

A Comparative Evaluation of Machine Learning Deployment Approaches in Real Term Environments using the Example of the Detection of Epileptic Seizure

Salima Houta

Fraunhofer Institute for Software and Systems Engineering, Germany

salima.houta@isst.fraunhofer.de

Abstract

The detection of epileptic seizures plays an important role in patient safety and therapy. Much research has been done in recent years to detect epileptic seizures using mobile devices. Although the variety of symptoms of certain types of seizures is challenging, progress has been made in identifying certain types of seizures. Machine learning is used in most work in an Experimental Environment. However, individual and situational aspects play an important role, especially in the detection of epileptic seizures. The improvement of seizure classification through machine learning in everyday life will play an important role in the further development of the technologies in the next few years. The EPItect project is researching the detection of epileptic seizures using an In-Ear sensor. A framework for machine learning for the Experimental and Real Term Environment was developed in the project. In this paper, we provide a comparative evaluation of different approaches to providing machine learning in the Real Term Environment.

1. Introduction

Epilepsies are among the most common neurological diseases worldwide. Depending on the degree of severity, affected persons can live a life with great restrictions on their autonomy. Characteristic symptoms are recurring epileptic seizures, which can be very stressful for the affected persons, relatives and carers due to the unpredictability of the time at which seizures occur, as well as the impairment of consciousness and the loss of control over different body functions. Among other things, the mortality of people with epilepsy is increased by a factor of 2-3 due to severe epileptic seizures (e.g., failure of the respiratory center) and seizure consequences (e.g., accidents, suffocation) [1][2]. The early detection of seizures can possibly help to take of appropriate safety measures for the person concerned and to reduce sudden unexpected death in

epilepsy (SUDEP). In addition to such early detection, an accurate recording of the seizures also helps in the individual planning of the therapy. In order to reduce seizure frequency or, at best, to achieve complete seizure control, a central component of medical treatment is the suppression of seizures by medication. Proper documentation of epileptic seizures by patients or relatives plays an important role in coordinating therapy. The documentation can be done on paper or web-based seizure calendars (e.g., EPI-Vista) [3]. However, previous studies show that approximately 50% of seizures are not documented and approximately two-thirds of patients provide incorrect data [4][5]. The main reasons for the faulty seizure documentation are, for example, the disturbed perception of one's own seizures, amnesia for seizure or later forgetting of the seizure that has taken place. The seizure documentation by relatives or caregivers is also prone to failure as relatives do not notice symptom-poor epileptic seizures [6]. Research on seizure detection using mobile devices has increased significantly recently. Pearl et al. [7] provide a literature review of seizure detection devices and their effectiveness in different types of seizures. In particular, success has been achieved in the detection of generalized tonic-clonic seizures (GTCS), which are associated with a bilateral, convulsive tonic contraction followed by generalized clonic muscle contractions [8], [9], [10], [11]. A combination of several modalities helps to achieve higher sensitivity and a lower false alarm rate. [9].

In this article we provide an evaluation for machine learning deployment approaches using the example of machine learning to detect epileptic seizures. The paper is structured as follows: Section 2 contains a brief literature overview of deployment approaches for machine learning projects. Section 3 describes the EPItect project including the components of the technical solution as well as the developed machine learning framework. Section 4 describes the method on which our assessment is based. You will also find information on the evaluation criteria used. Section 5

presents different approaches to the provision and use of machine learning with regard to the usage perspective in Real Term Environments and discusses advantages and disadvantages. Finally, section 6 describes the conclusion and some thoughts regarding future work.

2. Related Work

There is a patent for providing machine learning models [12]. However, no comparative evaluations of different deployment and operating models for machine learning could be found in the literature.

3. EPItect

The focus of the project EPItect is to develop a non-invasive sensor system, which reliably detects those bio signals that enable automated detection of epileptic seizures. The sensor is placed in the external auditory canal (similar to a classic hearing aid). The data are made available to selected persons via mobile devices. In this way, the personal environment can also be included if necessary. This specially developed in-the-ear sensor technology and a networking infrastructure based on (inter-)national communication standards (e.g., Integrating the Healthcare Enterprise, Elektronische Fallakte 2.0, HL7 Fast Healthcare Interoperability Resources) are the basis for several IT applications, which are also integrated into the existing medical-nursing processes. In addition, the vital parameters and events recorded by the In-Ear sensors and the data recorded by the mobile application, such as context information on seizures (e.g. mood diary, medication intake, activities), provide valuable data for clinical research. A sensor that can detect epileptic seizures ideally provides precise data on the frequency and severity of seizures. Context information about a recognized epileptic seizure can be recorded in the mobile patient application (e.g. activity before the seizure, consumption of alcohol, feelings). Additional documentation in the mobile application, which also represents a context of the seizure, is supported (e.g. medication intake, activities, mood). The inclusion of context information in the modeling supports the research of individual therapy methods, for example the suitability and effectiveness of a drug or a combination of drugs in a certain patient group or the identification of factors that trigger seizures.

3.1. Architecture

The components of the technological solution are shown in Figure 1: the In-Ear sensor (EPISENS), the mobile application (myEPI), the portal (EPICASE

Portal) and the networking infrastructure (EPICASE Infrastructure).

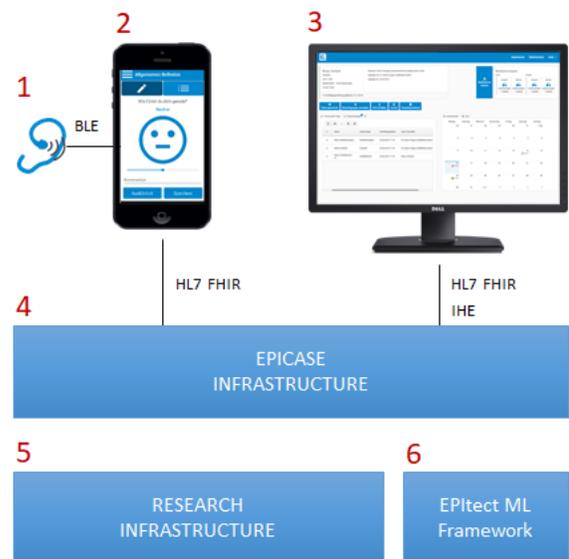


Figure 1. Components of the technological solution.

EPISENS (1) includes sensors to optimize seizure detection and seizure counting. It sends vital data, raw data and alarm events via Bluetooth Low Energy to the myEPI App. myEPI App (2) is a mobile companion for the patient. The app includes an alarm module. Upon receipt of alarm events, selected persons (e.g., parents or partners of an affected person) should be informed. The patient can use a simple action on the smartphone to confirm the seizure event or classify it as a false alarm. This information is used in the next step to optimize the algorithms developed. In the app, the patient also has the opportunity to collect additional data (contextual information on seizure events, mood, medication administration, side effects). He can selectively release data for doctors or relatives. The data is transmitted securely via the EPICASE infrastructure (4) and can be viewed by relevant actors via the EPICASE portal (3). The EPICASE portal is a case based communication portal for patients as well for professional and informal caregivers. It enables exchange of treatment-relevant data (e.g., medication order, medication administration, seizure documentation, diagnosis). The EPICASE infrastructure connects the IT applications. It is based on international standards and fully complies with data protection and data security requirements. The project EPItect also provides a research infrastructure (5) for pseudonymization, data capturing and integrating and storage of case based generated data. The integrated data is the basis for our machine learning framework (6).

3.2. Machine Learning Framework

3.2.1. Experimental Environment. The Experimental Environment is used to develop and optimize seizure detection models using controlled conditions for the duration of the clinical studies, which are carried out at the participating specialist clinics. 200 patients at the University Hospital Bonn and 26 children at the Department of Neuropediatrics of the University Kiel have been recruited to test the In-Ear sensor. For these patients Video-Electroencephalography (EEG), ECG (Electrocardiogramm), and In-Ear sensor data (Photoplethysmography (PPG), heart rate, temperature, acceleration data) were collected over an average period of four days. The In-Ear sensor measures the 3D acceleration with a sampling rate of approximately 50 Hz. Heart rate (HR) is measured once per second based on the previous 6 seconds of the PPG signal. A quality indicator is given for the calculated heart rate. Based on the Video-EEG physicians have recorded occurred seizures (period, type of seizure). 552 seizures were recorded. The main steps of the ML Framework in the Experimental Environment are: domain and data understanding, data preprocessing, feature extraction, model selection and evaluation. Domain understanding includes understanding the problem and the goal of the modelling. This means for example the understanding of the symptoms of an epileptic seizure. The understanding is obtained by literature review and by involving neurological experts and also affected persons in the project. While neurologists provide information on relevant features from the perspective of medical scientific knowledge, patients can report on factors that are subjectively perceived in everyday life (mood swings in relation to seizures, activities, symptoms). A main step is to understand the data, which plays a major role. A detailed analysis of the data and the understanding of the goal would help to avoid later problems. The project integrates a variety of data sources: ECG, EEG, PPG, vital data (heart rate, temperature), seizure labeling and classification, patient meta data (gender, age). The data preparation includes tasks like selecting, cleaning, integrating and formatting data. For example, we remove the HR data with a low quality value before feature calculation. The data integration and formatting are critical tasks for multi-modal approaches. The feature extraction depends on the data and the objective of the model. A feature set can be selected by experts opinion or feature selection algorithms. We test several different feature selection approaches considering their selected classification models. The next step is the modelling

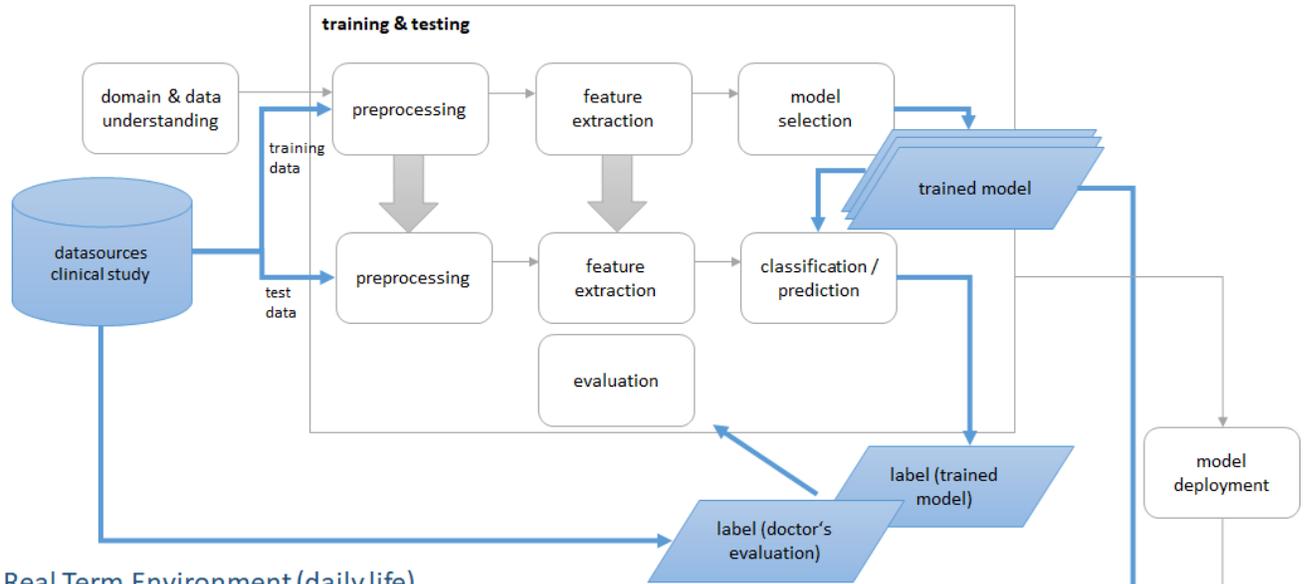
process. To select models for the experimentation, a literature review and the identification of similar research activities that have previously been successful are required. Each selected model can be trained on according features and feature sets. A continual evaluation and adjusting of features and models will identify the best model. It is important to preserve the order of the training data for the seizure detection problem, so that upcoming classification are based on previous results. From this we get a new classification model. To evaluate the trained model, one would classify match results into seizure and non-seizure and then determine sensitivity, specificity, positive and negative predictive value. In order to validate the performance of our models we use the method proposed by Beniczky and Ryvlin, which defines which outcome measures are to be considered to estimate the accuracy and performance of seizure detection [13]:

- Sensitivity number (and percentage) of all detected seizures/number of all seizures recorded during the study
- False alarm rate: the number of false alarms per 24 hours
- Detection latency: time from seizure onset to the detection time

In addition, we use the Receiver-Operating-Characteristic (ROC) curve to evaluate the classifier performance [21]. The representation is performed using the confusion matrix to estimate the performance of learning algorithms and the generated classifiers. The confusion matrix records the correct and misclassified features for each class. The ROC curve can be used to find the best possible value of a parameter [14] and to assess the trade-off between sensitivity and specificity [15]. The ROC analysis allows the rating of the classifier performance to be independent and complete rather than just accuracy.

3.2.2. Real Term Environment. Once a trained model is accepted, it is made available for use (Real Term Environment). For our project, this means deploying the models to mobile devices and sensors. Further training of the models, for example to take into account individual circumstances of the patient, is possible if the mobile devices and sensors provide good computing power and sufficient memory. Unlike the Experimental Environment, the verification of the model is done by the patients' labeling of alarm events which are triggered by the trained model. In order for the trained model to be accepted by the user, it is

Experimental Environment



Real Term Environment (daily life)

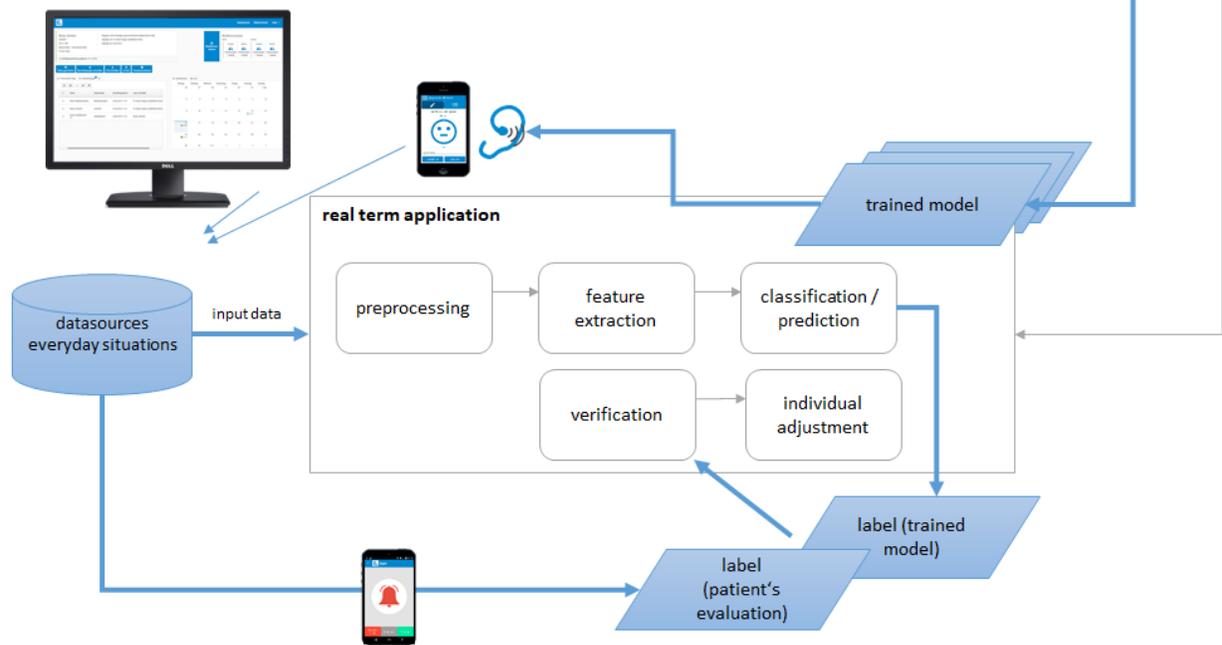


Figure 2. The ML-Framework.

important that there is no high amount of false alarms. In a survey of patients and care environment (n=305), we found that on average a maximum of 2/10 false alarms are accepted. The activities in everyday life (for example: sports, activities that trigger emotional states such as excitement) have a great influence on the signals and possibly on the applicability of the trained model. Further training of the model in a Real Term Environment is therefore an important task.

4. Method

Our assessment is based on the multi-criteria evaluation method. In the literature, important factors and corresponding outcome measures of results are suggested, which, in addition to the false alarm rate and sensitivity, are required to estimate the accuracy and performance of seizure detection devices [13]. These make reference to clinical studies; however, relevant criteria for evaluating deployment methods can also be derived from this. The continuous data acquisition and transmission plays a major role and should be ensured in order to avoid data loss. In addition, patient safety as well as data protection and data security play a major role in the transmission of patient data. The detection latency of seizures is especially important in the early detection of seizures. Dealing with the technologies as well as the effects of the technologies on daily life should also be considered.

In summary several criteria were defined to assess different deployment and operation models of machine learning: classification performance, memory, battery life, goal of machine learning (detection, prediction), data protection and data security, complexity of the system architecture, data sources to be integrated, integrability in the daily life an in organizational processes, costs, error susceptibility (e.g. data loss).

In the first step, the method is limited to the definition of the criteria (groups). The definition and application of metrics is planned for future work.

5. Results

5.1. Integration Approaches for ML in the Real Term Environment

The deployment of ML in Real Term Environment is very challenging in terms of ensuring performance and storage capacity, energy efficient transmission, network architecture, traffic engineering, maintaining strict security regulations and data management. Scenarios were created that show how the trained models can be used and adapted in the Real Term Environment. In view of the necessary situation-specific and patient-specific

Table 1. Scenario: classification based on the model including data acquisition for the training.

<i>Steps</i>	
1	Calculate vital signs from raw data
2	Feature extraction from vital signs
3	Classification of windows based on defined features
4	Caching the labeling result and data basis
5	Notification of the patient
6	Labeling by the patient
7	Save the label
8	Capturing context information by the patient
9	Integration into the training data base

Table 2. Scenario: Manual classification by the patient.

<i>Steps</i>	
1	Suffering a seizure
2	Document the seizure manually
3	Save the label
4	Capturing context information by the patient
5	Capture vital signs to corresponding time window
6	Integration into the training data base

further training of the models, aspects of capturing new training data from everyday life are also taken into account. Table 1 summarizes the scenario steps for the classification and collecting of data regardless of the selected technology approach.

Data on seizures not recognized by the model should also be recorded and included in the data basis for training according to the scenario steps described in Table 2.

The following approaches show technological approaches for the implementation of the ML. There is a rough differentiation in classification and training.

5.1.1. ML on In-Ear Sensor. This approach follows the application and training of models on the In-Ear sensor. Initially, a model trained and validated on the server (in the cloud) is deployed. Based on the vital parameters, the sensor itself detects the seizures and alarms independently of the other components such as patient's mobile application. To do this, the sensor must be expanded by memory and computing power and must have alarm functions (e.g. output of sound). In addition the patient must be able to use the interaction interfaces of the sensor to label the data (e.g. switch) (seizure yes

/ no). For the further training of the models, the labeling of the patient must be stored in the sensor memory. It is also possible to do the labeling via the mobile application and send this information via Bluetooth to the sensor. The advantage here is that the patient or relatives can give further information (time intervals, seizure type).

5.1.2. ML on Mobile Application. This approach follows the application and training of models on the mobile application. Initially, a model trained on the server is deployed on the mobile application. The sensor measures the vital signs and sends them via Bluetooth to the mobile application. The mobile application classifies windows as seizures or non-seizures. The alarm is triggered by the mobile application and can cause various actions depending on the patient's configuration settings (e.g. sound, sending an SMS). Labeling the event via the mobile application as seizure, no seizure or unsure if seizure ensures that the models can be trained further in the mobile application.

5.1.3. ML on the Server. If the entire machine learning runs on the server (in a cloud), all input parameters required for classifying windows and training models must be transferred to the server. The vital signs are sent via Bluetooth to the mobile application and transmitted from there using long-range technologies (e.g. WiFi) to the server. The server classifies a window as a seizure or not. If it is classified as a seizure, the server sends a notification to the mobile application. The patient confirms or rejects the presumption of a seizure. He can correct temporal information proposed by the server. The labeling and further seizure documentation on the mobile application is sent to the server.

5.1.4. Hybrid Approaches of Machine Learning in Real Term Environment. In the case of hybrid approaches, the classification and training are carried out on separate infrastructure components. Possible approaches are shown in Table 3.

Table 3. Hybrid Approaches.		
	Classification	Training
a	Sensor	Server
b	Sensor	App
c	App	Server

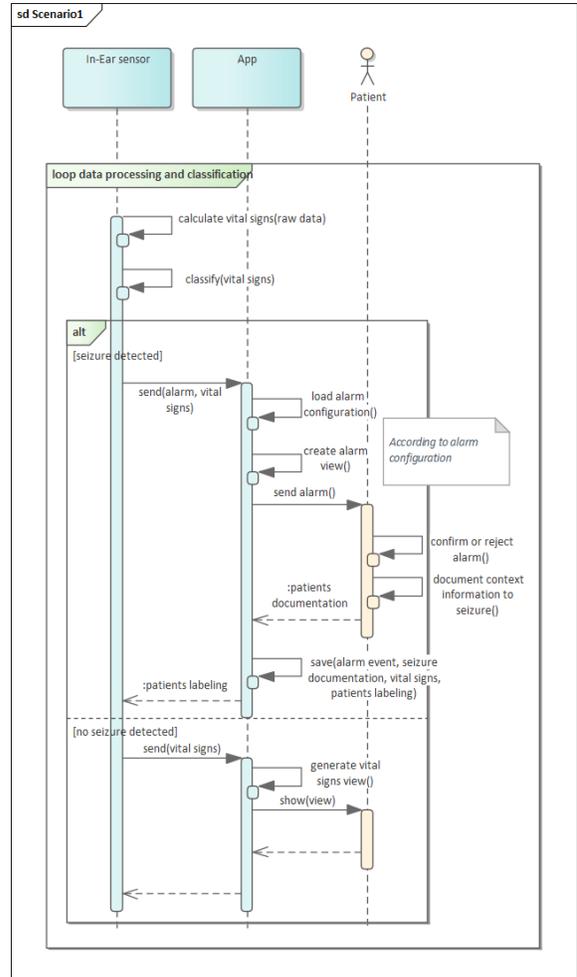


Figure 3. Scenario 1.

5.2. Comparison of Approaches

The classification on the sensor offers the advantage that an autonomous alarm is possible. The patient has more freedom of movement (e.g. when swimming). The classification can be done where the seizure takes place. However, this is also countered by the performance of the processors on the sensor, which can impair the classification performance. In addition to the performance, the storage capacity can also be a limiting factor for training depending on the classifier selected. The biggest disadvantage according to the current state of technology is the high energy consumption. Although the self-sufficient alarm system greatly increases usability, the short battery life limits this in turn. If the training does not take place on the sensor, but on external components with more memory and performance capacity, interfaces for the update of the model must be implemented in order to enable deployment of the trained model. Organizational and role concepts are completely lacking here and should be discussed in an interdisciplinary and user-oriented manner.

Performance and memory are higher for classification and training on the mobile application, but battery life is still a limiting factor. Context information that the patient records in the mobile application is directly available and can be taken into account as a feature (e.g. taking medication, well-being). In addition, configuration by the patient (e.g. when and how an alarm should be triggered) can be more easily linked to the model. Maintaining the Bluetooth connection between the In-Ear sensor and the mobile application is important so that the input parameters for the classification arrive on the app promptly and without data loss.

The classification and training of models on the server has the advantage of performance and memory. Because of the storage for data processing is very fast and cheap. Context data on the seizure, which was transferred to the server via the patient's mobile application, can also be included in the training of the models as a feature. The limiting factor of the battery is eliminated. However, the complexity of the solution is high and prone to errors and depends on the connection between the IT components. Data loss can occur. The topic of data protection and data security is also becoming more complex due to the distributed storage locations and the data transmission between the various IT components.

Hybrid solutions offer the possibility to combine advantages of different approaches, but require the interface specification between the components. Figure

3 shows a hybrid approach in which the classification is carried out on the sensor and the models are trained on the mobile application. Figure 4 shows how to deal with undetected seizures. The training data are enriched with further training data via the mobile application. Depending on the chosen approach, these can be used in the app (on the server) or on the sensor for training the model.

The choice of an approach depends on various parameters for example latency of detection and information, ease of use, traceability and frequency of adaption of the model.

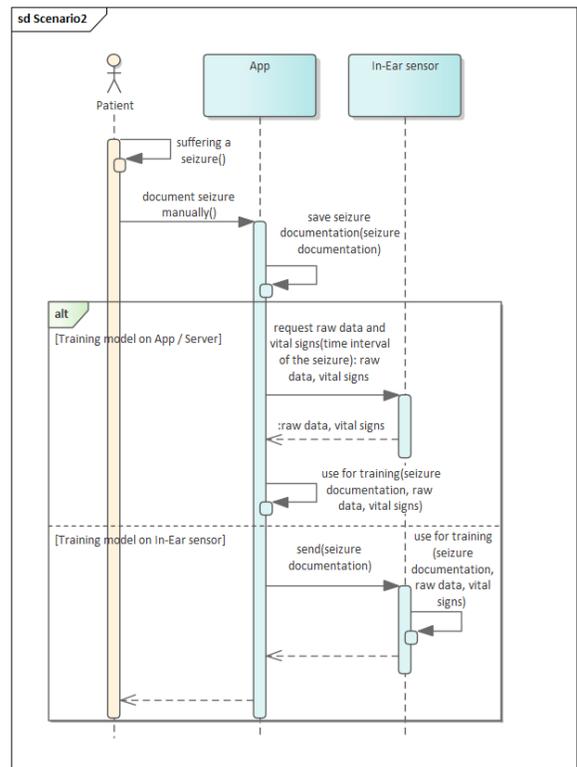


Figure 4. Scenario 2.

6. Conclusion and Future Work

The use of machine learning with mobile devices in healthcare is unstoppable. Most projects are still working in Experimental Environments. With the advent of digitization, which is also being enforced by legislation in many countries, the topic of deployment and operating models will become increasingly important. Using the example of the detection of epileptic seizures, we showed in our paper how possible deployment models can be designed. We compared them based on multiple criteria.

The results represent a first entry point from a theoretical-conceptual perspective. The elaboration of

the solution approaches is planned in future work. We want to further refine the evaluation method and define suitable performance metrics. Practical tests should also be carried out in order to actually be able to make statements on technical criteria. The topic of regulatory regulations also requires a precise analysis (e.g. Medical Device Regulation) and must be taken into account in the design of deployment processes.

7. Acknowledgment

The author acknowledges the financial support by the Federal Ministry of Education and Research of Germany in the framework of EPItect (project number 16SV7482).

References

- [1] L. Forsgren, E. Beghi, A. Oun, and M. Sillanpää, "The epidemiology of epilepsy in europe—a systematic review," *European Journal of neurology*, vol. 12, no. 4, pp. 245–253, 2005.
- [2] M. Thom, "Neuropathology of epilepsy: epilepsy-related deaths and sudep," *Diagnostic Histopathology*, vol. 25, no. 1, pp. 23–33, 2018.
- [3] G. Rabending and U. Runge, "The electronic treatment diary for epileptic patients (epivista[®]) - a new instrument for therapeutical treatment," vol. 7, pp. 273–280, 2007.
- [4] D. Blum, J. Eskola, J. Bortz, and R. Fisher, "Patient awareness of seizures," *Neurology*, vol. 47, no. 1, pp. 260–264, 1996.
- [5] C. Hoop, A. Poepel, and C. E. Elger, "Epilepsy: accuracy of patient seizure counts," *Archives of neurology*, vol. 64, no. 11, pp. 1595–1599, 2007.
- [6] A. C. Inan, M. A. Montenegro, S. Jacob, K. Eck, C. Chiriboga, and F. Gilliam, "Seizure frequency in children with epilepsy: factors influencing accuracy and parental awareness," *Seizure*, vol. 18, no. 7, pp. 524–529, 2009.
- [7] A. Ulate-Campos, F. Coughlin, M. Gainza-Lein, I. S. Fernández, P. Pearl, and T. Loddenkemper, "Automated seizure detection systems and their effectiveness for each type of seizure," *Seizure*, vol. 40, pp. 88–101, 2016.
- [8] A. Van de Vel, K. Cuppens, B. Bonroy, M. Milosevic, K. Jansen, S. Van Huffel, B. Vanrumste, L. Lagae, and B. Ceulemans, "Non-eeG seizure-detection systems and potential sudep prevention: state of the art," *Seizure*, vol. 22, no. 5, pp. 345–355, 2013.
- [9] F. S. Leijten, D. T. Consortium, J. van Andel, C. Ungureanu, J. Arends, F. Tan, J. van Dijk, G. Petkov, S. Kalitzin, T. Gutter, *et al.*, "Multimodal seizure detection: A review," *Epilepsia*, vol. 59, pp. 42–47, 2018.
- [10] J. B. Arends, "Movement-based seizure detection," *Epilepsia*, vol. 59, pp. 30–35, 2018.
- [11] E. G. Gutierrez, N. E. Crone, J. Y. Kang, Y. I. Carmenate, and G. L. Krauss, "Strategies for non-eeG seizure detection and timing for alerting and interventions with tonic-clonic seizures," *Epilepsia*, vol. 59, pp. 36–41, 2018.
- [12] J. Sirosh, M. K. Bulusu, V. Narayanan, R. Bhattacharya, S. Shoroff, P. Ardila, and A. Billing, "Easy deployment of machine learning models," May 26 2016. US Patent App. 14/554,413.
- [13] S. Beniczky and P. Ryvlin, "Standards for testing and clinical validation of seizure detection devices," *Epilepsia*, vol. 59, pp. 9–13, 2018.
- [14] S. Chandra, R. Bhat, H. Singh, and D. Chauhan, "Detection of brain tumors from mri using gaussian rbf kernel based support vector machine," *IJACT*, vol. 1, no. 1, p. 46, 2009.
- [15] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*, vol. 1. Springer series in statistics New York, USA, 2001.