A feasibility study into alternative pressure air matrass detection through machine learning

Joris H.J. Geurts[a,b]

[a] Open University, Heerlen, the Netherlands
[b] Fontys University of Applied Sciences, Eindhoven, the Netherlands

Abstract

Alternating pressure air mattresses play an essential role in the prevention of decubitus. Their presence and smoothless operation reduces the risk of developing decubitus. Nursing homes in the Netherlands have residents with varying physical conditions, and have therefore varying types of mattresses. The beds can be equipped with a sensor board that measures pressure and vibrations continuously. This report presents the investigation in machine learning models that predict the presence of alternative pressure air mattresses based on the sensor information. Working with a data set of a single nursing home, good accuracy results have been obtained. Future investigations on explainability of the trained models, as well as testing on larger data sets is needed to improve the robustness of the model.

Keywords: decubitus, supervised learning, machine learning, neural network, feature extraction, alternating pressure air mattress, fast fourier transformation, force sensitive resistor, Momo Bedsense

Contents

1 Introduction 1

2 Goal 2

3 Data analysis 2
- 3.1 Data sets 2
- 3.2 Exploratory Data Analysis 2
- 3.3 Feature extraction 3
- 3.4 Data augmentation 3

4 Methodology and Implementation 4
- 4.1 Research methodology 4
- 4.2 Design 4
- 4.3 Implementation 4

5 Evaluation and Results 4
- 5.1 Training baseline models 4
  - 5.1.1 Data set, features and hyper-parameters 4
  - 5.1.2 k-Nearest Neighbor 4
  - 5.1.3 Decision Tree 5
  - 5.1.4 Random Forest 5
- 5.2 Testing baseline models 5
- 5.3 Training neural network 6

6 Conclusions and Discussion 6
- 6.1 Interpretation of test results 6
- 6.2 Discussion 7
- 6.3 Future research 7

1. Introduction

Caregiving institutions like nursing homes are challenged by a global aging population since their residents need constant monitoring and support. In the Netherlands, this situation is aggravated by a growing shortage of caregivers. AI can support nurses in tasks considered to take place remotely from direct patient interactions (1). One of the severe health care risks is decubitus. Decubitus ulcers, also known as pressure ulcers or bed sores, are localized damage to the skin and/or underlying tissue that usually occur over a bony prominence as a result of usually long-term pressure, or pressure in combination with shear or friction. Often, enough damage is done to create the basis for a decubitus ulcer after as little as 2 h of immobility, a situation which may be difficult to avoid if the patient must undergo prolonged surgery or remain bedridden. Primary prevention is to redistribute pressure by regularly changing the person's position. This can be achieved by an Alternating Pressure Air mattress (APAM) (2; 3; 4). APAMs generate alternating high and low interface pressure between the body and support, by alternating inflation and deflation of air-filled cells (5). Investigations have shown that an APAM is more effective in decubitus prevention than other mattresses (3; 5). In various care homes in the Netherlands, the company QCare provides APAMs. Figure 1 shows a typical sample. Its operation is as follows: every other row is inflating simultaneously and the rows in between are deflating. After a couple of minutes, the actions reverse. The frequency of inflating and deflating is configurable, but typically 4-12 cycles per hour (cph).

A sensor system that detects the presence or absence of an APAM relieves the task of caregivers, especially in situations when there is a change of occupants in a room, or when an APAM is switched off accidentally. Such a sensor board is offered by Momo Medical, see figure 2.

The Momo Bedsense has 4 FSR sensors (Force Sensitive Resistors, measuring pressure/weight), 6 PE sensors (Piezzo-Electric sensors, measuring movement like heartbeat and res-

Figure 1: Alternating Pressure Air Mattress (source: QCare)

Figure 2: Momo Bedsense and its placement (source: AndersWerkenInDeZorg)

piration rate) and 1 accelerometer (measuring the angle of the bed) (6).

Momo Medical has made a initial implementation of an APAM detection system (unpublished); its accuracy is 0.89.

This research is part of the Europees Fonds voor Regionale Ontwikkeling (EFRO) OPZuid project 'Predictie in de Verpleeghuiszorg' (7).

2. Goal

In the participating nursery home, a (major) part of the beds have a regular, i.e. static, mattresses, the others have an APAM. All beds have a Momo Bedsense, which sensor data is constantly collected, and stored.

It is expected that the collected sensor data contains enough information such that a machine learning model is able detect the presence or absence of an APAM.

We try to answer the following research questions in this paper:

1. Does a Bedsense sensor registration carry enough information to detect an APAM?

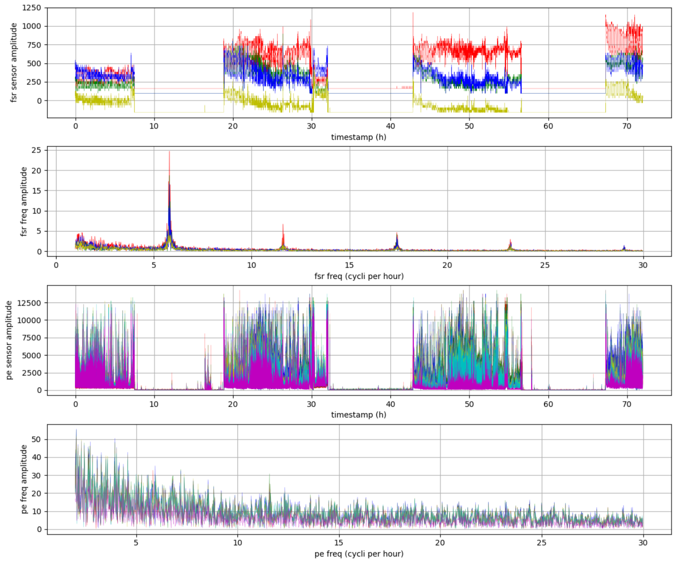

Figure 3: 72-hour time series of the sensors for an APAM (a) four FSR sensors (b) frequency spectrum of the FSR sensors (c) six PE sensors (d) frequency spectrum of the PE sensors

2. What detection accuracy can be achieved by a baseline machine learning model and a neural network (8)?
3. What hyper-parameter tuning is possible?

3. Data analysis

3.1. Data sets

A data set has been obtained from Momo Medical on 24 July 2024. It contains the bed sensor data of 860 beds of a nursing home in the Netherlands (province Noord-Brabant). The ID of the Bedsense has been replaced by its hash, and all other PII (Personally Identifiable Information, like names and room numbers) of the residents have been removed. The data is labeled with the type of mattress (either a regular mattress or an APAM). For each bed, the sensor information of the Momo Bedsense over a three day period is covered, with a sample time of one second. 96 of them have an APAM, 755 have a regular mattress, and in 9 samples the sensors are not switched on (so they are dropped in this investigation).

3.2. Exploratory Data Analysis

The inflating and deflating of the air cells in the APAM cause a weight shift of the nursing home resident. Our hypothesis is that the FSR sensors notice this as a regular pattern in sensor values. In this way, a distinction can be made with a regular mattress which doesn't yield such a pattern. With a Fast Fourier Transform a frequency spectrum of sensor time series can be obtained. (9; 10)

In figures 3 and 4 two examples are given for an APAM and a regular mattress. The FSR-frequency spectrum of the APAM shows a clear peak at 6 cph, together with its harmonics, while the FSR-frequency spectrum of the regular mattress does not. None of the PE-frequency spectra show a 6 cph peak. The PE

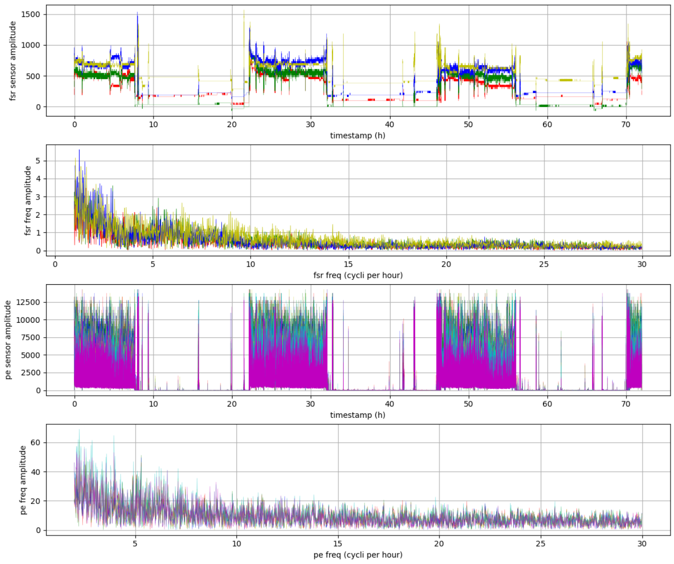

Figure 4: 72-hour time series of the sensors for a regular mattress (a) four FSR sensors (b) frequency spectrum of the FSR sensors (c) six PE sensors (d) frequency spectrum of the PE sensors

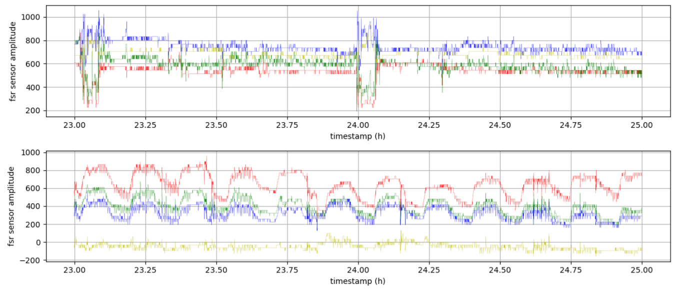

Figure 5: Zoomed in 1-hour time series of FSR sensors (a) a regular mattress (b) an APAM

sensor information is therefor not used in the remaining investigations.

Zooming in on a 1-hour time series (@23h) for the same examples shows indeed a repeating pattern of approximately 6 cph for the APAM, but not for the regular mattress; see figure 5.

Not all APAM samples show such a clear picture as in figure 3. The sensor values of one of the APAMs might highly resemble a regular mattress, see for example figure 6. In this figure, not all preferred peaks (i.e. the 6 cph and its harmonics) won't end up as features. So the machine learning algorithm will not be a trivial task.

3.3. Feature extraction

Based on the Exploratory Data Analysis (EDA), two major feature extractions have been performed.

For the baseline machine learning classifiers: sum the 4 FSR sensor values together (since we are not investigating a left-right position shift), perform a Fast Fourier Transform over the full 72h time period, followed by a peak detection with a certain prominence. For the largest peaks, their frequency and amplitude are stored as new features.

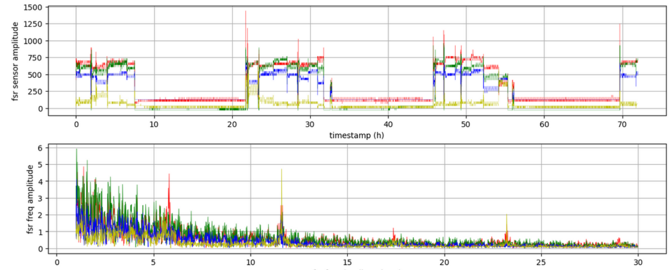

Figure 6: Exemplary APAM with less obvious peaks in the frequency spectrum pattern

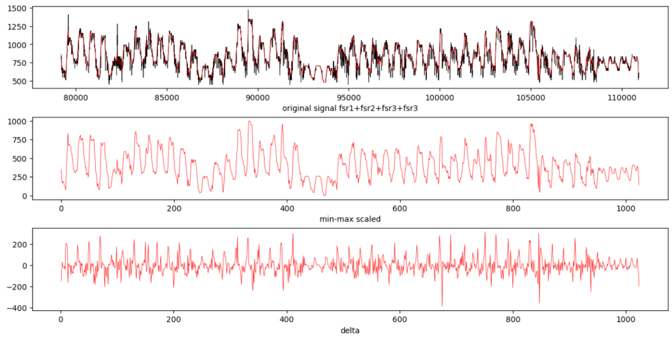

Figure 7: Sequential process of the downsampling feature extraction (a) original signal (black) and downsampled signal (red) (b) min-max scaled (c) difference-values (i.e. the final features)

In the used data set, a majority of APAMs have an (approx.) 6 cph, so the model might be unable to detect other cph's. Therefor additional features are calculated: a peak's relative frequency (to the first peak). Furthermore, the peaks' amplitudes relatively to the average amplitude of the FFT are calculated.

For the neural network classifier: The time series are limited to 400 elements (to limit the size of the layers). It is done by downsampling the original series with a period of 31 seconds (so the average sensor value of 31 seconds become one new time series element). A sample of 400 elements therefor covers $400 \cdot 31s = 3h26m40s$. For each original 72h time series, nine non-overlapping samples are created, each with 400 features. Since the sensor values vary over the samples (due to a.o. the occupants' weight), the values are minmax-scaled. Furthermore, offsets are removed by storing a feature's difference with its predecessor instead of its absolute value (example: two sawtooth occurrences 21,23,25,27,21 (etc.) and 400-402-404-406-400 (etc.) are both stored as features 0,2,2,2,-6 (etc.)).

Figure 7 visualizes the sequence of actions of this downsampling. The red line in plot (a) indicates that no essential information is lost in the downsampling. Please note that the x-values (representing the series-index) is 31x reduced in plot (b) and (c).

3.4. Data augmentation

Since the data set is unbalanced (755 regular mattresses vs. 96 APAMs), augmented APAM data is generated.

For the baseline machine learning classifiers: for each APAM sample three additional FFTs are calculated by splitting the 72h time series into the three 24h periods. They will slightly differ from the original sample because of physical movements of the occupant during the separate days.

For the neural network classifier: the time series are downsampled with other periods (being 27 and 23s, i.e. relative prime numbers to avoid repetitive patterns in the resulting downsamples). It has a further advantage that the new samples represent different cph's.

4. Methodology and Implementation

4.1. Research methodology

This research employs a quantitative approach to model the detection of alternating pressure air mattresses, as used in care homes in the Netherlands. The data is retrieved by Momo Bedsense devices, being in operation under all mattresses in said care homes. The company Momo collects that data and stores it for later use. In this investigation, exploratory data analysis is performed, which leads to meaningful feature extractions. Due to the unbalanced class distribution, data augmentation is performed by feature extraction over different time periods.

The data set is split over a training set and a test set (in a commonly used ratio 0.7/0.3).

The training of a classifier is done with a 5-fold cross-validation of the training set (where 5 is a commonly used value for cross-validation). This is repeated several times (with identical hyper-parameters but with different randomized partitioning of samples over the folds), and their results are averaged.

Testing on the test set yields various metrics, such as a confusion matrix, feature importance, ROC/AUC plots and a LIME analysis.

4.2. Design

The design is split into the following activities: (a) EDA, to characterize the data, (b) Fast Fourier Transformation, to investigate the feasibility of the resulting frequency spectrum (in fact an EDA activity as well) and the extraction of the peaks into new features including augmentation, (c) Downsampling, as a feature extraction and data augmentation for the neural network, (d) Baseline machine learning, to run models such as Decision Tree, Random Forest and Nearest Neighbor with various parameters, (e) Neural Network modeling, to train and run a Neural Network model with various constellations.

In selecting a Neural Network architecture for time-series analysis (11), a CNN has been chosen over an RNN. For RNN, samples with older timestamps have less influence on the predictions (which is beneficial in stock markets and weather forecasts), while in our investigation, all timestamps are equally contributing to the classification.

The neural network is constructed with the following consecutive components: (a) a Dropout (12), (b) a configurable number of hidden layers, each consisting of sequentially: a 1D Convolution with a configurable number of filters, a Normalization, a ReLU activation, (c) a Dropout, (d) a Global Average Pooling for final classification (13).

baseline classification	overall size	training	test
size	1111	777	334
regular mattresses	764	534	230
APAMs	347	243	104

Table 1: Used data sets for baseline classifiers

number of neighbours	accuracy
1	0.975
3	0.977
5	0.973
7	0.968
9	0.967

Table 2: Results of the training accuracy scores of knn models

4.3. Implementation

The models are implemented in Python with five Jupyter notebooks. The used libraries are: pandas, numpy, sklearn, matplotlib, seaborn, keras (14; 15; 16; 17; 18; 19).

The Jupyter notebooks and the data sets, including labels and extracted features, are available in a public git repository at https://gitlab.com/j.geurts/bedsense.

Methods are implemented to be configurable by their parameters, e.g. begin time, end time, downsample period, type of classifier, number of kernels/hidden layers. This facilitates further experimenting.

The execution of the Jupyter notebooks is done in the SURF Research Cloud environment (20), running Ubuntu 20.04 with two A10 GPU's; thanks to a Small NWO Grant.

5. Evaluation and Results

5.1. Training baseline models

5.1.1. Data set, features and hyper-parameters

The used data set is a concatenation of both original data set and augmented data set. The characteristics are depicted in table 1.

The features as used for training and testing are: the relative frequencies, relative amplitudes of the peaks, and the average FFT signal (which are extracted features as described in paragraph 3.3).

For each classifier model, various trainings are executed, each with its specific hyper-parameters. The range of values is extended such that an overfit is expected (i.e. that an increase of the model's complexity doesn't lead to an increase of accuracy any more). This is visible in the results below.

5.1.2. k-Nearest Neighbor

The k-Nearest Neighbor classifier is investigated for various settings of the number of neighbours. Table 2 shows the settings and the resulting accuracy of the training.

max depth	**accuracy**
1	0.931
3	0.963
5	0.968
7	0.966

Table 3: Results of the training accuracy scores of decision tree models

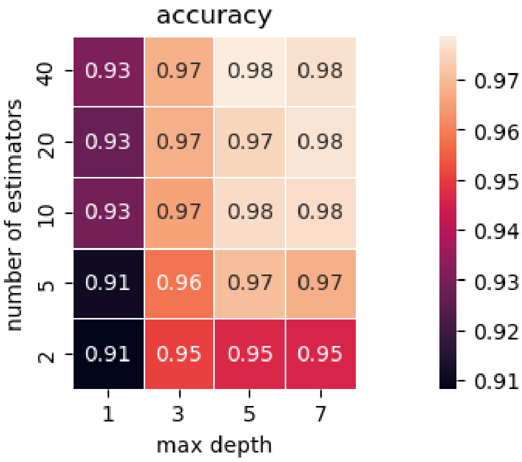

Figure 8: Grid search results of the training accuracy scores of random forest models

5.1.3. *Decision Tree*

The Decision Tree classifier is investigated for various settings of the maximum depth. Table 3 shows the settings and the resulting accuracy of the training.

5.1.4. *Random Forest*

For Random Forest, a grid search over the parameters max-depth and number-of-estimators is performed. Figure 8 shows the obtained accuracy values.

5.2. *Testing baseline models*

For each of the algorithm as evaluated above, the one with the best performance on the training is taken for testing its performance on the test set (i.e. unseen data). The results are collected in table 4.

In order to get a better insight of the operation of the models, auxiliary metrics are collected. For the sake of brevity, each metrics is only shown on one of the models.

The accuracy score (as mentioned in the sections above) is one of the metrics to charactery a learned model. Other metrics

model	**accuracy in training**	**accuracy on test set**
decision tree, max-depth=3	0.963	0.958
k-nn, nbr-neigh=3	0.977	0.979
random forest,max-depth=5, nbr-neigh=10	0.976	0.988

Table 4: Results of the accuracy score during training and on the test set for various baseline classification models

		Prediction		
		Positive	Negative	
Actual	Positive	true pos: 98	true neg: 6	*recall* : 0.94
	Negative	false pos: 1	true neg: 229	
		precision : 0.99		

Table 5: Confusion matrix of the test set results for a k-nn classifier with nbr-neighbours=3

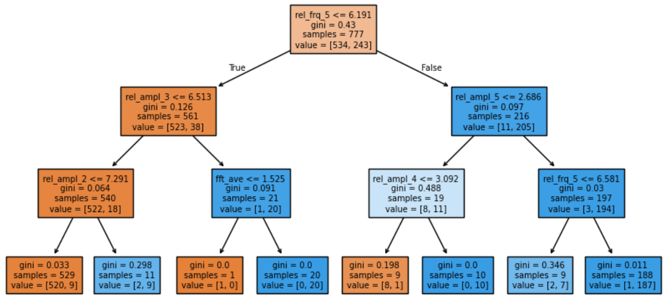

Figure 9: Internal tree of a trained Decision Tree with max-depth=3

are precision, recall and f1-score. They can be derived from a confusion matrix, as depicted in table 5. The corresponding f1-score is 0.96. A f1-score is particularly interesting when the class distribution of a data set is unbalanced.

Displaying the internal tree of a trained Decision Tree classifier (with max-depth=3) enhances the explainability of the model. Figure 9 shows such a tree.

Most classifier models provide the importance of the features during the training. Figure 10 shows a plot of the Random Forest with max-depth=5, n-estimators=5.

With the predicted probabilities of the test set samples, a ROC curve (Receiver Operating Characteristic) can be calculated; it shows the relation between the false positive and true positive rate for various thresholds. Figure 11 shows ROC curve

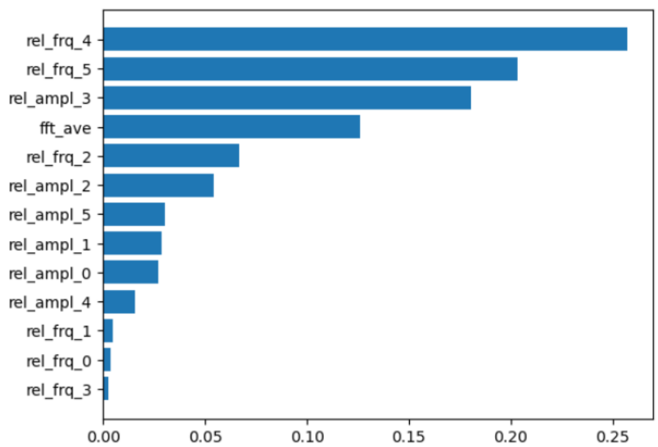

Figure 10: Overview of the importance of the involved features for a Random Forest with depth=5 and estimators=5

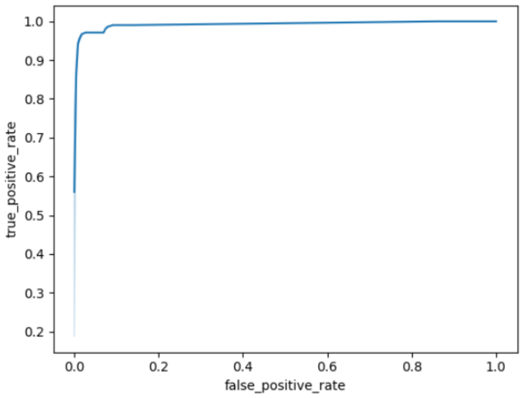

Figure 11: ROC curve for a Random Forest with depth=5 and estimators=5

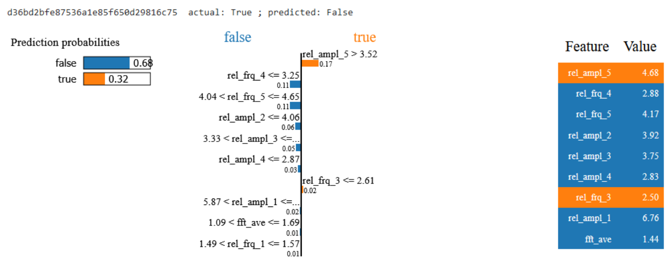

Figure 12: LIME-analysis of a false negative sample for a Random Forest with depth=5 and estimators=5

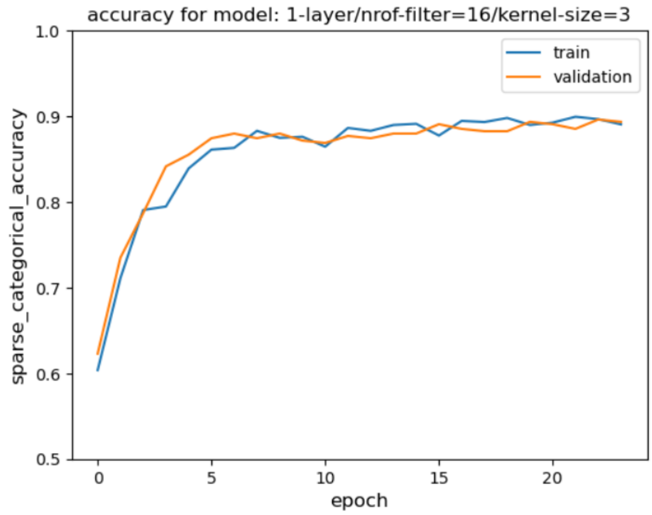

Figure 13: History over the epochs of achieved training and validation accuracies for a CNN model with settings: nbr-layers=1/nbr-filters=16/kernel-size=3

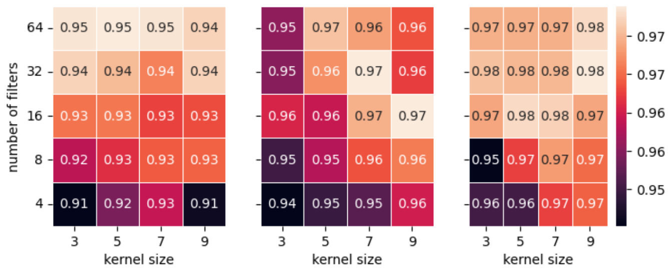

Figure 14: Accuracy results of the CNN grid search, split over the number-of-layers setting: (a) one convolution layer (b) two convolution layers (c) three convolution layers

of the Random Forest with max-depth=5, n-estimators=5. The corresponding AUC (Area Under the Curve) is 0.952.

LIME (Local Interpretable Model-Agnostic Explanations) (21) helps in explaining the predictions of a trained model locally around a single sample. Figure 12 depicts a plot of the LIME-analysis of a false negative sample in a Random Forest with max-depth=5, n-estimators=5. It shows how each feature contributed to the classification predictions.

5.3. Training neural network

The used data set is a concatenation of both the original data and the augmented data samples. The characteristics are shown in table 6.

Figure 13 shows the history of one training experiment. Since this figure shows that the results stabilize from epoch 12 onwards, for further experiments in this research the number of epochs is set to 12.

A grid search on the following hyper parameters is done: (a) the number of convolution layers, either 1, 2 or 3, (b) the number of filters per layer, either 4, 8, 16, 32, 64, (c) the kernel size, either 3, 5, 7, 9.

neural network classification	**overall**	**training**	**test**
size	6525	4567	1958
regular mattresses	5220	3654	1566
APAMs	1305	913	392

Table 6: Data sets as used for training and testing for the neural network

Figure 14 shows the results; for improved visibility of a three-dimensional grid, the data is split over the three settings for the number of layers.

6. Conclusions and Discussion

6.1. Interpretation of test results

With machine learning classifiers, APAMs can be detected by the sensor information as obtained from a sensor board like a Momo Bedsense. Furthermore, tuning the models' hyper-parameters influence the results according expectations, where overfitting can be avoided. The performed experiments show a high accuracy result value.

For the decision tree classifier, the model seems to overfit at a depth of 5. This is in line with the feature importance information, where the fifth feature's importance is approximately 5%. Initially, we expected that the decisions would focus on the first peak, being around 6 (cph), with rather high amplitude, combined with a low FFT average, as derived from pictures 3 and 4. It is surprising that the results showed that the fourth and fifth peaks are the most importance features (representing harmonics of approx. 24 resp. 30 cph). Since they hardly rise above the average FFT level, we fear that they don't really represent a key characteristic of the underlying signal. As a result, we are not yet fully convinced of the stability of the model. It resembles the urban legend about a tank image classifier model, where it turned out that the classification made the distinction between snowy/out-of-focus (Warsawpact tanks) versus glossy-advertisement (Nato tanks), instead of detecting the true tank characteristics (like size of the gun barrel, type of armour, etc.) (22).

The results of the Random Forest grid search are according expectations: an increase of estimators and depths initially increases the accuracy, but later gets overfitted. The top four features explain approx. 80% of the prediction.

The LIME analysis seems to be of little usefulness. Normally it indicates what level of change is needed for a feature value to switch to another prediction class (like in text sentiment analysis, where one more occurrence of a certain word turns the sample into e.g. a spam message). In our domain, it is not that obvious how the fourth or fifth harmonic can be influenced to improve prediction probabilities.

On the Convolutional Neural Network classifier, the performed experiments even show a higher accuracy. The third layers seems to overfit the data. In this research, it is not investigated what the trained features are in the filters of the various layers, i.e. if they represent key characteristics or accidental, overfitted artifacts.

6.2. Discussion

In a nursery home, an APAM is just one of the factors in decubitus prevention, besides (a.o.) skin inspection at regular intervals, balanced nutrition, amount of daily activities, and careful handling of the caregivers. The use of an APAM is triple-tradeoff between (a) effectiveness of the applied technology, (b) welfare of the resident, and (c) work load of the caregiver. After all: an APAM yields a proven relieve in skin pressure, but the constant movements of the air pressure chambers decreases the sleep quality of the resident, and intensifies the caregiver's task of helping a resident in- and out of bed.

The investigations in this report have been done with fully anonymized data. However, when applied in a larger decubitus prediction model (with the aforementioned factors), privacy issues become prominent. A resident's name and room number can be anonymized, but other information (even when techniques like Aggregation (23) have been applied) remains vulnerable for re-identification, e.g. by combining data sources.

6.3. Future research

Future research need be conducted into the following issues:

- In this research, a sample and its augmented children are treated as independent; so they might be divided over the training, validation and test sets during the experiments. It is to be investigated whether they are really independent, to avoid information leakage from training to either validation and/or test.
- Investigate if a prediction can be done based on shorter time series (e.g. based on a 30min time series), and establish the relationship between the length of the time series and the probability of correct detection.
- Perform a case study on various false positive and false negatives to find improvement opportunities for hyper-parameter tuning and/or additional feature extraction.
- Analyze the trained filters of the neural network (i.e. explain the model) to validate that the prediction is not based on coincidental effects.
- Perform exhaustive tests with larger data sets, over a longer time, from more nursery homes and with a larger variety of mattresses.
- Investigate how the current presented (offline) models can operate on real time data, and how it can be integrated efficiently in the daily flow of nursery homes.

References

[1] Kathrin Seibert, Dominik Domhoff, Dominik Bruch, Matthias Schulte-Althoff, Daniel Fürstenau, Felix Biessmann, and Karin Wolf-Ostermann. Application scenarios for artificial intelligence in nursing care: Rapid review. *J Med Internet Res*, 23(11):e26522, Nov 2021.

[2] Amit Gefen, David M. Brienza, Janet Cuddigan, Emily Haesler, and Jan Kottner. Our contemporary understanding of the aetiology of pressure ulcers/pressure injuries. *International Wound Journal*, 19(3):692–704, 2022.

[3] Cheryl Bansal, Ron Scott, David Stewart, and Clay J. Cockerell. Decubitus ulcers: A review of the literature. *International Journal of Dermatology*, 44(10):805–810, 2005.

[4] Wikipedia. Pressure ulcer — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Pressure%20ulcer&oldid=1264786341`, 2024. [Online; accessed 31-December-2024].

[5] Katrien Vanderwee, Maria H. F. Grypdonck, and Tom Defloor. Effectiveness of an alternating pressure air mattress for the prevention of pressure ulcers. *Age and Ageing*, 34(3):261–267, Mar 2005.

[6] Ilse Bierhoff, Hendrik Buimer, Bob Hofstede, Rachelle Rinzema, Tom van Hoesel, Johan Leeuw, and Henk Herman Nap. Eindrapportage bedsensor, Mar 2024.

[7] Stimulus Programmamanagement. Predictie in de verpleeghuiszorg. `https://www.stimulus.nl/opzuid-2021-2027/avada_portfolio/predictie-in-de-verpleeghuiszorg/`, Dec 2024. [Online; accessed 12-January-2025].

[8] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.

[9] Wikipedia. Fast Fourier transform — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Fast%20Fourier%20transform&oldid=1264174592`, 2024. [Online; accessed 31-December-2024].

[10] Henri J. Nussbaumer. *The Fast Fourier Transform*, pages 80–111. Springer Berlin Heidelberg, Berlin, Heidelberg, 1982.

[11] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. *CoRR*, abs/1611.06455, 2016.

[12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[13] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network, 2014.

[14] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[15] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

[16] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.

[17] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[19] François Chollet et al. Keras. `https://keras.io`, 2015.

[20] Jelle ten Hoeven. Surf research cloud. `https://www.surf.nl/en/services/surf-research-cloud`, Jan 1987. [Online; accessed 12-January-2025].

[21] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 1135–1144, 2016.

[22] Gwern Branwen. The neural net tank urban legend. `https://gwern.net/tank`, Sep 2011. [Online; accessed 24-January-2024].

[23] Suntherasvaran Murthy, Asmidar Abu Bakar, Fiza Abdul Rahim, and Ramona Ramli. A comparative study of data anonymization techniques. In *2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pages 306–309. IEEE, 2019.