International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 7: Issue 1 [ISSN : 2250-3765]
Publication Date : 06 April, 2017

Digitalization of Medical Questionnaires

Constraction of the medical questionnaire recognition system and the relevant algorithms.

Yuji Kuramoto, Yoshitaka Nakanishi, Koichi Udo, Carlisle St Martin, Shoko Otsuka

***Abstract*— Digitalizing medical questionnaires to process important clinical data can greatly contribute to the field of medicine by allowing medical practitioners to do direct statistical analyses on patient responses. The aim is to construct a system at various medical facilities where the answer sheets to medical questionnaires can be digitalized and the responses scanned and used for clinical analysis. A demonstration of this system has been conducted at the hospital. The hypothesis is that the system will have a readability rate of greater than 95% after improvements have been made to both the set-up and to the algorithm. This study demonstrated that it is possible to digitalize the answer sheet to the medical questionnaire. Moreover, constructing the system is inexpensive and isn't negatively affected by the kind of electronic medical record storage system used. In addition, due to the improvements made to the recognition algorithm, readability is higher, and the system doesn't increase the workload of the nurses and medical staff.**

***Keywords—digitalization, medical questionnaires*, OCR, EMR, readability**

I. Introduction

There are many advantages to digitalizing medical questionnaires. For example, various statistical analyses can be applied to study the relationship between such things as the patient's main complaint and hospital visitation, main complaint and diagnosis, main complaint and region, and main complaint and age. Moreover, it is easier for doctors and medical staff to include the patient's main complaint in the patient's own words into their electronic medical records (EMR). In this way, digitalizing medical questionnaires can make a huge contribution to the field of medical information.

Nowadays, medical questionnaires using iPads is also becoming more popular[1]. However, due to the aging society phenomenon in Japan, the number of patients over 60 years of age accounts for 60% of all the patients that require medical attention[2].

Yuji Kuramoto, Graduate School of Science and Technology
Kumamoto University
Japan

Yoshitaka Nakanishi, Professor, Dr. Eng.
Kumamoto University
Japan

Koichi Udo,Ph.D.,
Coloproctology Center Takano Hospital
Japan

Carlisle St Martin, MD
USA

Shoko Otsuka, CEO
Integra System, Inc.
Japan

For the younger generation it is possible to administer digitalized medical questionnaires using iPads or smart phones but they may pose a problem for the elderly and therefore, may be limited for use in dentistry and plastic surgery which many younger patients visit.

Therefore, a new digitalized system was developed that involves filling-out a medical questionnaire scanning the data into the system, and has proven to be user-friendly for even elderly patients. The aim of this paper is to discuss the medical questionnaire recognition system and the relevant algorithms.

II. Purpose

It is difficult to accurately digitalize medical questionnaires if you don't use the special mark sheet. However, for patients who are ill and for those who are elderly it is extremely difficult to fill-out the mark sheet. Therefore, the main purpose is to develop a digitalized medical questionnaire on a separate sheet of paper that anyone can fill-out. In addition, it is extremely important that both patients and nurses can easily use the system, that the data can be scanned with a more than 95% accuracy rate, and that it reduces the workload for nurses. Moreover, the goal is to create a system that is inexpensive, doesn't negatively affect companies that sell electronic medical record storage systems, and that can be used anywhere.

III. System Summary

The overall structure of the system is shown in Figure 1. The system is made up of the electronic medical records terminal, the medical questionnaire reading scanner, the scanned data storage device (NAS), the master data storage device, and the clinical information integrated management system.

First, the patient fills-out the medical questionnaire. Once that has been completed they give it to the nurse and they then verbally give more details about their main complaint. Second, the nurse digitalizes and scans the medical questionnaire and then compares the scanned version with the digitalized version to make sure that no input error occurred (Figure 2). The nurse also checks to see if it is necessary to input any important descriptive data in non-digitalized form that would help better understand the patient's condition. Finally, the nurse saves the medical questionnaire as well as the digitalized data and the essential part of the screen data onto the Network Attached Storage site (NAS). The original scanned data is sent to the clinical information integrated management system. The doctor can then access this system from his or her electronic medical records terminal and look over the data obtained from the above mentioned digitalized medical questionnaire when they are consulting with a patient in the examination room. After the data is carefully reviewed the doctor writes a brief

description about the patient's main complaint based on the digitalized data and then copy pastes it into the patient's electronic medical records. An additional advantage of using this system is that once the digitalized data accumulates it can be saved in mdb format and various statistical analyses can easily be done.

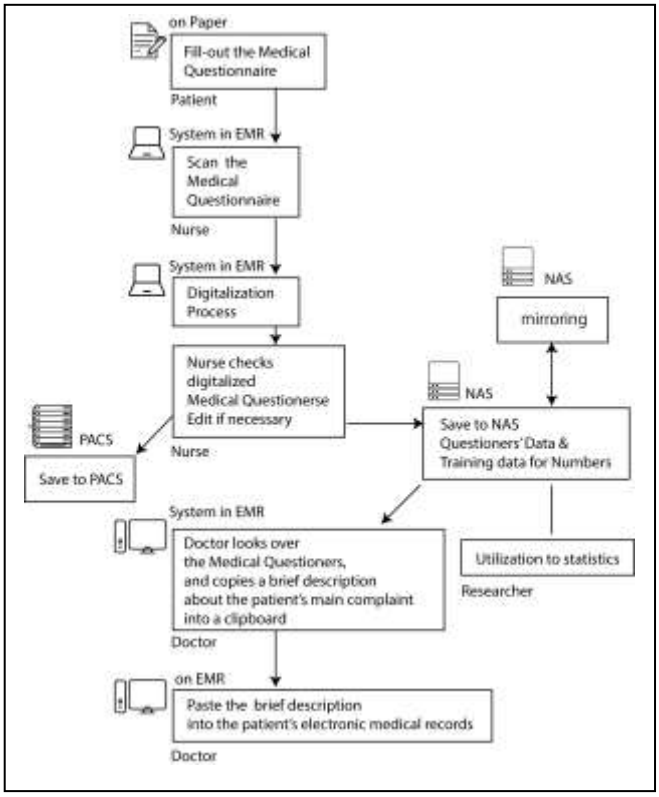

Figure 1. The overall structure of the system

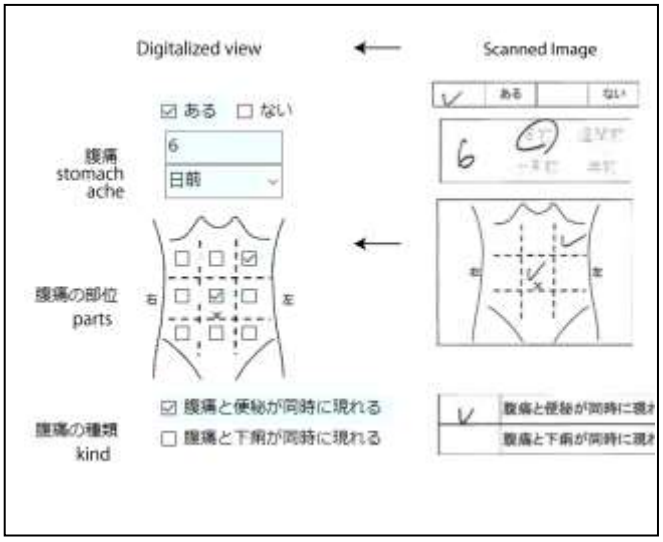

Figure 2. *Digitalized view and Scanned image*

A. *Inexpensive to set-up*

1) Incompatible link

The ideal situation is to be able to save the data from the digitalized medical questionnaire into the electronic medical record storage system. However, there are various different kinds of electronic medical record storage systems and it is difficult to deal with each one. Especially in cases where the electronic medical record storage system is incompatible with the digitalized system, and in cases where it is possible, the costs are much too high. Therefore, save the data from the medical questionnaires independently in NAS. Both the patient's records and the patient's history can be saved in NAS and the only requirement to be able to upload the data from there onto the electronic medical record storage system is to install a button to allow access to the above mentioned system. The obvious advantages of modifying the system include the relatively low cost to install the link and that both the doctors and nurses can see the data obtained from the medical questionnaires on the electronic medical record storage system.

2) Database unnecessary

Normally, health care data obtained from such sources as medical questionnaires are placed in one of the following databases; SQL Server, Oracle, or Cache. However, the software needed to create these databases is expensive and the hardware and maintenance costs are high. Therefore, all the important data is saved in a text file. The patient and the date of medical treatment can be quickly accessed by typing the name of the folder and/or the name of the file into the search engine and the information can then be used as an index. Moreover, the digitalized data stored in mdb can be copied and various statistical analyses can be conducted. The main reason why mdb is not used is because of the large number of people simultaneously using the system which greatly reduces the computing power and increases the possibility that the system could crash.

B. *The saved contents of the medical questionnaires*

Separate and save the following items of the scanned medical questionnaire.

① The "checked" items and the selected items. For example, Pain" Yes" " No",Intensity of Pain" Strong" " Weak". Digitalize the selected items h beforehand " True（1）" " False（0）"

② Number items. For example, items that can be digitalized include the date the symptoms began, frequency of the symptoms, and body temperature at the time of examination. However, due to obvious reasons (i.e. error in personal data), items such as personal phone numbers and birthdates should not be scanned if an accuracy rate of 100% cannot be achieved.

③ Non-number items with high statistical value manually put into the system. For example, the practice of putting detailed symptoms into the system by the nurse is encouraged.

④Non-number items with low statistical value saved into the system as jpeg images. For example, family occupation and family medical history.

C. *Master data from the medical questionnaire*

Coordinate the questions in the medical questionnaire with the type of question (same order as B above - ① ～ ④) when you prepare the master data. Moreover, save the coordinates of the special marks at the corners so that you can make adjustments if the answers are not centered or if they are in the wrong location. The system has a recognition processing function to help correct such errors.

IV. System Installation

At installation a test run will be conducted at the hospital to accommodate for the following items; the local language (regional dialects), variance in age (average age 47; median age 62), 166 beds, and 65,000 outpatients per year. In order for the system to produce a reading accuracy rate of over 95%, improvements in installation and improvements in the algorithms will be conducted to ensure optimal performance during the demonstration.

A. *Scan set-up*

First, it is important that the most suitable colors and image size of the scanner is selected and set-up. Black and white are the two colors normally used to read the "check" items and the number items but the level of readability in the clinical information integrated management system is low . Moreover, according to the Japanese e-Document Law, the image size of the scanner must be more than 200 dpi. However, if the image size is too big, it takes a lot of time to process and analyze the data (Table Ⅰ). Therefore, the important point here is that the color of the scanner should be set at gray scale and image size set at 300 dpi and then the data can be scanned and stored in the clinical information integrated management system. Once the data is in the system it can be condensed to 150 dpi, changed to black and white and then analyzed.

TABLE I. THE AVERAGE TIME FOR DIGITALIZATION OF 10 IMAGES

Resolution	Average time(sec)
300dpi	5.2
250dpi	4.5
200dpi	3.2
150dpi	2.7
100dpi	2

B. *Image adjustability*

If the image was not aligned properly when it was scanned it can be adjusted afterwards. In order to do that, use the up-down-right-left positioning control function of the master data to correct the misalignment. In addition, use the adjustable mark sheet with the medical questionnaire so that you can move the data if a misalignment error occurs.

1) Mark in the upper left hand corner

To get the first position of the medical questionnaire, print the mark "■" located in the upper left hand corner of the mark sheet. When the black mark is detected to be more than 3 pixels in size after pixel-by-pixel inspection starting from the upper left hand corner of the scanned image, the proper positioning process will begin. Compare the position of the scanned mark with the position of the mark on the master data sheet and then adjust the image of the scanned mark by using the up-down-right-left positioning control function. The accuracy rate for aligning the marks is 100%. However, the accuracy rate for aligning the master data with the overall image is only 85%.

2) Four corner marks

The 15% adjustment error mentioned in 2-1 is usually the result of mistakenly placing the mark sheet on an angle before it is scanned or by enlarging and condensing the data. Therefore, align the marks "■" in the four corners of the mark sheet with the marks "■" in the four corners of the master data and try using the perspective projection conversion method of adjustment. To begin, create an algorithm from the C# computer programming language that can adjust the position (up-down-right-left) of an image one pixel at a time. The algorithm needs about three seconds to adjust the image. For nurses three seconds is a long time. So use openCV Perspective projection conversion. Using openCV from C# can reduce the adjustment time to 0.2 seconds and is a 100% accurate.

C. *Reading of the "check" items and selection items*

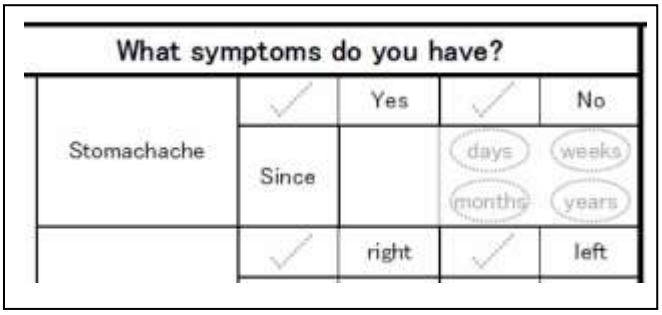

Figure 3. Watermarks image

TABLE II. THE AVERAGE TIME FOR DIGITALIZATION OF 10 IMAGES

Threshold value(%)	Recognition rate(%)	Note
80	20	All recognized as "False"
70	20	All recognized as "False"
60	20	All recognized as "False"
50	20	All recognized as "False"
40	70	
30	100	
20	80	All recognized as "True"
10	80	All recognized as "True"

Use the recognition processing and trimming functions of the scanned image after the position of the check items and selection items have been established. To read the check mark and the selected items use the black color contrast (0-

100 percent scale) setting of the trimmed image. In order to make it easier for the patients to fill-out the medical questionnaire scan the watermarks of the selected items and check items as shown in Figure 3. Image recognition is done by using the black and white color contrast function and not by trying to determine if the patient wrote a check mark, an "X" or any other possible mark in the designated answer box. This is the reason why it is important to use a color code (binary system) for black and white. Once a color code for black and white has been decided establish the range for color intensity (brightness). If the brightness is set too low then the image becomes too dark and if it is set too high then the image becomes too light. It is important to set the range for the check items and selection items so that the watermark disappears and only the patient's answers remain. If the range is set too high the patient's answers will disappear and you'll always get a "False" output. In contrast, if the range is set too low the watermark will be scanned and you'll always get a "True" output. Table II shows the results of a comparison that was done on the different rates of readability. The findings reveal that the range for normal readability is between 30-39% and that if the range is set at 30% the check mark readability rate is 100%.

D. *Recognition of number items*

Use the recognition processing and trimming functions of the scanned image after the position of the numbered items have been established. In this situation, there will be a 38% error rate if the OCR engine is used on the whole image. In order to avoid such errors, use the OCR engine to create a margin and trim each individual number as shown in Figure 4. Use the following algorithm to trim each individual number: select the consecutive black starting point; select the consecutive white position from the starting point; determine the consecutive white position as the end point and acquire both positons of the starting and end points; and then create a margin and trim after the positions have been acquired.

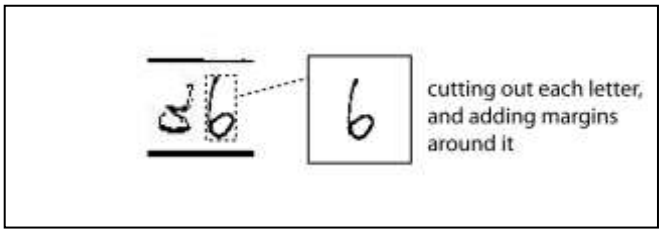

Figure 4. Create a margin and trim each individual number

After the nurse has finished and checked it, get him or her to use the OCR engine on each number of the trimmed image. The readability rate can be improved if the nurses use letters and numbers that look similar and then try to simulate common patient errors. Towards this end, a system was established to encourage the nurses to use the OCR engine in the above mentioned way and now the readability rate has improved to 96%.

After improvements were made to the above mentioned algorithm it was tested using data from 100 medical questionnaire responses. Table 3 shows that the recognition error for numbers has significantly dropped and that the reading accuracy rate is now 95%. However, continued practice is necessary to identify the various ways that numbers can be written.

TABLE III. THE RECOGNITION RATE OF 100 TYPES OF MEDICAL QUESTIONERS BY ITEM

	Check Items	**Select Items**	**Hand Written Number Items**	**Average recognition rate**
recognition rate(%)	98	96	92	**95.3**

V. Conclusion

This study has demonstrated the efficacy of digitalizing answer sheets for medical questionnaires and that mark sheets, iPads and other devices are unnecessary. Moreover, constructing the system is inexpensive and isn't negatively affected by the kind of electronic medical record storage system used. In addition, due to the improvements made to the recognition algorithm, readability is higher, and the system doesn't increase the workload of the nurses and medical staff.

Various devices used to digitalize medical questionnaires were examined in this study, and since Japan is an aging society, it was found that it is essential to develop recognition technology that is more efficient and to create answer forms for medical questionnaires that anyone can fill-out.

References

[1] Aki Maruo,Application of interview sheet using Linked Data, 2012.

[2] Health, Labour and Welfare Ministry, Comprehensive Survey of Living Conditions 2013.