**Predicting when patients leave their bed sixty seconds in advance using machine learning algorithms**

**Roy Duivenvoorden, Salah Abdulkader, Theo van den Berg, Nick Keereweer & Floris van Lingen**

**The Hague University of Applied Sciences**  
**Johanna Westerdijkplein 75, 2521 EN Den Haag**

**Corresponding author: Roy Duivenvoorden**

**Abstract**   
Predicting when a patient leaves their bed can be valuable if said patient for example suffers from dementia and forgets they have trouble walking, said patient might try to leave their bed and hurt themselves in the process. As such this research had the goal to determine the optimal machine learning algorithm to predict when a patient leaves their bed making use of the BedSense designed by Momo Medical. In order to achieve this data collected by Momo Medical was used and prepared for differing models such as Logistic regression and Linear Support Vector Machine using different features based on the FSR sensors to make predictions 60 seconds in advance. In conclusion the optimal machine learning algorithm with an accuracy of 81%, a precision of 95% and a recall of 64% is Linear Support Vector Machine.

**Introduction**

The ever-growing medical expertise has led mankind to a growth in the expected duration of a human life [1]. The increase in life expectancy increases the amount of elderly people. Their homes often do not fulfill their living requirements, which forces them to move to a nursing home. The muscle capacity decreases dramatically, and some become bed ridden. This increases the chance of pressure ulcers [2]. Momo Medical has created a sensor plate to notify nurses when a patient needs to be moved in a new position. This sensor plate consists of 8 force-sensing resistors and 6 piezoelectric sensors.

This research uses that sensor plate to predict when a patient wants to leave their bed. Nursing homes might also house patients who are suffering from dementia. These patients often forget where they are and that they are unable to walk without assistance. When those patients wake up and try to walk, there is a high chance they fall. When a nurse is using the BedSense application, that can predict when a person will try to leave their bed, this will give them enough time to assist the patient with getting out of bed. Furthermore, this gives the nurses more time to focus on other things than checking on patients. The main research question was: “What is the optimal machine learning algorithm that can predict when a patient leaves their bed, using the collected data from the BedSense (version 9) sensor plate created by Momo Medical?”.

**Original data**

The raw data consists out of sensor data gathered during the day using the BedSense sensor. This data is collected from 99 anonymous patients. These patients are either suffering from dementia, rehabilitating, or recovering from injuries. The original raw data files consist out of 82 columns an overview of the main columns can be found at table 1. The columns that are useful for predictions are the six piezoelectric (PE) sensors and the eight force-sensing resistor (FSR) sensors. Most of the received data contained a period of 90 minutes wherein the bed status changes to 0 at the 45-minute mark. A small selection of data contained the full 24-hour period. There are three different bed statuses; where 0 shows that the patient is out of bed, 1 is seen as the patient is sitting in bed and 2 means that the patient is lying in bed.

|  |  |  |
| --- | --- | --- |
| **Name** | **Number of columns** | **Used** |
| **time** | 1 | Yes |
| **device\_id** | 1 | No |
| **Sp\_accelero\_** | 2 | No |
| **Sp\_electric\_** | 72 | No |
| **Sp\_resisitive\_** | 8 | Yes |
| **Hour\_setting** | 1 | No |
| **Patient\_present** | 1 | No |
| **Percentage** | 1 | No |
| **Patient\_detection** | 1 | No |
| **Patient\_detect\_state** | 1 | No |
| **Bed\_status** | 1 | Yes |
| **Moment\_of\_detection** | 1 | No |

Table 1: An overview on the main columns found in the raw data

**Data Preparation**

As seen in figure 1 data preparation started with receiving the data.

Data received:  
The data has been received by the researchers; this is the starting point of the data preparation.

Converting PE values:  
The Pe value sampling rate is converted to the same sampling rate as the FSR-sensors. This reduces the amount of columns from 92 to 12.

Converting Date Time:  
The Unix time is converted to a UTC DateTime index. This is useful because the Unix time is not easily interpretable unlike a UTC DateTime index.

Data Frame Merge:  
After converting the PE values and DateTime there currently are three separate data frames, one for PE values, one for FSR values and lastly one for bedstatus. These three Dataframes are merged based on the DateTime index to create a single data frame holding the basic sensor data.

Data Slicing:  
The DataFrame holds 90 *minutes’* worth of data. This data will be sliced in order to speed up the learning process. Every row represents a situation in which a patient gets out of their bed and the bedstatus becomes 0. To include the former situation and ones where patients do not get out of their bed two slices have to be made. First the data is sliced to include nine minutes before and one minute after the first bedstatus 0. These slices are meant to represent situations where patients leave their bed. Secondly a section of ten minutes is sliced where the bedstatus is constantly 2 and is added to the DataFrame. These slices are used to represent situations where patients stay in their bed.

Create Features:  
Features are created over the data available. The features are created to help the model make predictions.

Add Features to Data Frame:  
The created features are added to the data frame and any NaN values generated by the features are removed. These NaN values occur due to the nature of certain features needing 10 to 15 seconds worth of data in the past to function and leave 150 rows worth of NaN values at the start of the data frame.

Create List of Data Frames:  
Each csv file was turned into a data frame and these data frames where then added to a list of data frames. This will be used in the following step.

Time shift:  
The list of Data frames is loaded in and shifted back in time to gain the closest sensor and feature values at -60 seconds to -120 seconds and add these into a row on a data frame.

Data Frame:  
The end results is a single data frame holding rows of data where each row is a situation where a person is either lying in bed or getting out of bed with the vales per column looking back at a 10 second interval.

Graphical user interface, application

Description automatically generated

Figure 1: Flowchart data preparation

**Data splitting**

The dataset has been split into a training, validation, and a test set.

**Methods**

**Linear support vector machine (LSVM)**

Linear Support Vector Machine is a model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is as follows: The algorithm creates a line or a hyperplane which separates the data into classes. [3]. They are widely used for classification tasks [4]. Since the objective is to estimate a class it is a classification problem. Based on other research regarding classification tasks that use a SVM like the one below it was decided to experiment with SVM’s.[5, 6]

**Two class**

Two class or binary classification means that there are two classes, in case of this research a bedstatus with either a value of zero or one. The positive class zero means a patient has left their bed. The negative class 1 means a patient is in their bed.

**Logistic regression**

Logistic regression is a classification algorithm, used when the value of the target variable is categorical in nature. This model is most commonly used when the data in question has binary output, so when it belongs to one class or another, or is either a 0 or 1. Based on other research regarding classification tasks that use logistic regression to get good results like the one below it was decided to experiment with this kind of regression.[7, 8]

**Feature**

Forward selection has been used to select the model features [9].

The objective is to predict when a patient leaves their bed, hence the feature values are sliced so they can be used for predictions. This means the features used contain values starting two minutes before the moment of leaving the bed and going down to one minute before. The bed status is used as the ground truth.

*Bed\_status2* is a feature which uses the bed status one minute before the ground truth as a feature. This can be useful in case somebody is already sitting in their bed. When somebody is sitting in their bed there is a high chance the person will also try to get out of their bed.

The *left* feature uses the two most left FSR sensors. The value of this feature is the mean of both sensors. The *middle* feature uses the four most middle FSR sensors. The value of this feature is the mean of all the middle sensors. The *right* feature uses the two most right FSR sensors. The value of this feature is the mean of both sensors. A decrease or increase in these features can be interpret as a patient moving in their bed. This could mean that a patient is getting out of their bed.

*Avg\_column* is a feature which takes all the mean values of the sensors. Even though this does not mean much if the person stays in their bed, it would mean a lot if a person rolls over to a side of their bed. This could indicate a patient is trying to get out of their bed.

The *slope* is a feature which calculates the direction and steepness of the mean values of all FSR sensors. The values are calculated over the so called *Avg\_column* at present time and the value of the *Avg\_column* feature one second before present time. When people move heavily in their bed the slope value will give significantly higher values at that point which could indicate a person trying to leave their bed.

*FSR\_15s\_variance* is a feature which takes the variance over 15 seconds. The variance is the mean of the squared deviation. This feature tells us that FSR values are changing, which could indicate a patient is moving in their bed. Usually before a patient tries to get out of their bed, we observe a huge deviation in the variance.

The *Slopelmr\*slope* is a feature which calculates the slope over the following existing features *left, middle* and *right.* After calculating the slopes those values are multiplied with each other. After this the outcome is multiplied with the slope feature.

**Results**

The models have been tested on predicting whether patients are going to leave their bed or not. Those models have been tested in combination with the features described above. The performance of the models is measured by the following scores: accuracy, precision and recall [10] . The score from the linear support vector machine model is shown in table 2. The score from the logistic regression model is shown in table 3. The confusion matrixes show the distribution of TP, FP, FN and TN as shown for the LSVM in figure 2 and the Logistic regression model in figure 3

*Linear support vector machine*

Test set LSVM

|  |  |
| --- | --- |
| Accuracy score | 0.8050541516245487 |
| Precision score | 0.9516483516483516 |
| Recall score | 0.6358296622613803 |

Table 2: Test LSVM

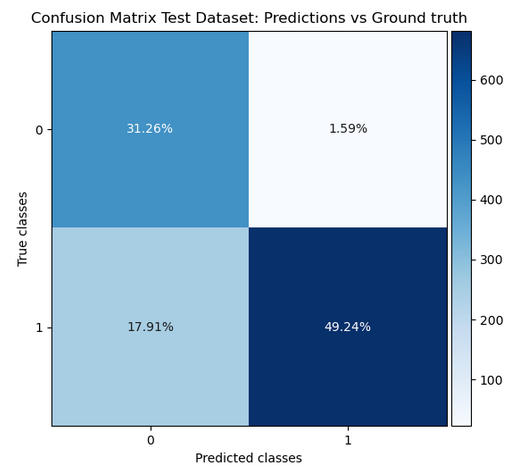


Figure 2: Confusion matrix LSVM

*Logistic regression*

Test set Logistic

|  |  |
| --- | --- |
| Accuracy score | 0.8064981949458484 |
| Precision score | 0.843956043956044 |
| Recall score | 0.6609294320137694 |

Table 3: Test Logistic

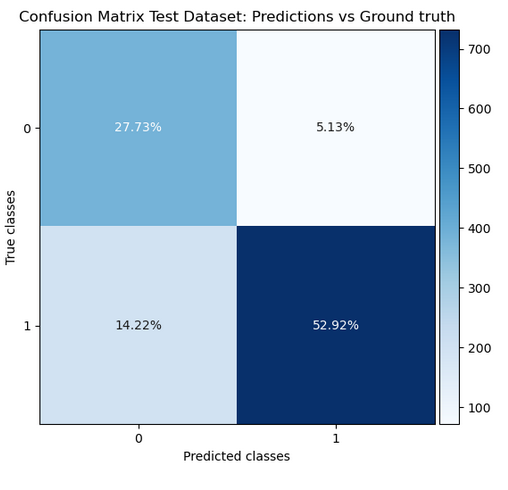


Figure 3: Confusion matrix Logistic Regression

**Discussion**

In this paragraph the validity of this research will be discussed. The data used to make predictions comes from three different nursing home departments: Psychogeriatrics, Somatic and Rehabilitation. The dataset includes information of 78 to 99 anonymized patients. The nature of the conditions the patients have is only generalizable since it is seen as personal data and was not included in the dataset. Therefore, we cannot be certain the results will be consequent since the researched sample might not represent the population.

Also, the machine learning model was trained to predict when a patient is getting out of bed, not in their bed. Hence when the model is used in real time when patients might get into their beds the results may vary. Configuring the correct settings in a real time program is necessary to get the best results.

Another thing to note is that the ground truth is based on the bed status. This algorithm is developed to calculate if a patient is in bed according to sensor data but not by having an actual observer in the patients’ room. This does affect the validity of the results, but this was known up front.

The methods used during this research are reliable because they were based on best practices like (insert cross validation, balancing dataset, desk research). The literature used was trustworthy because it consisted mostly of peer-reviewed papers.

One of the expectations of this research was to be able to make better predictions than Momo Medical could do as well as extend the prediction time to a minute. Both goals were achieved as expected and it was determined what the best machine learning model is for this application.

A big limitation of the research was the time factor. Due to uncertainties of getting the right results with neural networks this research was started with machine learning experiments. Unfortunately results with machine learning that were good enough to start experimenting with neural networks came in too late for it to be feasible to step over to neural networks. Because of this we excluded neural networks from the scope of this project.

Based on a study about human activity recognition [11] the expectation was that with the aid of machine learning it is possible to recognize and predict when someone is going to get out of bed. In the end the expectations were met since the machine learning model was able to make accurate predictions.

This research shows that a linear support vector machine using time shifted features is the best model to be able to make correct predictions for this application.

Momo medical can use this model to aid in preventing fall accidents among elderly and or revalidating patients in nursing homes. Due to the recall, accuracy and precision of the predictions personnel hopefully will not lose their trust in the use of the model when this gets implemented in real time.

**Conclusion**

The goal of this research was to answer the following question: “What is the optimal machine learning algorithm that can predict a patient leaving their bed, using the collected data from the BedSense (version 9) sensor created by Momo Medical?”.

Optimal was defined as the following:  
- The least number of false positives and false negatives.  
- Predicting as far ahead as possible.

The conclusion is that the optimal algorithm found was a Linear Support Vector Machine. This can be seen in the result scores for LSVM and Logistic regression. The accuracy and recall scores for the LSVM where higher compared to Logistic regression. For the purposes of this research the accuracy and recall score outweigh the precision score to predict with less false positives and negatives that a patient is leaving their predicted sixty seconds in advance.

**Recommendations**

There are a few recommendations that can be made to further improve the current model and its features. These are:

* Heartrate feature using the PE sensor
* Deep learning neural networks

**Heartrate feature using the PE sensor:**  
It is possible to use the piezoelectric sensors on the device to filter out the heartrate of patient[12, 13], this is valuable data for the heartrate of a sleeping person is different between stages of sleep and being awake. Being able to use the piezoelectric sensor data to map the heartrate of a person could prove valuable in correctly predict when a patient is leaving or planning to leave their bed at an earlier time then the current minute.

**Deep learning** **neural networks:**  
The data format used for our machine learning algorithm can benefit from the ability a deep learning neural network provides. Neural networks can make use of more data and does not require the data to be structured [14]. With the sheer amount of data that could potentially be given for predictions, all sensor data per time interval of around 100ms this is the time interval for measurements from the used sensors. Using this data, it is expected that a neural network would be the way to proceed on a larger scale in order to make predictions on the more minute movements within a time period of the last few minutes.

**Bibliography**

[1] World Health Organization. (2020, September 28). Life expectancy and Healthy life expectancy.  
<https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/life-expectancy-and-healthy-life-expectancy>

[2] Calcagnini, G., Biancalana, G., Giubilei, F., Strano, S., & Cerutti, S. (1994). Spectral analysis of heart rate variability signal during sleep stages. Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1252–1253. <https://doi.org/10.1109/iembs.1994.415418>

[3] Kavitha S, Varuna S, & Ramya R. (2016). A comparative analysis on linear regression and support vector regression. 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 1–5. <https://doi.org/10.1109/get.2016.7916627>

[4] Gandhi, R. (2018, July 5). Support Vector Machine — Introduction to Machine Learning Algorithms. Towards Data Science. <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>

[5] Bhavsar, H., & Panchal, M. H. (2012). A Review on Support Vector Machine for Data Classification. International Journal of Advanced Research in Computer Engineering & Technology, 1(10), 185–189.<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1039.2508&rep=rep1&type=pdf>

[6] Thabtah, F., Abdelhamid, N. & Peebles, D. A machine learning autism classification based on logistic regression analysis. Health Inf Sci Syst 7, 12 (2019). <https://doi.org/10.1007/s13755-019-0073-5>

[7] V. H. Cene and A. Balbinot, "Upper-limb movement classification through logistic regression sEMG signal processing," 2015 Latin America Congress on Computational Intelligence (LA-CCI), Curitiba, 2015, pp. 1-5, doi: 10.1109/LA-CCI.2015.7435940.

[8] K. Nurhanim, I. Elamvazuthi, L. I. Izhar and T. Ganesan, "Classification of human activity based on smartphone inertial sensor using support vector machine," 2017 IEEE 3rd International Symposium in Robotics and Manufacturing Automation (ROMA), Kuala Lumpur, 2017, pp. 1-5, doi: 10.1109/ROMA.2017.8231736.

[9] Stephanie Glen. "Forward Selection: Definition" From StatisticsHowTo.com: Elementary Statistics for the rest of us! Visited on (11/01/2021) <https://www.statisticshowto.com/forward-selection/>

[10] Powers, D. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. ArXiv, abs/2010.16061.

[11] Schrader, L.M., Toro, A.V., Konietzny, S.G., Rüping, S., Schäpers, B., Steinböck, M., Krewer, C., Mueller, F., Güttler, J., & Bock, T. (2020). Advanced Sensing and Human Activity Recognition in Early Intervention and Rehabilitation of Elderly People. Journal of Population Ageing, 13, 139-165.

[12] N. Bu, N. Ueno and O. Fukuda, "Monitoring of Respiration and Heartbeat during Sleep using a Flexible Piezoelectric Film Sensor and Empirical Mode Decomposition," *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Lyon, 2007, pp. 1362-1366, doi: 10.1109/IEMBS.2007.4352551.

[13] T. Klap and Z. Shinar, "Using piezoelectric sensor for continuous-contact-free monitoring of heart and respiration rates in real-life hospital settings," *Computing in Cardiology 2013*, Zaragoza, 2013, pp. 671-674.

[14] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, 2017, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629.