METHOD ARTICLE

Health status classification model for medical adherence system in retirement township [version 1; peer review: awaiting peer review]

Abubaker Faisal Abubaker Sherif, Wooi Haw Tan, Chee Pun Ooi, Yi Fei Tan

Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, 63100, Malaysia

V1 First published: 20 Oct 2021, 10:1065
<https://doi.org/10.12688/f1000research.73332.1>
Latest published: 20 Oct 2021, 10:1065
<https://doi.org/10.12688/f1000research.73332.1>

Open Peer Review

Approval Status Awaiting Peer Review

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Medical adherence and remote patient monitoring have gained huge attention from researchers recently, especially with the need to observe the patients' health outside hospitals due to the ongoing pandemic. The main goal of this research work is to propose a health status classification model that provides a numerical indicator of the overall health condition of a patient via four major vital signs, which are body temperature, blood pressure, blood oxygen saturation level, and heart rate. A dataset has been prepared based on the data obtained from hospital records, with these four vital signs extracted for each patient. This dataset provides a label associating each patient to the number of medical diagnoses. Generally, the number of diagnoses correlates with the patient's medical condition, with no diagnoses indicating normal condition, one to two diagnoses suggest low risk, and more than that implies high risk. Thus, we propose a method to classify a patient's health status into three classes, which are normal, low risk and high risk. This would provide guidance for healthcare workers on the patient's medical condition. By training the classification model using the prepared dataset, the seriousness of a patient's health condition can be predicted. This prediction is performed by classifying the patients based on their four vital signs. Our tests have yielded encouraging results using precision and recall as the evaluation metrics. The key outcome of this work is a trained classification model that quantifies a patient's health condition based on four vital signs. Nevertheless, the model can be further improved by considering more input features such as medical history. The results obtained from this research can assist medical personnel by providing a secondary advice regarding the health status for the patients who are located remotely from the medical facilities.

Keywords

Medical adherence, home hospitalization, machine learning

This article is included in the **Artificial Intelligence and Machine Learning** gateway.

This article is included in the **Research Synergy Foundation** gateway.

Corresponding author: Wooi Haw Tan (twhaw@mmu.edu.my)

Author roles: **Abubaker Sherif AF:** Conceptualization, Data Curation, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation; **Tan WH:** Funding Acquisition, Investigation, Project Administration, Software, Supervision, Validation, Writing – Review & Editing; **Ooi CP:** Project Administration, Supervision; **Tan YF:** Validation

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Abubaker Sherif AF *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Abubaker Sherif AF, Tan WH, Ooi CP and Tan YF. **Health status classification model for medical adherence system in retirement township [version 1; peer review: awaiting peer review]** F1000Research 2021, 10:1065 <https://doi.org/10.12688/f1000research.73332.1>

First published: 20 Oct 2021, 10:1065 <https://doi.org/10.12688/f1000research.73332.1>

Introduction

Background

Medical adherence is described as a patient's willingness to follow their doctor's treatment plan by taking the medications recommended to them.¹ Failure to follow a treatment plan might result in negative clinical outcomes and a significant rise in hospitalisation cost.² According to a recent study, most patients do not follow their doctors' advice, resulting in a considerable rise in hospitalizations and medical visits.³

Many researchers and healthcare firms have been focusing on home hospitalization technology in recent years.⁴ Many approaches are used to identify medical adherence: some employ blood testing, some employ patient self-assessment feedback, and others employ pill dispensers to determine whether the medicine was taken. These solutions can give important information on a patient's behaviour, but they are ineffective, as some solutions are costly to operate, while others do not provide information on the type of illness.

By proposing a method for continuous monitoring of the patient's vital signs, this study intends to develop a solution that can aid doctors in monitoring the patient's health state and medication intake. We used machine learning to determine the patient's health status and match it to the pharmaceutical consumption schedule defined by doctors. Major vital indicators are taken by off-the-shelf medically approved devices. The data is logged and analysed, and the result is presented to doctors in the form of a suggestion, which they may use to alter medicine dose or frequency, or to schedule a doctor visit for the patient.

Literature review

A lot of research has been focused on medical adherence and home hospitalisation.^{5–12} Patients' in-hospital days were reduced when they were given a home hospitalisation option.⁵ The abilities to remotely monitor the patient's health state and check their adherence to prescriptions are critical factors for enabling home hospitalisation. Tripathi *et al.*⁶ proposed to gather information about the patient via tracking sensors and wearable devices, then transmitting the data to an Internet server, where the decision is taken whether to contact family members, an ambulance, or clinical aid. To enable a smart health environment, Zulkifli *et al.*⁷ created a health monitoring and information system. Patients may use their cell phones to submit feedback to physicians, and doctors will reply to their reports to determine if the patient needs an appointment or can continue therapy from home.

Due to the benefits of reducing hospital stays, cutting treatment costs, and freeing up doctors' time, home hospitalisation has garnered a lot of attention. Early release care, which permits patients to stay at home for a portion of their inpatient therapy, was studied by Hernández *et al.*⁸ They urged nurses to visit these patients on a regular basis and record their vital signs in an online system that doctors could access. Federman *et al.*⁹ assessed hospitalised patients and sent some of them home to undergo treatment. Nurses and health experts evaluated patients' vital signs before asking all patients to rate their treatment experience, with the results indicating that those who received therapy at home obtained a higher rating. Sherif *et al.*¹⁰ suggested a system to facilitate home hospitalisation, by tracking patients' medicine consumption using integrated electronics. Patients were instructed to record their medication adherence using an alarm button that was linked to a monitoring dashboard. While this strategy is useful for the monitoring of patients, it cannot assure that the medicines have been taken.

Similarly, Daramola *et al.*¹¹ used a smartphone application to report medicine consumption and relied on patient self-reporting. Kumar *et al.*¹² presented a similar method for measuring medication adherence using medicine dispensers. However, these approaches do not guarantee that the medicines have been consumed. To overcome these limitations, this paper intends to verify medicine intake by monitoring patients' vital signs and health status regularly.

Motivation

The study's main goal is to improve patients' medication adherence without the need of special technology, nurse or caregiver monitoring. Thousands of people have been hospitalised because of the current COVID-19 pandemic. Many healthcare institutions have run out of resources to handle the rising number of patients.¹³ Thus, most people with minor symptoms have been recommended to stay at home and monitor their health.¹³ This pandemic has highlighted the critical necessity for home hospitalisation. Consequently, we propose a complementary tool that assists doctors in monitoring patients' health from home.

Ethical considerations

This study received ethical approval from the ethical Review Board of the Multimedia University, Technology Transfer Office, Malaysia (EA0962021).

As healthy patients' data was collected retrospectively and anonymised, written informed consent was waived.

Methods

The method proposed in this paper consists of two sections. The first section is data preparation and pre-processing, while the second section is creating a classification model to identify the health status category based on the vital signs.

Data preparation

Blood pressure, body temperature, heart rate and blood oxygen saturation level can be considered as good indicators to predict the health status of the patient. The initial stage in data preparation was to assemble a labelled dataset collected by physicians to anticipate the patient's health status. We adopted a publicly available medical dataset called Medical Information Mart for Intensive Care III (MIMIC-III) [Clinical Database](#), which contains health data from over 40,000 patients hospitalised in intensive care units.¹⁴ The dataset consists of numerous tables with patients' information, such as vital signs and doctor diagnoses. The caregivers documented the data collected on an hourly basis. However, the data from a single patient is saved in different tables across the dataset. As a result, the data must be restructured before they can be used for model training. Patients who had more than four daily measurements, including blood oxygen saturation, blood pressure, body temperature, and heart rate, as well as a doctor's diagnosis, were selected. These readings were divided into two classes: the first class included those with two or fewer diagnoses while the second class comprised those with more than two diagnoses. To predict the health status of the patients, we needed to train the machine learning model with data from healthy people, who were not diagnosed with any diseases. Since vital sign data from healthy people were not widely available, we decided to collect these data with the help of a general physician. The resulting dataset consists of three classes: the first class corresponding to healthy people; the second class includes patients with up to two diagnoses; the third class includes patients with more than two diagnoses.

Training the machine learning model

Three multiclass classification models were proposed to help doctors by providing a secondary opinion based on vital signs, to determine how serious a patient's health condition is. The prediction was based on the labelled dataset, which consisted of three classes. The classifiers were chosen, based on the size of the dataset, were k-nearest neighbour classifier, linear support vector machine (SVM) classifier, and logistic regression classifier. [Scikit-learn](#) v. 0.23 was used to train these models, and is well-known for its classification methods.¹⁵ The purpose of the classifiers was to forecast the patient's health status class based on their vital signs. Diagnoses used in our prepared dataset were related to heart disease, respiratory sickness and hypertension, which the doctors would subsequently verify or reject in order to offer feedback data that could be used to train a more accurate model.

Results and discussion

A well-labelled and filtered dataset with five variables and one label was prepared. The mean values for blood oxygen saturation level, diastolic blood pressure, systolic blood pressure, heart rate, and body temperature were used as variables in the prepared dataset, as well as a label that indicated whether the measurements were normal, abnormal or abnormal with a serious condition. The variables included in the prepared dataset are listed below:

- Diastolic blood pressure
- Systolic blood pressure
- Heart rate
- Blood oxygen saturation level
- Temperature
- Label

A total of 1,382 rows were collected after filtering data for those patients who had been diagnosed with diseases related to the vital signs, including only those who had more than three readings per day. A similar approach was utilised to get 102 "normal" readings manually with the aid of a medical expert. Both tables were then concatenated; the patient ID was removed, and labels reading "0" were assigned to normal readings, while "1" was assigned to abnormal readings up to two diagnoses, and "2" was assigned to abnormal readings with more than two diagnoses.

Although we had collected as many data for healthy people as possible, the dataset was still imbalanced, with the majority of the data samples coming from the MIMIC-III dataset, resulting in biasing.

As shown in Table 1, the dataset consisted of 102 rows of healthy data samples (Class 0), 813 rows of low-risk data samples (Class 1) and 569 rows of high-risk data samples (Class 2). A 10% fraction of the data samples was used for testing. The rest of the dataset was split using repeated stratified K-Fold cross-validation, with four splits, and ten repetitions. One significant aspect of this validation technique is its balanced training weights and its testing of data while handling unbalanced data, since the percentage of observations with a particular categorical value is the same.¹⁴ Multiple splits and repetitions allowed the training and evaluation of different iterations of the same model, with training and test outcomes differing depending on the sampling data. After the data were divided into training and test sets, we trained three distinct models to assess their performance. These processing stages are shown in the system architecture of the proposed solution as depicted in Figure 1.

The first trained model was linear SVC. The settings of the hyperparameters used were 'C' set to 1, 'max_iter' set to 1000 and 'intercept_scaling' set to 1. This model exhibited a relatively low performance in predicting patients with more than three diagnoses, as shown in the confusion matrix shown in Table 2. The evaluation results in terms of precision and recall for the first model are presented in Table 3.

Table 1. Prepared dataset information.

	Class 0	Class 1	Class 2
Training	92	732	512
Testing	10	81	57

Figure 1. System architecture.

Table 2. Confusion matrix of linear support vector machine (SVM) classifier.

		Predicted class		
		0	1	2
True class	0	10	0	0
	1	0	50	31
	2	0	36	21

Table 3. Precision and recall values of linear support vector machine (SVM) classifier.

	Precision	Recall
Class 0	1.00	1.00
Class 1	0.58	0.62
Class 2	0.40	0.37

The second model trained was based on logistic regression, constructed with 'max_iter' set to 1000 and 'C' set to 1. As shown in [Table 4](#) and [Table 5](#), the logistic regression model demonstrated a modest performance in relation to the previous classifier with, better precision and recall for Class 2. Nevertheless, the recall for Class 1 was considerably lower.

The third model was the k-nearest neighbour classifier, trained with 'n_neighbors' set to 5, 'weights' set to 'uniform' and using Euclidean distance. The confusion matrix generated by applying the k-NN classifier on the test data is shown in [Table 6](#). The results for the trained model are presented in [Table 7](#). The performance of the model is on par with the first model, but worse than the second model for Class 2 in terms of recall.

After training three different classification models, we trained a majority-vote classifier, which aggregated the predictions of the three trained models and made the predictions based on the class that obtained the greatest number of votes. The results are presented in [Table 8](#) and [Table 9](#).

In this study, the majority-vote classifier was chosen according to its overall performance in terms of precision and recall, in comparison to the results of the other three trained models. Further improvements will be made in the future regarding imbalanced data training, or by adding more minority data to the dataset. The positive results suggest that the proposed classification model is feasible in predicting severity of health condition, especially between normal people and those with medical risks.

Table 4. Confusion matrix of logistic regression classifier.

		Predicted class		
		0	1	2
True class	0	10	0	0
	1	0	21	60
	2	0	13	44

Table 5. Precision and recall values of the logistic regression classifier.

	Precision	Recall
Class 0	1.00	1.00
Class 1	0.62	0.26
Class 2	0.42	0.77

Table 6. Confusion matrix of k-nearest neighbour classifier.

		Predicted class		
		0	1	2
True class	0	10	0	0
	1	0	58	23
	2	0	40	17

Table 7. Precision and recall values of k-nearest neighbour classifier.

	Precision	Recall
Class 0	1.00	1.00
Class 1	0.59	0.72
Class 2	0.43	0.30

Table 8. Confusion matrix of the majority-vote classifier.

		Predicted class		
		0	1	2
True class	0	10	0	0
	1	0	45	36
	2	0	29	28

Table 9. Precision and recall values of majority-vote classifier.

	Precision	Recall
Class 0	1.00	1.00
Class 1	0.61	0.56
Class 2	0.44	0.49

Conclusions

The goal of this study was to improve medical adherence among patients who get therapy at home, by tracking their health status using supervised machine learning models. We are confident that the proposed solution will increase medicine adherence and provide a way to enable home hospitalisation, which is now in great demand. The trained model can predict how serious a patient's health status is, as determined by the four measured vital signs. Clinicians should be able to accept or reject the model's predictions, providing feedback which can then be utilised to train and enhance another version of the model. To further improve the performance of the model, additional input variables such as medical history can be considered.

Data availability

Underlying data

The experiments in this work were carried out using data from the MIMIC-III Clinical Database: <https://doi.org/10.13026/C2XW26>.¹⁴

Extended data

Zenodo: Five vital signs of normal people, <https://doi.org/10.5281/zenodo.5549632>.¹⁶

This project contains the anonymised vital signs data from healthy patients collected retrospectively.

Analysis code available from: <https://github.com/abubakerSherif/A-Health-Status-Classification-Model/tree/v1.0>

Archived analysis code as at time of publication: <https://doi.org/10.5281/zenodo.5551854>

License: MIT

References

- Ho PM, Bryson CL, Rumsfeld JS: **Medication adherence: its importance in cardiovascular outcomes.** *Circulation*. 2009; **119**(23): 3028–3035. [PubMed Abstract](#) | [Publisher Full Text](#)
- World Health Organization, Healthtopics: chronicdiseases. [Reference Source](#)
- Span P: **A dose of confusion.** 2011. [Reference Source](#)
- Polese F, Carrubbo L, Caputo F, et al.: **Managing healthcare service ecosystems: abstracting a sustainability-based view from hospitalization at home (HaH) practices.** *Sustainability*. 2018; **10**(11): 3951. [Publisher Full Text](#)
- Heidenreich PA, Ruggiero CM, Massie BM: **Effect of a home monitoring system on hospitalization and resource use for patients with heart failure.** *Am. Heart J.* 1999; **138**(4): 633–640. [Publisher Full Text](#)
- Tripathi V, Shakeel F: **Monitoring health care system using internet of things – an immaculate pairing.** 2017 *International conference on next generation computing and information systems (ICNGCIS)*. IEEE. 2017; pp 153–158.

7. Zulkifli FY, Mustika IW: **Development of monitoring and health service information system to support smart health on android platform**. *2018 4th international conference on nano electronics research and education (ICNERE)*. IEEE. 2018; pp 1–6.
8. Hernaández C, Aibar J, Seijas N, *et al.*: **Implementation of home hospitalization and early discharge as an integrated care service: a ten years pragmatic assessment**. *Int. J. Integr. Care*. 2018; **18**(2), 12.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Federman AD, Soones T, DeCherrie LV, *et al.*: **Association of a bundled hospital-at-home and 30-day postacute transitional care program with clinical outcomes and patient experiences**. *JAMA Intern. Med.* 2018; **178**(8): 1033–1040.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Sherif S, Tan WH, Ooi CP, *et al.*: **LoRa driven medical adherence system**. *Bull Electr Eng Inf.* 2020; **9**(6): 2294–2301.
[Publisher Full Text](#)
11. Daramola O, Nysaulu P: **A digital collaborative framework for improved tuberculosis treatment adherence of patients in rural settings**. Cape Town, South Africa: Open innovations (OI); 2019; 297–303.
[Publisher Full Text](#)
12. Kumar MP, Nelakuditi UR: **IoT and I2C protocol based M-health medication assistive system for elderly people**. *IEEE 16th India Council international conference (INDICON)*, Rajkot, India. 2019; pp 1–4.
[Publisher Full Text](#)
13. Nundy S, Patel KK: **Hospital-at-home to support COVID-19 surge – Time to bring down the walls**. 2020.
[Reference Source](#)
14. Johnson AEW, Pollard TJ, Shen L, *et al.*: **MIMIC-III, a freely accessible critical care database**. *Scientific Data*. 2016; **3**: 160035.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Pedregosa F, Varoquaux G, Gramfort A, *et al.*: **Scikit-learn: machine learning in python**. *J. Mach. Learn. Res.* 2011; **12**(Oct): 2825–2830.
16. Sherif A, Tan WH, Ooi CP, *et al.*: **Five vital signs of normal people [Data set]**. *Zenodo*. 2021.
[Publisher Full Text](#)

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research