy of the test data set increased from 72.1% to 75.0%. Nonetheless the F1 score for the users is at 0.00. The gain in accuracy resulted in the model incapacity to classify the users. A more detailed analysis revealed that the approach with the best F1 score (0.06) for users is using the Gini criterion and a maximum tree depth of 4. Without good classification results for the users with the Decision Tree algorithm the next step is to use the Logistic Regression.

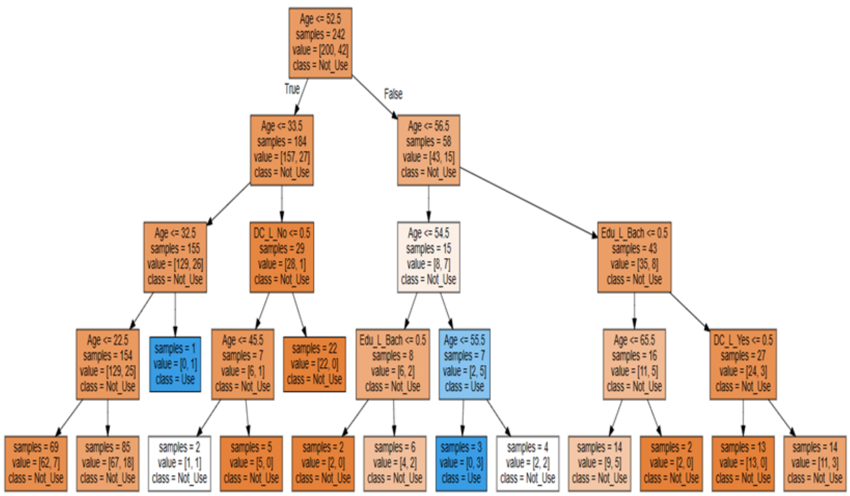

Fig. 1. DecisionTreeClassifier (max tree depth =4). Graphical output elaborated with graphviz package

Logistic Regression was used from sklearn.linear_model – LogisticRegression. The accuracy of the training data set was 82.0% and the test data set was 76.0%. The F1 score for the test data set for non-users was 0.86 and for users was 0.00 (see table 2). Although Logistic Regression showed a slightly higher accuracy with the test data set compared to the Decision Tree classifier, the F1 score for the users shows that the model is not effectively capable of classifying the users.

The Decision Tree classifier showed that it is a better option than the Logistic Regression to classify both classes, although with very poor results with the users. The Random Forest algorithm acts as an improved version of the Decision Tree algorithm [8], so we decided to use it to improve the study results. Random Forest algorithm was used from sklearn.ensemble- RandomForestClassifier. Using the default parameters of the random forest classifier, that usually work quite well [8], no accuracy improvement was obtained if more than 10 trees were used. The accuracy (for 10 trees) of the training data set was 89.3% and the test data set = 69.2%. The F1 score for the test data set for non-users was 0.81 and for users was 0.11 (see table 2). The Random Forest classifier was able to improve the classification of the users, still the F1 score is regarded as low for this class.

Table 2. Classification report for the 3 different classification algorithms used.

	Precision	Recall	F1-Score
	Decision Tree		
Non-Users	0.75	0.95	0.84
Users	0.20	0.04	0.06
	Logistic Regression		
Non-Users	0.76	1.00	0.86
Users	0.00	0.00	0.00
	Random Forest		
Non-Users	0.74	0.90	0.81
Users	0.20	0.08	0.11

To better understand which variables influence more the classification in each one of the algorithms, a visual representation of the features importance for each one is shown in figure 2. Particularly with Logistic Regression is advisable to drop one dummy variable to protect from the dummy variable trap [8].

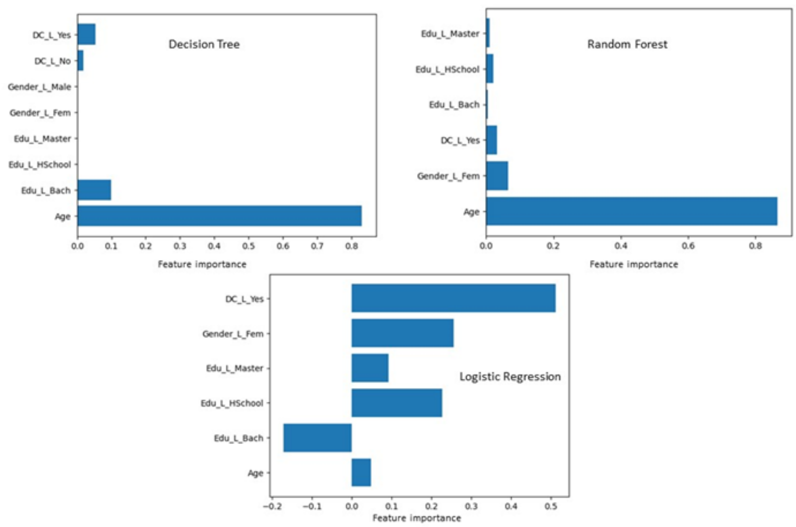

Fig. 2. (a) Decision Tree, (b) Random Forest, (c) Logistic Regression Feature Importance Equations

As a result of creating dummy variables, we obtained Gender_L_Male and Gender_L_Fem, respectively encoding males and females. DC_L_Yes and DC_L_No, for people with or without chronic disease. Education (Edu_L) was coded, for high school (HSchool), bachelor (Bach), master (Master) and Doctor (PhD).

The Decision Tree and Random Forest algorithms identified Age as being the feature with greatest importance. Conversely the Logistic Regression identifies DC_L_Yes as the feature with the greatest importance. Nonetheless none of the algorithms did a great work in classifying the users of video consultations.

As a next step, to better understand the socio-demographic factors that are associated with the use of video consultations, a more specific analysis between the variables in the data set and their direct interaction with the target variable (VC) was performed.

SciPy package was used to analyse the relation between the target variable and the nominal variables: gender, chronic disease, and education. The chi2_contingency test from SciPy was used. The Chi Square test was only statistically significant for chronic disease ($X^2(2)=6.196$; $p=0.045$) and an association between having chronic disease and using video consultations confirmed. Age was regarded as a feature with very relevant importance both in the Decision Tree and Random Forest algorithms. Nonetheless, the difference between the mean age from the participants that don't use video consultations (34.6 years) and use video consultations (35.7 years) is small. The observation of figure 3 shows a very asymmetrical age histogram, with a very high frequency of both users and non-users between 20 and 30 years old. Mann-Whithey U Test from SciPy was used to compare the ages of users and non-users of video consultations and, as expected, no statistically significant difference between the two groups was found (U=0.496; $p=0.62$).

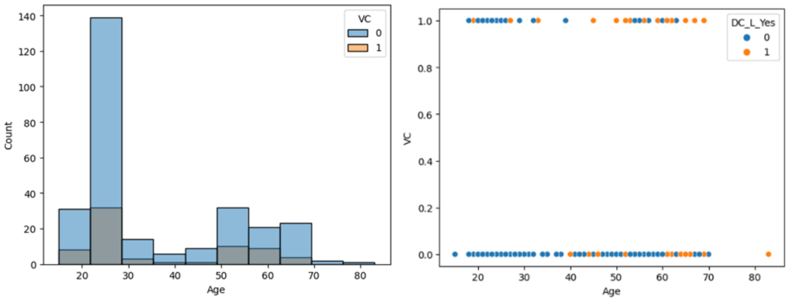

Fig. 3. (a) Age histogram. VC 0: non-users and VC 1: users of video consultations, (b) VC use taking into account Age and the existence of chronic disease (DC_L_Yes)

Additional exploratory analysis was conducted to understand if any further interaction between parameters could be associated with the use of video consultations (VC). The Decision Tree and Random Forest algorithm identified Age as the most relevant parameter and the Logistic Regression identified chronic disease. Figure 3(b) visualization suggests a potential interaction between them. Specifically, if video consultations increase more with age for people with chronic disease (DC_L_Yes). We used SciPy to test the interaction that was not statistically significant ($p=0.277$). Additional analysis using gender and education also rendered non-statistically significant interactions with age.

4. Conclusions

Approximately 20% of the participants that answered our survey mentioned that they used video consultations. During Covid-19 pandemic the usage increase of video-consultations enabled a better patient access to healthcare when the healthcare system was overwhelmed [3]. Understanding the socio-demographic characteristics of video consultations users may help to define specific strategies to promote usage for specific groups with less adoption. Not only for big and complex data sets ML algorithms are useful, in fact even for research in-progress projects, like our study, they can be helpful to obtain insights and refine the on-going study [8]. The usage of AI and ML in healthcare can raise ethical concerns [10]. It is important that the models used are self-explanatory so that the healthcare providers can understand the actions proposed and explain them to patients [12]. The models should be patient/human centric and should be able to predict not only if someone is a user or not, but which are the drivers of the outcome [10]. The overall accuracy of the 3 ML models was acceptable with an overall accuracy between 69-76% in the test data set. The main issue relates to the ability of the 3 models to classify the users. The Decision Tree and Random Forest algorithms identified age as being the feature with greatest importance. Conversely the Logistic Regression identified chronic disease as the feature with the greatest importance. Additional statistical analysis (Chi-Square test) proved an association of having chronic disease and using video consultations. Although age appeared to be a relevant feature for the Decision Tree and Random Forest algorithms, the asymmetrical distribution and high frequency of younger participants (20-30 years old) potentially hinders the algorithm's ability to effectively discriminate the users.

From our analysis it is suggested that future on-going research should include a more balanced age distributed sample, a specific question related to digital literacy [9] and attitudinal questions. Our analysis enabled us to understand that the main driver to use video consultations in a younger sample of participants with higher-than-average education is the presence of one or more chronic diseases.

References

[1] Arrogante, O. (2020.) "Increase in video consultations during the COVID19 pandemic: healthcare professionals' perceptions about their implementation and adequate management."*Int. J. Environ. Res. Public Health* **17** (14): 1–14. https://doi.org/ 10.3390/ijerph17145112.

[2] Dzau, V.J., Cohen, M., McGinnis, J.M. (2020) "Vital directions for health & health care." *N. C. Med. J* **81** (3): 167–172. https://doi.org/10.18043/ncm.81.3.167.

[3] Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 (2020). "Transforms health care through telemedicine: Evidence from the field." *J Am Med Inform Assoc* **27(**7):1132-1135. doi: 10.1093/jamia/ocaa072.

[4] Jiménez-Rodríguez D, Santillán García A, Montoro Robles J, Rodríguez Salvador MDM, Muñoz Ronda FJ, Arrogante O (2020) "Increase in Video Consultations During the COVID-19 Pandemic: Healthcare Professionals' Perceptions about Their Implementation and Adequate Management." *Int J Environ Res Public Health* **17**(14):5112. doi: 10.3390/ijerph17145112. PMID: 32679848; PMCID: PMC7400154.

[5] Ahmad F, Wysocki RW, White N, Richard M, Cohen MS, Simcock X (2021) "Telemedicine Use during the COVID-19 Pandemic: Results of an International Survey." *J Wrist Surg* **11(**4):367-374. doi: 10.1055/s-0041-1731820.

[6] Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, Lewis R, Iyer G, Macleod J, Horwood J. (2021) "Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study." *Br J Gen Pract*. **71**(704):e166-e177. doi: 10.3399/BJGP.2020.0948.

[7] Matthew Helm J., M. Swiergosz, A., S. Haeberle H., M. Karnuta, J., L Schaffer J., E. Krebs, V., I. Spitzer A., N. Ramkumar, P. (2020) "Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions." *Current Reviews in Musculoskeletal Medicine* **13**, 69–76 https://doi.org/10.1007/s12178-020-09600-8

[8] Müller, A. C., & Guido, S. (2018). Introduction to Machine Learning with Python (4th ed.). O'Reilly.

[9] Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. (2022) "Sociotechnical Factors Affecting Patients' Adoption of Mobile Health Tools: Systematic Literature Review and Narrative Synthesis" *JMIR Mhealth Uhealth* **10**(5):e36284, URL: https://mhealth.jmir.org/2022/5/e36284, DOI: 10.2196/3628

[10] Rasheed, K., Qayyum, A., Ghaly, M., Al-Fuqaha, A., Razi, A., & Qadir, J. (2022). "Explainable, trustworthy, and ethical machine learning for healthcare: A survey. In Computers in Biology and Medicine" **49**: p.06043). Elsevier BV. https://doi.org/10.1016/j.compbiomed.2022.106043

[11] Morais, A., Bugalho, A., Drummond , M., Ferreira, A., Oliveira, A., Sousa, S., Winck, J., & Cardoso, J. (2023). "Teleconsultation in respiratory medicine - A position paper of the Portuguese Pulmonology Society." *Pulmonology,* **29(**1): 65-76. https://doi.org/10.1016/j.pulmoe.2022.04.007

[12] Rudin, C. (2019) "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead" *In Nature Machine Intelligence* **1**(5): 206–215. Springer Science and Business Media LLC. https://doi.org/10.1038/s42256-019-0048-x