Article

Sentiment Analysis and Text Mining of Questionnaires to Support Telemonitoring Programs

Chiara Zucco [1] [iD], **Clarissa Paglia** [2], **Sonia Graziano** [3] [iD], **Sergio Bella** [2] [iD] **and Mario Cannataro** [1,4,*] [iD]

[1]   Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, 88100 Catanzaro, Italy; chiara.zucco@unicz.it

[2]   Unit of Cystic Fibrosis, Bambino Gesù Children's Hospital, 00165 Rome, Italy; clarissa.paglia@opbg.net (C.P.); sergio.bella@opbg.net (S.B.)

[3]   Unit of Clinical Psychology, Bambino Gesù Children's Hospital, 00165 Rome, Italy; graziano.sonia@gmail.com

[4]   Data Analytics Research Center, University Magna Graecia of Catanzaro, 88100 Catanzaro, Italy

[*]   Correspondence: cannataro@unicz.it

check for updates

Abstract: While several studies have shown how telemedicine and, in particular, home telemonitoring programs lead to an improvement in the patient's quality of life, a reduction in hospitalizations, and lower healthcare costs, different variables may affect telemonitoring effectiveness and purposes. In the present paper, an integrated software system, based on Sentiment Analysis and Text Mining, to deliver, collect, and analyze questionnaire responses in telemonitoring programs is presented. The system was designed to be a complement to home telemonitoring programs with the objective of investigating the paired relationship between opinions and the adherence scores of patients and their changes through time. The novel contributions of the system are: (i) the design and software prototype for the management of online questionnaires over time; and (ii) an analysis pipeline that leverages a sentiment polarity score by using it as a numerical feature for the integration and the evaluation of open-ended questions in clinical questionnaires. The software pipeline was initially validated with a case-study application to discuss the plausibility of the existence of a directed relationship between a score representing the opinion polarity of patients about telemedicine, and their adherence score, which measures how well patients follow the telehomecare program. In this case-study, 169 online surveys sent by 38 patients enrolled in a home telemonitoring program provided by the Cystic Fibrosis Unit at the "Bambino Gesù" Children's Hospital in Rome, Italy, were collected and analyzed. The experimental results show that, under a Granger-causality perspective, a predictive relationship may exist between the considered variables. If supported, these preliminary results may have many possible implications of practical relevance, for instance the early detection of poor adherence in patients to enable the application of personalized and targeted actions.

Keywords: text mining; sentiment analysis; Web-based questionnaire; telemedicine; telemonitoring; telehomecare

1. Introduction

Telemedicine can be defined as the set of health services providing medical care in patients' daily living environment, which is possible thanks to the support of information and telecommunication technologies [1].

Common goals of telemedicine programs are substantially threefold [2–4]:

- to increase self-management skills for patients whether they have a chronic condition or, for instance, during recovery or a rehabilitation phase after surgery or in the follow-up after a long time hospitalization and also during the treatment for depression and other mental health conditions;
- to improve patient surveillance and medication adherence by remote monitoring of their health status; and
- to reduce healthcare costs by optimizing the doctor's work by reducing the number of accesses to the emergency room and reducing the average of hospital admissions per patient.

A subfield of telemedicine is telehomecare, or home telemonitoring, which enables the rapid exchange of information between health systems and patients. Patients enrolled in a telehomecare program are provided with bio-monitoring devices and Internet reporting systems installed in their daily living environment. The devices can be used to autonomously measure vital signals, specifically related to the specific condition of the patient. These measurements are then transmitted and evaluated by health professionals (physicians and nurses) who will subsequently re-contact patients via phone call or message to check their symptoms and, eventually, provide an early medical response.

Telemedicine and, specifically, telehomecare systems have shown themselves to be cost-effective [5] and to provide an improvement in patients' quality-of-life, in terms of significant reduction of both mortality and length of stay of patients in progressive care unit [6], significant improvement of glycemic control for patients with diabetes [7], etc.

However, two variables that may affect a telemonitoring program's effectiveness are adherence levels and the degrees of drop-out. Adherence levels may be measured in different suitable ways. Here, adherence is intended as the rate of performed monitoring events with respect to the ideal number of events, suggested by the telemonitoring protocol, while the degrees of drop-out refer to the percentage of patients who abandon the telemedicine or the telehomecare program they were enrolled in, generally due to poor adherence. In particular, in [8], a systematic review among 37 healthcare programs for Heart Failure and Chronic Obstructive Pulmonary Disease was carried out. In the study, rates of refusal of almost one-third of patients were reported. Moreover, among patients who took part in the telehomecare program, one-fifth abandoned the program after enrollment.

Another interesting study is related to cystic fibrosis patients' follow-up at home [9].

Cystic fibrosis is the most common life-threatening genetic disease in the Caucasian population [10]. It is characterized by recurrent episodes of a respiratory infection that cause progressive lung deterioration, with a long-term decline in lung function. The spirometry test is a simple test used for lung function monitoring. It is known that continuous monitoring of lung function during the follow-up of patients with cystic fibrosis can reduce lung damage by preventing bronchopulmonary exacerbations and, consequently, prevent patient's exitus for lung insufficiency [11].

The authors reported that, of 39 enrolled patients, 15 dropped out of the program (38.46%). The percentage decreases to 31.4% if considering voluntary drop-out. Eighty-one percent of drop-out was due to poor program adherence [9].

The most frequently used approaches for conducting research studies in the social sciences make use of surveys [12]. Among the methods for collecting survey data, questionnaires represent a widely used tool. Thanks to the availability of tools and systems that facilitate the development and administration phases through online platforms, their popularity has grown significantly. Compared to face-to-face or telephone interviews and to questionnaires-on-paper, online questionnaires provide several advantages: (i) a cost reduction; (ii) the collection of a greater number of data in a shorter time; (iii) the possibility for the individual to manage the place and time to take the questionnaire; and (iv) responses are already digitized and exportable in formats which are suitable for a subsequent analysis [13].

Thus far, closed-ended questions in questionnaires have dominated the scene in the social sciences and, consequently, in the psychological and health sciences. This choice is justified by the easiness of data collection, reliability and simplicity of analysis, and the possibility of standardizing the collection to compare results between different populations [14]. On the other hand, the possibility of using open-ended questions would allow performing fine-grained analysis by offering new and interesting insights, capable of detecting slight differences, especially, for example, in the context of patient monitoring.

The investigations carried out to verify if a significant benefit may be obtained by introducing open-ended questions in questionnaires led to different results. In [15,16], it is shown that answers which received a high response rate in closed-ended questionnaires were not mentioned when the same question was formulated in an open-ended form, whereas the study conducted in [17] showed no benefit in introducing open-ended questions.

With the availability of many textual data coming from social platforms, noteworthy developments concerning the automated analysis of texts have been registered during the last decade. Above all, an increasing interest in the field of sentiment analysis, which aims at the automatic extraction of emotions and opinions, mainly from text [18], has been witnessed.

This work aims to present an integrated software architecture for the online provision and collection of questionnaires or surveys, which exploits a sentiment analysis-based approach to monitor patients' adherence to telehomecare programs. The idea is that the sentiment, i.e., the degrees of positiveness/negativeness, expressed by patients through their responses to questionnaires, may be related to their adherence and used to predict drop-out.

The present architecture proposal is intended as a contribution that can help the context of home telemonitoring programs. The basic idea is to integrate within a telehomecare system an online survey instrument to investigate the polarities of patients' opinions in relation to their experience.

The proposed system also encompasses a novel analysis approach that leverages lexicon-based sentiment analysis techniques and exploits the inferred polarity as a numerical feature to enhance further statistical or machine learning analysis.

To the best of our knowledge, no specific research has been published, nor has a system architecture been proposed that would explicitly monitor changes in patient's opinion across time through the repeated administration of a questionnaire, using the polarity associated with answers to open-ended question as a numerical feature, in a telehomecare system.

Additionally, the paper presents a case study application of the system architecture to discuss whether a predictive relationship, in terms of Granger-causality test modeling, may be assessed between patient adherence in a cystic fibrosis telehomecare program and their opinion about the program they are enrolled in.

The rest of the paper is organized as follows. Section 2 describes the methodology behind the proposed approach and the case-study application. Section 3 provides insights related to collected data and presents the Granger-causality hypothesis tests results and discusses it. Finally, Section 4 concludes the paper and outlines future works.

2. Materials and Methods

In this section, some preliminary information related to the case study, a description of the experimental protocol used and the analysis pipeline's proposal are presented.

2.1. Preliminary Information

Since 2001, a home telemonitoring program has been provided by the Cystic Fibrosis Center of the "Bambino Gesù" Pediatric Hospital in cystic fibrosis patients' follow-up. Patients are provided with Spirotel instrumentation from MIR (Medical International Research), which transmits data from the spirometry test and overnight pulse oximetry remotely, following the clinical workflow detailed in [11]. Patients are suggested to send spirometry transmission at least twice a week.

After the transmission, physicians contact patients by performing a telephone interview involving questions about some pulmonary symptoms and more general health conditions. Patients included in the telemonitoring program are treated with standard follow-up protocols, similar to those not enrolled in the program. A detailed description can be found in [19]. Despite the promising results, a significant percentage of abandonment for poor adherence has been constantly registered.

Table 1 reports some statistics related to patient enrollment, related to a nine-year period (2010–2018). As shown in Table 1, the drop-out patients enrolled in the telehomecare program represent 41% of the total number. Table 2 further illustrates the composition of patients who leave the program. In particular, 81.25% of patients' drop-out is due to voluntary drop-out: 50% of patient abandonment is related to poor adherence, while 31.25% of intentional abandonment is related to other reasons.

Table 1. Balance of enrolment during the period 2010–2018.

Patients	n.	Percentage
Enrolled	78	
Active	46	58.97
Drop-out	32	41.03

Table 2. Proportion of patients drop-out during the period 2010–2018, grouped by abandonment causes.

Drop-Out Patients	n.	% of Total	% of Dropout
poor adherence	16	20.51	50
died	6	7.69	18.75
other	10	12.82	31.25

2.2. Experimental Protocol and Dataset Description

The data analyzed in the present case study application were collected from the Cystic Fibrosis Unit, Bambino Gesù Children's Hospital, Rome, Italy. In this study, 169 online surveys sent by 38 cystic fibrosis patients (F/M = 20/18, age = 28.7 ± 9.91, age range = 14–49) recruited among patients already enrolled in a telemedicine program (years of enrollment = 5.9 ± 3.9) were collected and analyzed at five different survey epochs.

The enrollment criteria included patients more than 12 years old with cystic fibrosis who access the Cystic Fibrosis Unit in ordinary, daytime, or outpatient hospitalization. All patients who have undergone a transplant (liver/lung) were excluded from the study.

The study was formally approved by the local Medical Research Ethics Committee.

2.3. Administration of Questionnaire

From June 2019, 38 enrolled patients were asked to complete, every three months, an online questionnaire designed ad-hoc by the clinical team. In the following, each set of surveys submission is indicated as an epoch.

The Telemedicine Drop-Out (TDO) questionnaire consists of 15 blocks of closed, mixed, and open-ended items with yes/no constraints, and it was administered through a self-hosted web-based survey instrument built on top of LimeSurvey. The TDO survey was designed as an online, structured version of the interview led by the medical team within the telemedicine program, extended with a series of open-ended questions, whose objective was to infer polarity or, in perspective, to extract emotions from the relative answers [20]. The TDO questionnaire is reported in Appendix A.

To administer surveys to patients, LimeSurvey (https://www.LimeSurvey.org/) [21] is set up as a highly customizable, free, and responsive online survey tool. It also provides various API functions through the LimeSurvey RemoteControl 2 (LSRC2). The survey structure and the participants are created through the user interface provided by LimeSurvey. The collection of survey answers is automatized using the Python library Limepy that provides a Python wrapper for the LSRC2 API and the Python library Schedule to automatically update the responses. The DBMS server is MySQL.

As already stated, adherent patients need to transmit the results of the spirometry test at least twice a week. For each survey administration, i.e., survey epoch, the patient's adherence score (Adh-score) to the telemonitoring program was assessed as the total number of spirometry transmissions sent during a three-month window starting from the month before until the month subsequent to the survey administration, averaged by twice the total number of weeks following. More in details, suppose that a survey was carried at month t, then:

$$\text{Adh-score}_t = \frac{nS_{t-1} + nS_t + nS_{t+1}}{2(w_{t-1} + w_t + w_{t+1})}$$

where nS_{t-1}, nS_t, and nS_{t+1} refer to the number of spirometry transmissions sent in month $t-1$, t, and $t+1$, respectively, while w_{t-1}, w_t, and w_{t+1} refer to the number of weeks in months $t-1$, t, and $t+1$, respectively.

For instance, to calculate the Adh-score related to the first epoch submission, t = June 2019. Therefore, each patient's total number of spirometry transmissions from May 2019 to July 2019 was considered. Moreover, since the three-month window encompasses 13 weeks, the total number of spirometry transmissions was averaged by twenty-six.

By definition, patients who strictly follow medical advice have a related Adh-score ≥ 1. In the following, the percentage of Adh-score, i.e., Adh-score (%), is considered. Therefore,

$$0 \leq \text{Adh-score (\%)} = \text{Adh-score} * 100$$

and Adh-score (%) > 100 for patients with high rates of adherence. The clinical team provided the number of transmissions per month.
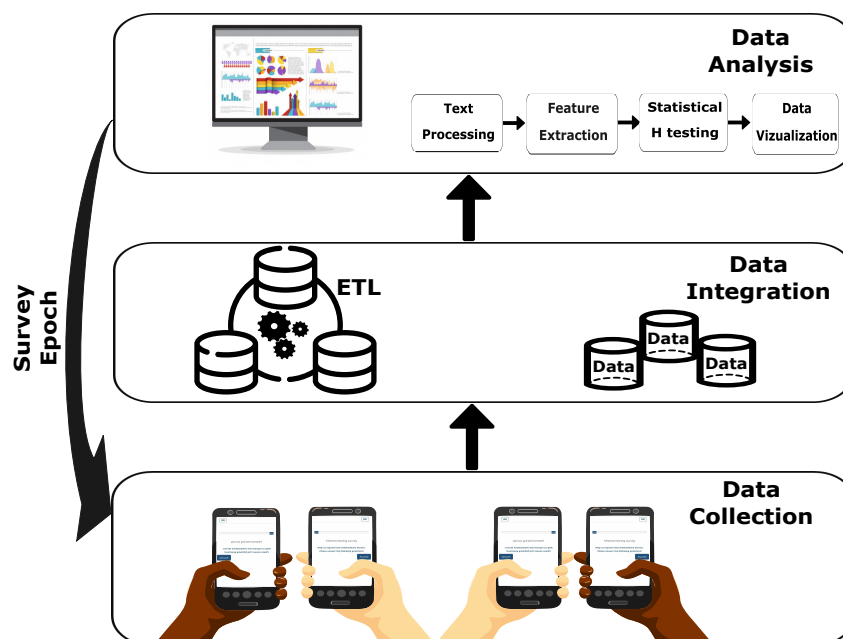
2.4. System Architecture

The system architecture encompasses three independent modules, connected in a cascade-fashion. In future works, the modules are supposed to be integrated using a unique user interface. Figure 1 shows the overall architecture for the system, which is organized as three logical levels:

- **Data collection:** Only completed survey data are automatically downloaded and cleaned to eliminate redundant information and duplicates. If the same respondent at the same administration epoch sent two surveys, the most recent one is considered.
- **Data integration:** Collected data are joined with adherence data according to a common schema. All the steps, typical of an ETL (Extraction, Transformation and Loading) approach, have been implemented through the Pandas library [22]. The integrated data table may be downloaded as a comma separate value (csv) file.
- **Data analysis:** Collected and integrated data encompass both structured information and free texts, resulting from patients' answers to the open-ended TDO questions. Therefore, this level implements different analysis pipelines, depending on the integrated data and analysis type.

2.4.1. Data Analysis Pipeline

The general pipeline for the analysis of textual data, i.e., answers to open-ended questions, involves:

- **Text preprocessing:** It includes standard NLP (Natural Language Processing) techniques, i.e., tokenization, stop word removal, and lemmatization. The preprocessing step was executed by using *SpaCy* (https://spacy.io/), a popular library for NLP in Python, which provides a set of preprocessing algorithms also for the Italian language.
- **Feature extraction:** To each open-ended free-text answer, a polarity score in the range $[-1, 1]$ was assigned through the VADER [23] lexicon-based method adapted to the Italian language and considered as a numerical feature.
- **Statistical hypothesis testing:** Data were sorted by respondents and survey submission date and, for each open-ended question, the sequence of assigned polarity was modeled as a time series, as the sequence of Adh-scores. Augmented Dickey–Fuller Test [24] was used to check for stationarity, while Granger-causality hypothesis test model [25] was examined to discuss the existence of directed causal interactions between the polarity score associated with free-text answers and adherence.
- **Data visualization:** To provide useful insights and summarize patient answers, different visualization techniques were used. In particular, preprocessed free-text answers are visualized through word clouds, while a graph shows the time-series of polarity scores at different submission epochs.

Figure 1. The modules of the system architecture, implemented as three independent levels connected to each other in cascade. The architecture is designed to be cyclical, as the system is used for each scheduled administration of the survey.

2.4.2. Sentiment Polarity Extraction

Valence Aware Dictionary for sEntiment Reasoning (VADER) [23] is a lexicon-based sentiment analysis engine that combines lexicon-based methods with a rule-based modeling consisting of five human validated rules.

The benefits of VADER's approach are: it does not require a training phase, and, consequently, its application is feasible even in low resource data domain; it works well on short text; it is fast and, therefore, may be suited for near real-time application; being related on general "parsimonious" rules, it is basically domain-agnostic; and it constructs a white box model, thus is highly interpretable and adaptable to different languages.

The starting point of the VADER system is a generalizable, valence-based, human-curated gold standard sentiment lexicon, built on top of three well-established lexicons, i.e., LIWC [26], General Inquirer, and ANEW [27], expanded with a set of lexical features commonly used in social media, which include emoji, for a total of 9000 English terms subsequently annotated in a $[-4, 4]$ range through the Amazon Mechanical Turk's crowd-sourcing service.

The VADER engine's second core step is the identification of some general grammatical and syntactic heuristics to identify semantic shifters, i.e., words that increase, decrease, or change the polarity orientation of another word. In particular, five heuristics for sentiment polarity shifters have been identified:

- Punctuation: The exclamation mark (!) is a valence booster, i.e., it increases sentiment intensity without affecting sentiment orientation.
- Capitalization: Uppercase words in the presence of lower cases words should be treated as a sentiment intensifier without affecting sentiment orientation.
- Degree modifiers: Nouns, adjectives, and adverbs, as well as idioms, are known as intensifiers or down-toning, which impact sentiment intensity by increasing or decreasing it.
- Contrastive particles: The "but" conjunction between two sentences shifts sentiment polarity in favor of the second part of the text.
- Negation: Negations reverse the polarity orientation of the lexical particles they are referred to. The investigation of Trigram preceding sentiment-laden terms enables the identification of negations for that specific term.

To extend VADER to the Italian language, Sentix [28], a lexicon that automatically extends the SentiWordNet annotation to the Italian synsets provided in MultiWordNet [29], was considered.

Among the five heuristics designed in VADER, only three needed to be adapted to the Italian language since the shifter role of capitalization of words and exclamation marks is used as intensifiers for both languages. Words belonging to the VADER set of negation words were translated in the Italian language, and the set was then extended by retrieving MultiWordNet synset terms for each word, while contrastive particle "but" was simply translated to Italian.

Among the intensifier sets, VADER also considered a few idioms, but, due to discrepancies across different languages, idioms were not considered.

2.4.3. Granger-Causality Testing

Granger-causality is a statistical hypothesis testing model to determine if there is a directed relationship between two time series [25]. A time series X is said to Granger-cause Y if it can be shown that there is a statistically significant improvement in predicting future values of Y by using past values of X (i.e., lagged values of X) and Y, compared to predictions based only on past values of Y.

The possibility to relate past values of X to Y's actual values is in virtue of a lag factor. Here, the Granger-causality test was computed for X's lagged values. All the lags ranging from one to four were tested, where four is the number of considered submission epochs minus one.

Here, the considered alternative hypothesis is that the polarity-score time series associated with each considered open-ended question Granger-cause the time series of adherence. The level of significance was set at 5%, i.e., p-value < 0.05. The Granger-causality test assumes the hypothesis that the investigated time-series are stationary. Therefore, the augmented Dickey–Fuller method was exploited to check stationarity conditions [24].

3. Results and Discussion

In this section, we present the results related to the Granger-causality testing model to assess the plausibility of the existence of a predictive relationship between a score representing the opinion polarity of patients about telemedicine and their Adh-score. Moreover, to gain useful insights about the collected data, a preliminary exploratory data analysis was performed by following the pipeline discussed in the previous section and by summarizing data through suitable visualization.

3.1. Exploratory Data Analysis

In this study, 169 answers to the TDO survey were collected and analyzed following the system architecture described in the previous Section.

The present exploratory data analysis aims to provide useful insights into the data collection and integration processes.

In particular, the collected data were sent by 38 cystic fibrosis patients through five subsequent submissions, scheduled every three months on average.

Figure 2 shows a violin plot describing the distribution of Adh-score in percentage associated with each submission epoch, while, in Table 3, the same information is provided in a tabular form. Although mean values of Adh-score (%) are in the range $[38.67\%, 51.45\%]$ for each epoch, the standard deviation and minimum and maximum values of Adh-score (%) show a considerable variation of Adh-values, with patients who sent zero spirometries and patients who transmitted three times more than the medical advice.
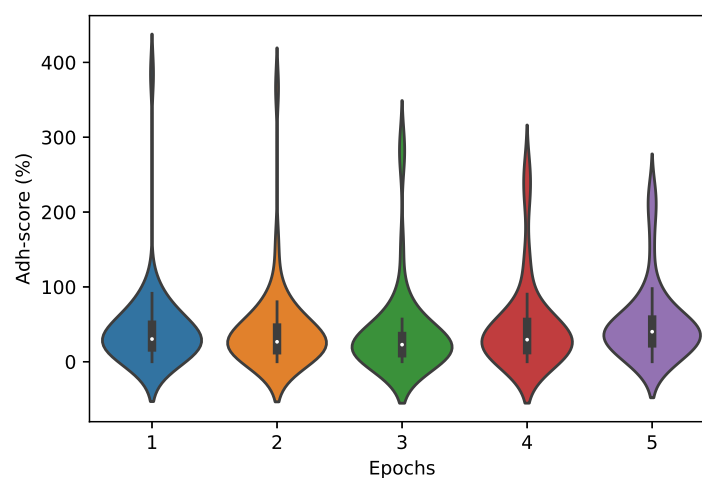
Figure 2. Distribution of Adh-score in percentage across the five subsequent submission epochs.

Table 3. Adh-score (%): descriptive statistics across five subsequent epochs.

	Epoch_1	Epoch_2	Epoch_3	Epoch_4	Epoch_5
mean	42.67	41.77	38.67	47.39	51.45
std	57.25	57.70	59.64	59.54	51.82
min	0.00	0.00	0.00	0.00	0.00
max	384.62	365.76	293.55	260.96	229.50

A comprehensive analysis of the responses to the TDO survey is beyond the scope of this paper. Instead, only answers to two open-ended questions collected from the TDO survey are discussed:

- **Q1 -** "What do you think about telemedicine?"
- **Q2 -** "Since you joined the telemonitoring program, what has improved the quality of your life?"

A polarity score ranging in $[-1, 1]$ was inferred by adapting the VADER framework to the Italian language and considered as a numerical feature for each set of answers. In Figure 3, the sentiment polarity with respect to the TDO survey Question Q1 is shown through time. In particular, the polarity intensities for the five different epochs are shown in different colors. The results show an overall positive opinion about telemedicine. In Figure 4, the sentiment polarity with respect to the TDO survey Question Q2 is shown through time. Answers related to this question show a more negative polarity score with respect to Question Q1.
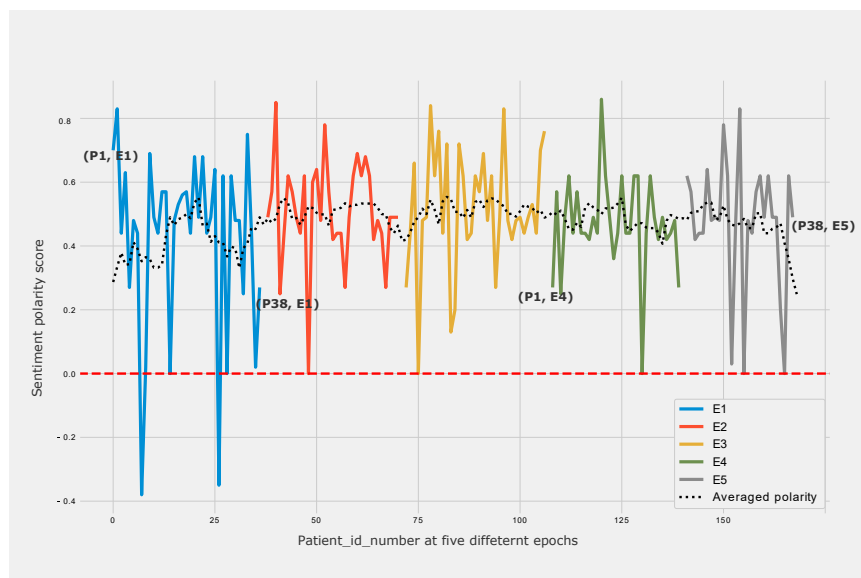
Figure 3. Sentiment polarity associate to Question Q1 are visualized through time. Responses are represented with the relative patient id and the questionnaire session epoch, i.e., E1, E2, E3, etc. On the y-axis, the compound polarity score related to the patient answer at the epoch E_j is inferred by the VADER framework's adaption to the Italian language.
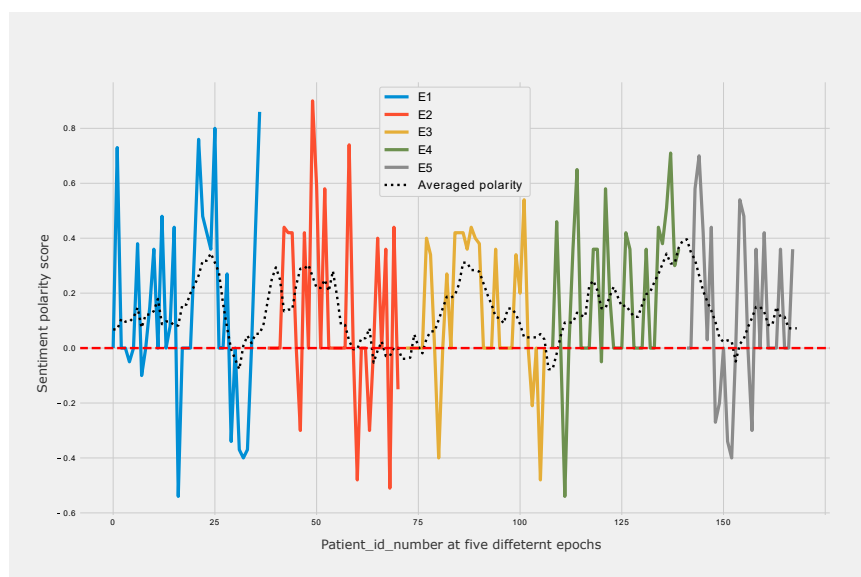
Figure 4. Sentiment polarity associate to Question Q2 are visualized through time. Responses are represented with the relative patient id and the questionnaire session epoch, i.e., E1, E2, E3, etc. On the y-axis, the compound polarity score related to the patient answer at the epoch T_j is inferred by the VADER framework's adaption to the Italian language.

To further provide some insights about the latent aspects more frequently mentioned by patients, in Figure 5, the 50 words resulting more used by patient are shown. The set of free-text answers was pre-processed with standard NLP techniques, i.e., tokenization, stop word removal, and lemmatization.

Figure 5. Word cloud showing the most frequent tokens with respect to the answers to question "What do you think about telemedicine?". Tokens with the largest font size are the most frequent.

The results show how "excellent", "useful", "tool", "health", and "patient" are the most common words in the patient response through time.

3.2. Testing Granger-Causality

Three time-series were considered, i.e., polarity scores related to Q1 and Q2 and the time series of Adh-scores. The Augmented Dickey–Fuller Test showed that for all the three considered time series the stationarity condition holds (p-value$= 4.6124 \times 10^{-18}$, p-value$= 3.2185 \times 10^{-7}$, and p-value $= 0.0035$, respectively).

Two Granger-causality tests were performed to check whether Q1 Granger-causes Adh-score and whether Q2 Granger-causes Adh-score. Moreover, since all the three series are considered contemporaneously, we also need to check whether Adh-score Granger-causes Q1 and whether Adh-score Granger-causes Q2. Three different test-statistics, i.e., F-test, chi2, and likelihood-ratio, were considered, with the number of lags varying from one to four. Tables 4 and 5 show the results in terms of p-values. It can be seen that both Q1 and Q2 Granger-cause Adh-score for lag = 1. On the other hand, Adh-score appears to not Granger-cause Q1 or Q2.

Therefore, the results suggest the existence of a predictive relationship between the polarity scores series associated with Q1 and the polarity score series associated with Q2 with respect to the Adh-score.

Table 4. Q1 and Adh-score: p-value of Granger-causality test performed with three different statistics and four different lags.

	Q1$\longrightarrow$ Adh-Score			Adh-Score$\longrightarrow$ Q1		
	F-test	Chi2 test	Likelihood-ratio	*F*-test	Chi2 test	Likelihood-ratio
1	0.0368	0.0339	0.0350	0.4076	0.4027	0.4031
2	0.0998	0.0908	0.0936	0.6667	0.6585	0.6592
3	0.1063	0.0917	0.0963	0.6628	0.6479	0.6496
4	0.2032	0.1760	0.1833	0.8200	0.8061	0.8064

Table 5. Q2 and Adh-score: *p*-value of Granger-causality test performed with three different statistics and four different lags.

	Q2⟶ Adh-Score			Adh-Score⟶ Q2		
	F-test	Chi2 test	Likelihood-ratio	*F*-test	Chi2 test	Likelihood-ratio
1	0.0020	0.0016	0.0028	0.9970	0.9969	0.9969
2	0.0174	0.0141	0.0155	0.7455	0.7390	0.7394
3	0.0547	0.0446	0.0482	0.8235	0.8148	0.9154
4	0.0865	0.0685	0.0744	0.7566	0.7387	0.7406

3.3. Discussion

The survey instrument and the analysis pipeline were applied to a real case study related to the remote follow-up of patients with cystic fibrosis, held in collaboration with the Cystic Fibrosis Unit, at Children's Hospital "Bambino Gesù", Rome, Italy.

In particular, 169 online surveys sent by 38 patients enrolled in a home telemonitoring program provided by the Cystic Fibrosis Unit at the "Bambino Gesù" Children's Hospital in Rome, Italy, were collected and analyzed through five subsequent questionnaire submissions.

Only answers to two open-ended questions were considered, i.e., Q1 "What do you think about telemedicine?" and Q2 "Since you joined the telemonitoring program, what has improved the quality of your life?".

The time-series of polarity score inferred through the adaption of VADER to the Italian language were used as numerical features to perform the Granger-causality testing model to investigate whether a predictive relationship between the polarity score of open-ended questions and Adh-score may exist.

The experimental results reported in Tables 4 and 5 therefore suggest that, under a Granger-causality perspective, the existence of a predictive relationship between the polarity scores series associated to Q1 and the Adh-score (lag = 1, *p*-value = 0.0339, statistic = Chi2 test) and between the polarity score series associated to Q2 and the Adh-score (lag = 1, *p*-value = 0.0016, statistic = Chi2 test).

The results are consistent with the hypothesis that the polarities extracted from patients' opinion on telemedicine may help predict their average adherence one epoch after the survey administration.

If supported, these results may enable the possibility to intervene early, in a targeted and individual way, to avoid drop-out and continue with the home telemonitoring program which represents a valid aspect of care for patients with Cystic Fibrosis.

Moreover, the early recognition of the reasons that lead the patient to drop-out and intervene immediately may result in:

- better assistance to the cystic fibrosis patient, both clinical and psychological (quality of life and self-management of the disease); and
- improvement of the quality of the telemonitoring program.

4. Conclusions

In the present paper, a system architecture for the extraction of emotional states from textual contents, designed to support the monitoring of patients with chronic disease, is presented. The main goal of the proposed system is to present a methodology to capture the underlying opinions that chronic patients have about the program they are enrolled in and to investigate whether these features may help in the early prediction of patient drop-out from the telemedicine program.

The proposed system is designed in an end-to-end fashion to provide support through the whole process, including the implementation of the questionnaire, the survey administration at scheduled intervals, as well as the analysis. Specific contributions are:

- the design of a self-hosted web-based survey instrument built on top of LimeSurvey for the management of online inquiries over time;
- an analysis pipeline that exploits sentiment analysis techniques to infer a sentiment polarity score for each open-ended answer and uses it as a numerical feature (to the best of our knowledge, this is the first time that this kind of approach has been proposed); and
- the validation of both the survey instrument and the analysis pipeline, which were applied to collect and analyze 169 TDO survey responses sent by 38 patients enrolled in a home telemonitoring program provided by the Cystic Fibrosis Unit at the "Bambino Gesu" Children Hospital in Rome, Italy.

In particular, in the present study, we focused on three variables modeled as time series: the polarity scores extracted from the responses to Question Q1, the polarity scores extracted from the responses to Question Q2, and an adherence score (Adh-score) defined for each epoch starting from the number of spirometries over a three-month window provided by the medical team.

The Granger-causality testing model was performed to assess whether a predictive relationship between the polarity scores series associated to Q1 and the Adh-score (lag = 1, *p*-value = 0.0339, statistic = Chi2 test) as well as between the polarity score series associated to Q2 and the Adh-score (lag = 1, *p*-value = 0.0016, statistic = Chi2 test).

Limitations to the present analysis may be found in the small number of data collected up-to-date, which does not allow the investigation of changes for a single patient through time.

Moreover, not every patient answered each survey session, which may have an impact on the number of lags.

Nevertheless, the promising results encourage us to further investigate the potentiality of the proposed architecture and the analysis pipeline with the aim to develop, as future work, a predictive system for the early detection of poorly-adherent patients that may also alert doctors to contact patients and eventually update/personalize their telemedicine program (e.g., in terms of timing, technological equipment, psychological counseling, etc.).

Appendix A

The Telemedicine Drop-Out (TDO) survey

1. What do you think about Telemedicine?
2. What improvement have you had since you have been in Telemonitoring? Choose an answer for each question

I felt safer/more peaceful about my health	Strongly disagree	Disagree	indifferent	Agree	Strongly agree
I felt more followed	Strongly disagree	Disagree	indifferent	Agree	Strongly agree

3. Since you joined the Telemonitoring program, what has improved your quality of life?
4. Since you joined the Telemonitoring program, do you find that the time spent on disease management has decreased?
   □ Yes   □ No
   In yes: Why, choose among the sequel
   □ I do fewer intravenous therapies □ Fewer days of hospitalization
5. Have you ever had difficulty sending spirometry?
   □ Yes, why?   □ No, why?
6. In your opinion, when you don't send spirometry results, what are the related reasons?
   Choose one or more of the following options
   □ Instrument malfunction   □ Anxiety
   □ Lack of time   □ Forgetfulness
   □ I recently performed spirometry at the hospital   □ I do not feel well
   □ Other:
7. When you feel physically unwell, do you find it difficult to transmit?
   □ Yes, why?   □ No, why?
8. When you are emotionally down, do you find it difficult to perform the spirometry test?
9. When you don't send the spirometry test, there are any prevailing emotions related to this event?
   Choose one or more of the following options
   □ Anxiety   □ Worry
   □ Fear   □ Anger
   □ Sadness   □ Other:
10. If there have been times that you have not performed and / or have not sent spirometry, what prompted you not to do it?
11. Would you like to be contacted after you have transmitted the data?

Strongly disagree	Disagree	indifferent	Agree	Strongly agree

12. Has your relationship with doctors / nurses changed since you followed in Telemonitoring?
    □ Yes, why?   □ No, why?
13. Do you share the sending of data with your family?
    □ Yes, why?   □ No, why?
14. Do you share the sending of data with your family?
    □ Yes, why?   □ No, why?
15. What would you change in the Telemonitoring program you are enrolled in?

References

1. Ryu, S. Telemedicine: Opportunities and Developments in Member States: Report on the Second Global Survey on eHealth 2009 (Global Observatory for eHealth Series, Volume 2). *Healthc. Inf. Res.* **2012**, *18*, 153. [CrossRef]
2. White, L.A.E.; Krousel-Wood, M.A.; Mather, F. Technology meets healthcare: Distance learning and telehealth. *Ochsner J.* **2001**, *3*, 22–29. [PubMed]

3. Suter, P.; Suter, W.N.; Johnston, D. Theory-based telehealth and patient empowerment. *Popul. Health Manag.* **2011**, *14*, 87–92. [CrossRef] [PubMed]

4. Nielsen, M.K.; Johannessen, H. Patient empowerment and involvement in telemedicine. *J. Nurs. Educ. Pract.* **2019**, *9*, 54–58. [CrossRef]

5. Delgoshaei, B.; Mobinizadeh, M.; Mojdekar, R.; Afzal, E.; Arabloo, J.; Mohamadi, E. Telemedicine: A systematic review of economic evaluations. *Med J. Islam. Repub. Iran (MJIRI)* **2017**, *31*, 754–761. [CrossRef]

6. Armaignac, D.L.; Saxena, A.; Rubens, M.; Valle, C.A.; Williams, L.M.S.; Veledar, E.; Gidel, L.T. Impact of Telemedicine on Mortality, Length of Stay, and Cost Among Patients in Progressive Care Units: Experience From a Large Healthcare System. *Crit. Care Med.* **2018**, *46*, 728. [CrossRef]

7. Polisena, J.; Tran, K.; Cimon, K.; Hutton, B.; McGill, S.; Palmer, K. Home telehealth for diabetes management: A systematic review and meta-analysis. *Diabetes Obes. Metab.* **2009**, *11*, 913–930. [CrossRef]

8. Gorst, S.L.; Armitage, C.J.; Brownsell, S.; Hawley, M.S. Home telehealth uptake and continued use among heart failure and chronic obstructive pulmonary disease patients: A systematic review. *Ann. Behav. Med.* **2014**, *48*, 323–336. [CrossRef]

9. Tagliente, I.; Solvoll, T.; Murgia, F.; Bella, S. Telemonitoring in cystic fibrosis: A 4-year assessment and simulation for the next 6 years. *Interact. J. Med. Res.* **2016**, *5*, e11. [CrossRef]

10. Sibley, C.D.; Rabin, H.; Surette, M.G. Cystic fibrosis: A polymicrobial infectious disease. *Future Microbiol.* **2006**, *1*, 53–61. [CrossRef]

11. Bella, S.; Murgia, F.; Tozzi, A.; Cotognini, C.; Lucidi, V. Five years of telemedicine in cystic fibrosis disease. *La Clinica Terapeutica* **2009**, *160*, 457–460. [PubMed]

12. Crombie, I.K. *Research in Health Care: Design, Conduct and Interpretation of Health Services Research*; John Wiley & Sons: Hoboken, NJ, USA, 1996.

13. Allery, L.A. Design and use questionnaires for research in medical education. *Educ. Prim. Care* **2016**, *27*, 234–238. [CrossRef]

14. Popping, R. Analyzing open-ended questions by means of text analysis procedures. *Bull. Sociol. Methodol. De Méthodologie Sociol.* **2015**, *128*, 23–39. [CrossRef]

15. Schuman, H.; Presser, S. The open and closed question. *Am. Sociol. Rev.* **1979**, *44*, 692–712. [CrossRef]

16. Schwarz, N. Self-reports: How the questions shape the answers. *Am. Psychol.* **1999**, *54*, 93. [CrossRef]

17. Friborg, O.; Rosenvinge, J.H. A comparison of open-ended and closed questions in the prediction of mental health. *Qual. Quant.* **2013**, *47*, 1397–1411. [CrossRef]

18. Zucco, C.; Calabrese, B.; Agapito, G.; Guzzi, P.H.; Cannataro, M. Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, e1333. [CrossRef]

19. Murgia, F.; Cilli, M.; Renzetti, E.; Majo, F.; Soldi, D.; Lucidi, V.; Bella, F.; Bella, S. Remote telematic control in cystic fibrosis. *La Clinica Terapeutica* **2011**, *162*, e121–e124.

20. Zucco, C.; Bella, S.; Paglia, C.; Tabarini, P.; Cannataro, M. Predicting Abandonment in Telehomecare Programs Using Sentiment Analysis: A System Proposal. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, 3–6 December 2018*; Zheng, H.J., Callejas, Z., Griol, D., Wang, H., Hu, X., Schmidt, H.H.H.W., Baumbach, J., Dickerson, J., Zhang, L., Eds.; IEEE Computer Society: Washington, DC, USA, 2018; pp. 1734–1739. [CrossRef]

21. Team, L.; Carsten, S. LimeSurvey: An open source survey tool. *LimeSurvey Project* **2012**. Available online: https://www.limesurvey.org/en/ (accessed on 25 November 2020).

22. McKinney, W. Pandas: A Foundational Python library for Data Analysis and Statistics. *Python High Perform. Sci. Comput.* **2011**, *14*. Available online: https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf (accessed on 25 November 2020).

23. Hutto, C.; Vader, G.E. A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social, Ann Arbor, MI, USA, 1–4 June 2014; pp. 1–4.

24. Dickey, D.A.; Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Assoc.* **1979**, *74*, 427–431.

25. Granger, C.W. Testing for causality: A personal viewpoint. *J. Econ. Dyn. Control* **1980**, *2*, 329–352. [CrossRef]

26. Pennebaker, J.W.; Boyd, R.L.; Jordan, K.; Blackburn, K. The Development and Psychometric Properties of LIWC2015. Technical Report. 2015. Available online: https://repositories.lib.utexas.edu/handle/2152/31333 (accessed on 25 November 2020).

27. Bradley, M.M.; Lang, P.J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*; Technical Report for C-1; The Center for Research in Psychophysiology, University of Florida: Gainesville, FL, USA, 1999.

28. Basile, V.; Nissim, M. Sentiment analysis on Italian tweets. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Atlanta, Georgia, 14 June 2013*; Association for Computational Linguistics: Atlanta, Georgia, 2013; pp. 100–107.

29. Pianta, E.; Bentivogli, L.; Girardi, C. MultiWordNet: Developing an Aligned Multilingual Database. In Proceedings of the First International Conference on Global WordNet, Mysore, India, 21–25 January 2002; pp. 293–302.