

# SUPPORTING UTILITY MAPPING WITH A DEEP LEARNING DRIVEN ANALYSIS TOOL

Christian Versloot, Maria Iacob and Klaas Sikkel

## ABSTRACT

*Utility strikes have spawned companies specializing in providing a priori analyses of the underground. Geophysical techniques such as Ground Penetrating Radar (GPR) are harnessed for this purpose. However, analyzing GPR data is labour-intensive and repetitive. It may therefore be worthwhile to amplify this process by means of Machine Learning (ML). In this work, harnessing the ADR design science methodology, an Intelligence Amplification (IA) system is designed that uses ML for decision-making with respect to utility material type. It is driven by three novel classes of Convolutional Neural Networks (CNNs) trained for this purpose, which yield accuracies of 81.5% with outliers of 86%. The tool is grounded in the available literature on IA, ML and GPR and is embedded into a generic analysis process. Early validation activities confirm its business value.*

**Keywords:** Utility mapping; ground penetrating radar; intelligence amplification; machine learning; convolutional neural networks; business value

## INTRODUCTION

With the length of utilities exceeding 1.7 million kilometers, the Netherlands has a complex underground infrastructure (Rijksoverheid, 2015). Utility strikes occur in approximately 5.7% of all excavation work (Rijksoverheid, 2015). This equals 33,000 incidents annually, that is once every 3–4 minutes. Annual damages exceed 25 million Euros (Rijksoverheid, 2015). The risk of strikes can be mitigated by inspecting the underground a priori. Unsurprisingly, therefore, a business model has emerged for companies performing

---

Smart Industry - Better Management

Advanced Series in Management, Volume 28, 171–189



Copyright © 2022 Christian Versloot, Maria Iacob and Klaas Sikkel

Published by Emerald Publishing Limited. This work is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this book (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>.

ISSN: 1877-6361/doi:10.1108/S1877-636120220000028011

underground mapping. Ground Penetrating Radar (GPR) is used by analysts to detect underground utilities non-destructively (Cassidy & Jol, 2009). GPR analysis tasks are however reported to be repetitive and cost-intensive. This partially occurs due to the scarcity of GPR analysts and the steep learning curve preceding one's mastery. This chapter addresses this problem by designing and developing an intelligent system capable of supporting GPR analyses by means of Machine Learning (ML) models. We identify well-performing ML algorithms for predicting the material type of underground utilities detected with GPR, in response to an open automation problem posed by our industrial partner, Terra Carta. In doing so, we explicitly take an Intelligence Amplification (IA) approach in which the analyst is not replaced. Rather, through our smart ML agent, our goal is to amplify the analyst's intelligence. This allows analysts to quickly identify the material types of simple objects while using more creative, human intelligence-based approaches for complex ones. The main argument for this approach is that it reduces substantially the repetitiveness of the analyst's work and makes on-the-shelf knowledge available repetitively, while creating a true human-machine symbiosis and smart working environment. Our research has high applicability in industries such as critical infrastructures (e.g. gas, water, communication) and oil industry. It constitutes a typical example of business process improvement within the Industry 4.0 (I4.0) paradigm: two core I4.0 technologies, smart sensing and ML are combined with the goal of achieving smart working in which the worker is augmented with AI, and the nature of work changes (also one of I4.0's main goals). This work further contributes to science and practice in multiple ways. First, it introduces IA to the literature on GPR analysis by means of the intelligent system. Second, it designs and validates three novel classes of ML algorithms. Third, it presents a review of the literature on ML for underground mapping. Fourth, it embeds the system into a general human-in-the-loop analysis process by means of an architectural model. Fifth, it presents early validation comments demonstrating industry demand. As mentioned earlier, the chapter has emerged in cooperation with an industry partner. The chapter is structured as follows. Section 'Background' presents the literature reviewed. Section 'Research Methodology' discusses the methodology used. Section 'ML Models Underlying the Intelligent System' presents the design of the novel ML algorithms. Section 'ML Performance' demonstrates model performance. Section 'ML Driven Intelligent System' discusses the design of the intelligent system. Section 'Discussion' discusses and explains our findings. Finally, Section 'Conclusion' concludes this chapter.

## BACKGROUND

### *Origins of Excavation Damage*

Utility strikes pose a significant problem to Dutch construction given the costs and dangers involved. There are multiple causes for this problem:

- Quickly rebuilding damaged infrastructure was the primary concern after World War 2 (Eng, 1987). No central registry for utilities existed then, leaving many unregistered.
- The accuracy of methods for geospatial positioning was lower compared to today's ones. This sometimes resulted in large deviations between registered and actual utility positions.

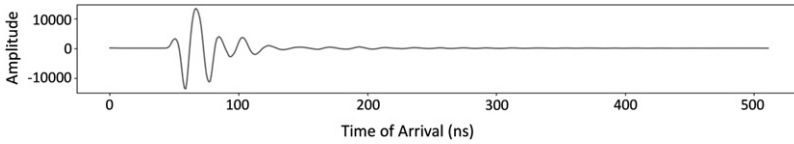


Fig. 1. A-Scan.

- Utilization of registries, which emerged in the late 1960s, remained optional until 2008. Only then, the government introduced legislation requiring their use (Kadaster, 2008; Rijksverheid, 2019).

Today, the registries are centralized into the KLIC utility registry. Here, all excavation work must be reported. In return, the reporter will receive all available information regarding utilities near their construction site. Consequently, as any request is forwarded to network operators, they are required to register their utilities and maintain accurate maps. Additionally, to protect critical utilities, network operator experts must provide physical identification when excavation work is performed near them. Although much progress has been made, inconsistencies from the past ensure that incidents continue to occur today (Rijksverheid, 2015).

### *Underground Mapping and GPR*

Mitigating the risk of utility strikes is possible through a priori analysis. Companies specializing in underground mapping have harnessed technologies such as the GPR for this purpose. A GPR is equipped with antennas transmitting and receiving electromagnetic waves and is moved over the Earth's surface. When waves propagate into the subsurface medium, they are echoed back when they hit objects buried in this medium (Cassidy & Jol, 2009). These are subsequently received by the GPR. This behaviour allows one to analyze the underground non-destructively. In fact, GPR is the de facto standard method used in underground mapping today.

#### *Visualizing GPR Data: A-Scans and B-Scans*

Echoes received can be visualized in A-scans (Scheers, 2001). In those, the amplitudes of the echoes are plotted against time of arrival (ToA), often in nanoseconds. Analysts can already derive certain object characteristics from A-scans. For example, they can identify object depth and, possibly, the nature of object contents. Fig. 1 presents an A-scan. Since a GPR moves horizontally, consecutive A-scans produce a richer image called B-scan (Scheers, 2001). In those, the horizontal axis represents the time domain across multiple A-scans. The vertical axis contains A-scan echo backscatters. Fig. 2 presents an exemplary scan. On B-scans, underground objects are visible as hyperbolae. This signature results from spherical GPR wave emission. ToAs are longer first, but get shorter when the operator gets closer to the object. ToA is shortest when the GPR is directly above it. Moving away again produces longer ToAs, resulting in the characteristic hyperbolic signature. Analysts can derive richer insights from B-scans, especially when they can combine GPR data with additional information sources like the KLIC registry. This makes the utility strike problem manageable.

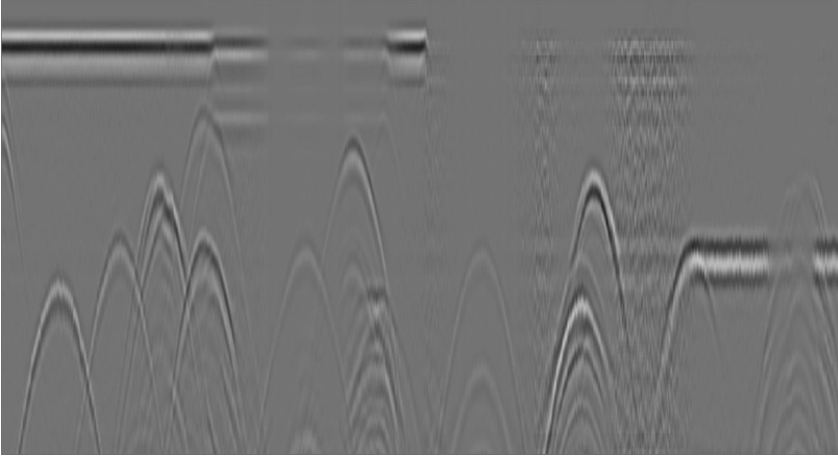


Fig. 2. B-Scan.

### *Challenges Experienced During Utility Mapping*

Despite being widely used, GPR-based analysis presents certain challenges (Cassidy & Jol, 2009). Those are technical and organizational in nature. Key challenges experienced by GPR practitioners are:

- The interface between the air and the subsurface is small but significant. In fact, strong echoes are produced that are known as ground bounce.
- Waves emitted attenuate over time. As a result, shallow objects cause relatively strong echoes while deep objects are scantily visible. Compensating attenuation can be done with time-varying gain filters (Cassidy & Jol, 2009).
- GPR analysis tasks are reported to be labour-intensive. The industry partner reports that they are also repetitive. Human beings generally do not excel at such tasks (Cummings, 2014).
- Becoming a GPR analysis expert requires significant training in geophysics. Consequently, analysts are scarce (Versloot, 2019).

This work contributes to reducing those challenges by designing an intelligent system that amplifies the underground mapping process. Machine learning algorithms are used for this purpose. Therefore, we next position our work with respect to the Industry 4.0 paradigm, introduce the possible benefits of symbiotic or IA relationships between humans and machines and finally review the literature on applying ML to underground mapping.

### *Industry 4.0 and Intelligence Amplification*

The Industry 4.0 paradigm was introduced to secure the competitiveness of the German manufacturing industry. It has now spread into a global development and comprises base technologies which, integrated with business processes, accelerate the fourth industrial revolution (Kagermann, Wahlster, & Helbig, 2013).

### *Base Technologies*

Technology has become increasingly important since World War 2 (Isaacsson, 2014). The nascence of the transistor has created unprecedented technology growth. Computers, once

available only at the most sophisticated research institutes, have moved into peoples' homes and hands. Increased interconnectivity between devices through the internet has democratized information which resulted in the third industrial revolution (Isaacsson, 2014; Lasi, Fettke, Feld, & Hoffmann, 2014). In the fourth revolution, which is currently underway, the role of technology is similarly significant. This time, however, different technologies serve as its primary driver. Specifically, the Internet of Things, cloud services, big data and analytics are the key technologies in this revolution (Kagermann et al., 2013). They are the so-called base technologies (Frank, Dalenogare, & Ayala, 2019). Machine Learning is an instance of analytics.

#### *Business Impact through Front-End Technologies*

According to Frank et al., front-end technologies interface between base technologies and business actors (Frank et al., 2019). They can also be considered to be application areas. The initial Industry 4.0 document considered one front-end technology, Smart Manufacturing (Kagermann et al., 2013). Frank et al. added three others, yielding these Industry 4.0 application areas all driven by base technologies:

- *Smart Manufacturing*: considers internal production operations.
- *Smart Working*: considers operational activities.
- *Smart Products*: considers end products.
- *Smart Supply Chain*: considers improved processes through supply chain integration.

Frank et al. argue that Smart Working comprises 'technologies [supporting] worker's tasks, enabling them to be more productive and flexible to attend [to their] (...) requirements' (Frank et al., 2019). This definition aligns with the operational improvements for GPR analysts' tasks discussed previously. This work can thus be cast as an attempt to produce a Smart Working system.

#### *The Need for Integration*

Traditionally, organizations have developed in a siloed fashion (Britton & Bye, 2004). Organizations distributed responsibilities for IT acquisition and operations to individual departments that would guard these strongly. This impacted organizational technology and application landscapes, which have traditionally been very scattered. This results in for example interoperability problems and vendor lock-ins. The Industry 4.0 principle radically breaks with this view-point. Rather, it prescribes that technologies are integrated both with other technologies and into production processes (Diez-Olivan, Del Ser, Galar, & Sierra, 2019). This integration occurs at various levels:

- *Horizontally*: distinct base technologies are combined to create integrated solutions (e.g. combining the IoT for acquiring data with ML for creating predictive models).
- *Vertically*: individual or integrated base technologies are integrated with business processes through front-end technologies to provide business value.
- *Circularly*: horizontal and vertical technology integration is combined to create a sound and relevant IT-based solution for a business problem, considering their lifecycles as well.

The intelligent system designed in this work integrates circularly. Horizontal integration is provided by means of integrating ML with big data and cloud services, while vertical

integration is achieved by embedding it into a generic analysis process in Section ‘[Embedding the System Into GPR Analysis Process](#)’.

*Intelligence Amplification: A Symbiotic Relationship Between Humans and Technology*

Although base technologies can automate human tasks, one should consider the degree of automation a priori. For example, our industry partner claims that GPR analysts cannot be fully replaced by Industry 4.0 base technologies; they must remain in the loop. In the dawn of the computing era, technologists however considered automation to be binary: problem-solving was either entirely automated or fully left to human beings (Cummings, 2014). This viewpoint shifted in the early 1950s. Scholars, attempting to characterize the field of human-computer interaction, proposed a set of heuristics to distinguish between what ‘men are better at’ and what ‘machines are better at’: the MABA-MABA heuristics (Fitts, 1951).

Those were later expanded into Levels of Automation (LoA), which explain to what extent humans interact with information systems in a decision-making situation (Cummings, 2014; Parasuraman, Sheridan, & Wickens, 2000). They extend the binary view of automation and allow humans and machines to work together. Machines can increase human intelligence by amplifying it (Ashby, 1957). This observation emerges from machine capabilities for problem-solving, which Ashby argues it comes down to a suitable selection (Ashby, 1957). Freely interpreted, the core of his argument is that solving a problem equals picking the best solution out of a set of candidates. Since he claims that intelligence is measured as one’s ‘power of appropriate selection’, and that devices can amplify this power (i.e. assisting in picking a solution), he analogizes that intelligence can be amplified. Ashby’s comparisons would be the basis of further research and the nascence of the research area known as IA. Intelligence Amplification, by means of the front-end technologies interfacing between base technologies and business actors, is intrinsically related to the Industry 4.0 paradigm discussed before. Intelligently supporting GPR analysts by means of ML algorithms, which we cast as an Industry 4.0 instance, can namely also be cast as an IA instance. Following the concerns raised by our industry partner, this work explicitly takes the point of view that GPR analysts should be amplified rather than replaced. This way, the operational aspects of their tasks can be improved for simple objects while human creativity is still required for complex ones. We next review the literature already available for this research goal.

*Machine Learning Approaches for Underground Mapping*

Recently, ML algorithms, today especially deep learning (DL) ones, have been used to eliminate repetitive tasks in various business domains. They allow for ‘[discovering] regularities in data through the use of computer algorithms and [by using them] to take actions’ (Bishop, 2006). Analyzing GPR data is essentially a classification problem: the analyst, given contextual data, classifies an object with respect to its material type. Since GPR analysis is often repetitive and GPR analysts are scarce, applying ML here can be worthwhile. In fact, many studies have validated ML approaches for underground mapping. They can be grouped into four distinct groups (Pasolli, Melgani, & Donelli, 2009a). Primarily, studies report on (1) detection and localization, rendering it trivial today. Less work has focused on (2) material recognition and estimation of (3) dimension and (4) shape.

### *Machine Learning for Recognizing Material Type, Shape and Size*

Given the popularity of Support Vector Machines (SVMs) during the early 2010s, they were primarily used then. Consequently, various SVM approaches were identified. For example, El-Mahallawy and Hashim combine noise reduction and discrete cosine transform (DCT) A-scan signal compression with SVM classification (El-Mahallawy & Hashim, 2013). DCT-based features yielded superior results over time series and statistical ones. Shao et al. also use SVMs but apply other signal processing methods. They first sparsely represent an A-scan by ‘[expressing] a signal as a linear combination of elementary waves’ (Shao, Bouzerdoum, & Phung, 2013). In another study, Pasolli et al. combine SVMs with B-scans (Pasolli et al., 2009a). They also demonstrate that estimating object size is possible as well (Pasolli, Melgani, & Donelli, 2009b). The subfield of DL experienced a breakthrough in 2012 (Jordan & Mitchell, 2015). Since then, scholars have applied DNNs to material recognition. Zhang et al. validate an architecture of three neural networks for recognizing object shape, material and size (Zhang, Huston, & Xia, 2016). Their network also computes object depth and medium conductivity. It however only supports a limited number of material types. More recently, Almaimani successfully applied Convolutional Neural Networks (CNNs) to material recognition (Almaimani, 2018). Contrary to previous approaches, no feature extraction was performed. Rather, a B-scan slice is used as a feature vector. Although her results are promising, she welcomes more research that demonstrates the applicability of CNNs to material recognition.

## RESEARCH METHODOLOGY

Creating an intelligent system is a typical design problem in which an artefact that aims to improve a problem context is designed and developed (Wieringa, 2014). Design science methodologies can be used to attain scientific rigor during such research. They ensure that artefacts are both theoretically sound and practically relevant. Several methodologies exist for design research. Choosing one partially relies on the practicality of the research at hand, since certain methods rely more heavily on theory than others, while those often allow researchers to align their work with practice more easily. Sein et al. argue that methodologies like the Design Science Research Methodology by Peffers et al. ‘fail to recognize that the [artefact] emerges from interaction with the organizational context even when its initial design is guided by the researchers’ intent’ (Peffers, Tuunanen, Rothemberger, & Chatterjee, 2007; Sein, Henfridsson, Puro, Rossi, & Lindgren, 2011). They would produce insufficient agility when working with industry partners. Inspired by Action Research, they conceptualize the Action Design Research (ADR) methodology. It iteratively interweaves artefact development with organizational intervention and evaluation and is especially relevant for business problem-oriented research. The research carried out in this work has been triggered by the business problem discussed in Section ‘Introduction’. Additionally, artefact design and development was performed in strong collaboration with an industry partner. The ADR methodology, therefore, guided this work.

## ML MODELS UNDERLYING THE INTELLIGENT SYSTEM

### *Rationale*

Histograms can be used to count the number of instances across a range of values in a statistical sample. Weia and Hashim created histograms based on A-scan signal backscatters

and a thresholding algorithm, demonstrating that various material types can be discriminated for human analysis (Weia & Hashim, 2012). We apply their feature extraction approach for training ML models. Therefore, one of the classes of CNNs trained is a histogram-based one. El-Mahallawy and Hashim harness the Discrete Cosine Transform (DCT), which is known for signal compression, for training SVM classifiers (El-Mahallawy & Hashim, 2013). CNNs could however perform better for multiple reasons. First, training SVMs requires the configuration of a kernel function a priori. Kernels are generic functions for computing similarity and may not be entirely suitable to the ML problem at hand. This cannot be known in advance. Second, SVMs cannot be used for multiclass predictions (i.e. when the number of material types is  $>2$ ). Third, SVMs do not scale well with larger datasets. CNNs do however learn kernels themselves, are capable of generating multiclass predictions and do scale with larger data volumes. This work therefore replicates the application of the DCT with CNNs. Generally speaking, however, the ML community suggests that minimum feature extraction must be applied when training CNNs (Chollet, 2018). That is, since they can learn filters themselves, data should be input as raw as possible. This work therefore also validates a CNN trained on slices of slightly pre-processed B-scans. In total, therefore, three classes of CNNs are validated in this work: a histogram-based class, a DCT-based class and a B-scan window-based one.

### *CNN Architecture*

The CNNs contain various components combined into an architecture. Specifically, it uses a convolutional block and a densely connected block. Fig. 3 presents the architecture. It contains those components:

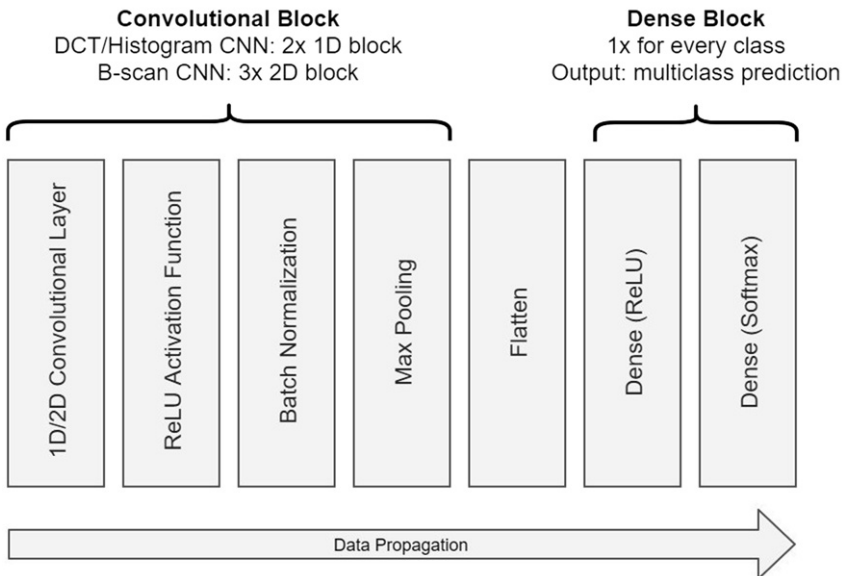


Fig. 3. CNN Components.



- The convolutional block contains a convolutional layer, batch normalization and max pooling. Those are appropriate for learning from image-like data (Chollet, 2018). The DCT and histogram-based CNNs have two since data are already sparsened; the B-scan one has three.
- The densely connected block contains two Dense layers. They convert the patterns identified by the convolutional block into a multiclass prediction. All CNNs have one densely connected block attached to the final convolutional block. To interface, a Flatten layer is added in between.

### *GprMax Simulations*

A training set was generated using gprMax, which implements the finite-difference time-domain (FDTD) method for simulating GPR imagery (Giannopoulos, 2005). In total, 770 B-scans were generated using gprMax Python scripts. A custom wavelet generated by the GPR used by our industry partner was emitted in the simulations to mimic the real world as much as possible. Every simulation represents a B-scan composed of 150 A-scan traces. One A-scan is composed of 1,024 signal backscatter amplitudes. In total, six target classes were simulated: a concrete sewage, high-density polyethylene (HDPE), iron, perfect electricity conductors (PECs) like steel and copper, tree roots and stoneware pipelines. In the simulations, object contents were varied and objects were buried at various depths. The soil was randomly varied over the entire spectrum of available soil types and signal interference was introduced by adding noise.

### *Data Pre-processing*

Before training, the simulations were pre-processed as follows:

- (1) The gprMax output file was first converted to a readable GSSI file.
- (2) Ground bounce was removed with a median-based filter (Versloot, 2019).
- (3) Energy decay (i.e. exponential) gain was used to reduce signal attenuation.
- (4) Feature-wise normalization was applied to reduce amplitude variance without changing the A-scan waveform shape. In DL, this benefits model performance (Chollet, 2018).
- (5) Feature extraction was applied which was dependent on the algorithm.
  - For the histogram-based CNN, histograms computed using the interval  $[-5\sigma, 5\sigma]$  were included. Specifically, since the bin size was  $\sigma/10$ , the feature vector extracted contained 101 features.
  - For the DCT-based CNN, the DCT was computed using SciPy's signal processing package. Inspired by (El-Mahallawy & Hashim, 2013), only the 14 first DCT coefficients out of 1,024 compose the feature vector.
  - For the B-scan-based CNN, a window of 25 traces was sliced left and right of the hyperbola. Since an A-scan is composed of 1,024 amplitudes, feature vector shape was (51 and 1,024).

### *Training, Validation and Testing Data*

DL datasets must be split into training, validation and testing data (Chollet, 2018). With training data, predictions are generated that can be compared to actual targets. Validation data are used to identify the effectiveness of subsequent optimization. Finally, final performance is measured with testing data the model has not seen before. This way, one can assess its predictive and generalization powers. Creating these sets can be done naively by

simply holding out certain proportions for training, validation and testing data (Chollet, 2018). However, with imbalanced datasets, this could produce overly confident model performance. Using K-fold cross-validation, performance is computed as the average over K training attempts. This yields more accurate performance metrics, but is K times more expensive. Typically, based on empirical results, the value of K ranges between 5 and 10. We use  $K = 10$ .

### *Hyperparameter Tuning*

DL architectures must be configured before training them. This can be achieved through hyperparameter tuning (Chollet, 2018). It involves parameter (i.e. neuron) initialization, choosing a loss function and other performance metrics as well as an optimizer, learning rate (LR), batch size and a number of training iterations (epochs). This work tunes model hyperparameters manually based on evidence from the literature. To accommodate ReLU, we used He uniform initialization for neuron initialization, since it performs best (Kumar, 2017). Categorical cross-entropy loss is used for multiclass predictions (Chollet, 2018). We also utilized accuracy which is more intuitive to humans. Adam optimization is used, striking a balance between sound methods and novel approaches (Ruder, 2016). With the LR Range Test, an optimum default LR is found and then decayed linearly (Smith, 2018). Before training the models with maximum computational resources and with the full data set, we used *KERAS\_LR\_FINDER* to perform the LR Range Test. This allows us to find the maximum learning rate with which the model does not overfit. We performed this test for all three algorithms and per algorithm chose this learning rate as the base learning rate. We subsequently apply a linear decay rate. Batch size is set to 70 given hardware constraints.

Finally, the number of epochs is 200.000. However, training is stopped early when the model has not improved for 30 epochs. The best model is saved to disk. This way, the training process stops exactly in time (Chollet, 2018). More details can be retrieved from (Versloot, 2019).

## **ML PERFORMANCE**

### *Data Pre-processing*

Fig. 4 presents the results of data pre-processing. The upper part presents a raw A-scan. Clearly, the air-ground interface is strong and signal attenuates with time. After pre-processing, ground bounce is no longer present, signal strengths are relatively equal and amplitudes are normalized. Likely, resembling a real-world scenario, regular noise is still present in some A-scans.

### *Initial Model Performance*

Table 1 presents initial model performance across 10 training folds. Multiple hypotheses have emerged why model performance is mediocre:

- (1) Primarily, we considered the model to be underfit – that is, every unique object appears only once in the data set. It is hypothesized that expanding the data set results in better performance.

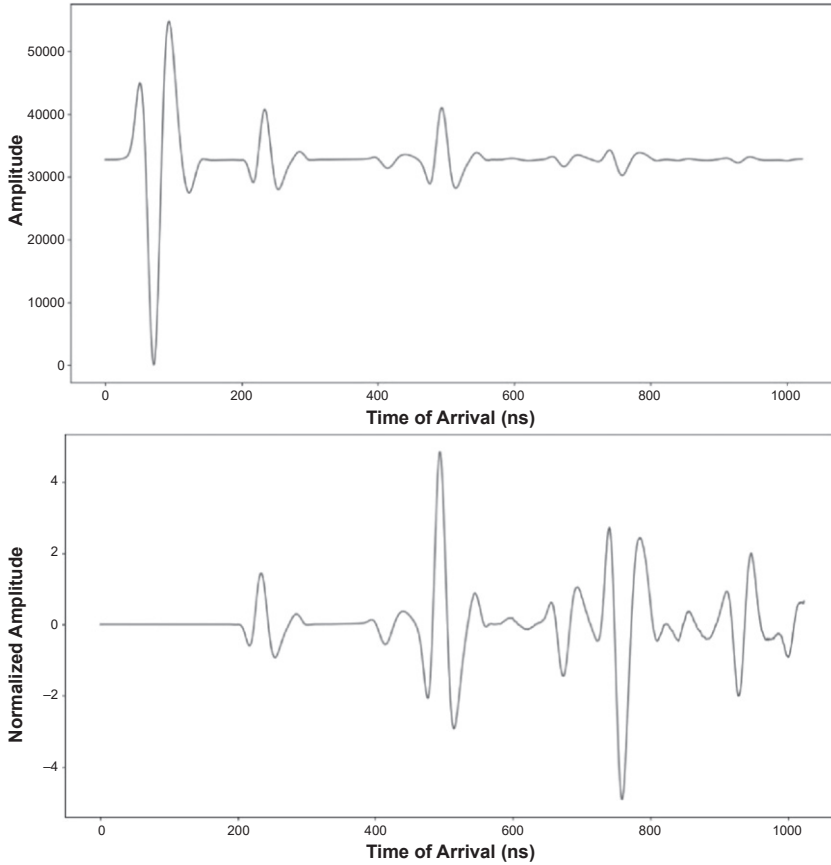


Fig. 4. Unprocessed (Above) and Pre-processed A-Scan (Below).

**Table 1.** Initial Performance of the DL Models.

Model	Average Cross-Entropy Loss	Average Accuracy
Histogram based CNN	1.3162	61.30%
DCT based CNN	1.2176	61.82%
B-scan based CNN	1.3613	67.53%

- (2) Different material contents produce different echoes. Initially, the model did not separate material types with respect to content. We hypothesize that by using material types and contents as targets, performance increases.
- (3) The training process converged quickly. This could be caused by slow LR decay and a consequentially overshoot optimum towards the end of the training process. Increased LR decay may produce better models.
- (4) Pre-processing including feature extraction currently applied may be sub-optimal.
- (5) Generally, different hyperparameter tuning could yield better performing ML models.

### *Variation Performance*

Initially, model performance for all three algorithms was mediocre, with accuracies ranging between 60% and 70%. The three CNN classes were retrained with various variations for testing these hypotheses. The first variation we performed was to expand the dataset to approximately 2,425 simulations, since we observed that one simulation was unique with respect to the object's unique characteristics and by consequence could only be present in one set (i.e. training/validation or test set).

The histogram-based and DCT-based CNNs did not improve any further. The B-scan based CNN, however, did improve, albeit primarily through dataset expansion. Other variations subsequently improved performance incrementally. Those variations included studying the effect of separating the material and its contents when generating target classes, increasing the decay of the learning rate and studying the effects of varying input (i.e. applying different gains). The effects of varying the number of DCT coefficients, the number of histogram bins and the width of B-scan windows were also studied. Similarly, various model variations were studied, with variations in batch size across the three algorithms as well as differences in hyperparameters. Combining those variations into one yielded the most promising results in terms of loss. For these variations, most B-scan CNN accuracies were in the range of 77–82%. Some variations produced outliers to 86% on individual folds. [Table 2](#) presents the performance of the variations to the B-scan CNN.

## ML DRIVEN INTELLIGENT SYSTEM

The Industry 4.0 paradigm combines base technologies like ML with front-end technologies to provide business value. This section discusses such an interface between ML models and GPR analysts. It also integrates it into a generic analysis process using an ArchiMate model.

### *System Design and Instantiation*

The intelligent system is a web application that is capable of analyzing GPR imagery uploaded by the user. When started, a GPR radar file can be uploaded, which is interpreted by the back-end and presented on-screen. Subsequently, the user can fine-tune signal processing applied to the image, altering time-varying gain and ground bounce removal as desired. The browser window immediately adapts the visualization. The user can also click on a hyperbolic signature. When doing so, a window is sliced around the mouse pointer and input into the ML model running in the background. Its prediction is displayed in a popup message. A line drawn on-screen shows the user where he has clicked. [Fig. 5](#) illustrates the tool when used in practice.

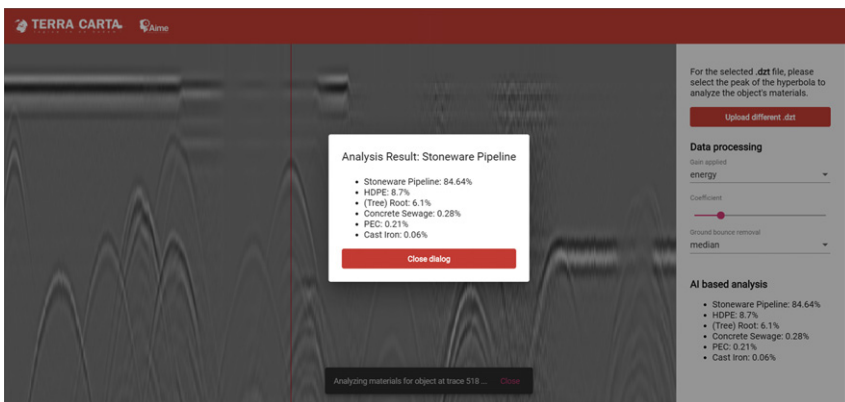
### *Embedding the System Into GPR Analysis Process*

[Fig. 6](#) shows how the intelligent system can be embedded into a generic analysis process. It starts when a customer requests a quote for utility mapping.

This is followed by negotiations and contractual agreement. The project is then added to project planning. At the planned date, a GPR operator records data on site. For this, he configures the GPR and performs measurements. When finished, data are downloaded in the office and sent to the GPR analyst. An analysis request is then added into project planning. When analysis is due, a GPR analyst loads the data into a specific analysis tool. Which tool is used is dependent on the GPR manufacturer; it is often proprietary. First,

**Table 2.** Performance of Variations to Initial B-Scan Based CNN.

Variation	Average Loss	Average Accuracy
Expanded dataset to 2,426 simulations.	0.8834	77.57%
Used separate materials and contents as targets.	1.1695	70.20%
Increased LR decay 175,000 times.	0.8665	77.57%
Reduced shape to (25, 1,024).	2.0215	79.18%
Expanded shape to (75, 1,024).	0.9112	79.18%
Original B-scan input scaled down to 33% of original image size.	0.9706	78.23%
Swish with (101, 1,024) shape.	0.8798	78.81%
Leaky ReLu ( $\alpha = 0.10$ ) with Glorot uniform initializer and (101, 1,024) shape.	0.9805	77.62%
Tanh activation function with Glorot uniform initializer and (101, 1,024) shape.	0.8783	81.16%
Batch size = 5.	2.8527	48.23%
Batch size = 15.	1.1110	66.16%
Batch size = 25.	0.9294	77.87%
Batch size = 35.	0.8599	80.96%
Batch size = 50.	0.8108	81.54%
Batch size = 90.	0.9900	79.39%
Batch size = 115.	0.8369	79.72%
Batch size = 140.	0.7931	80.63%
No gain applied in pre-processing.	1.3299	67.65%
Strong exponential gain applied.	1.2134	67.35%
Linear gain applied instead of energy gain.	0.8517	79.93%
Combined (101, 1,024) shape, Tanh/Glorot, batch size = 50, linear gain, 175.000 $\times$ LR decay.	0.8390	78.44%
Combined (101, 1,024) shape, ReLU/He, batch size = 50, linear gain, 175.000 $\times$ LR decay.	0.7419	79.31%
Combined (101, 1,024) shape, ReLU/He, batch size = 50, original gain, 175.000 $\times$ LR decay.	0.7758	79.76%

**Fig. 5.** Hyperbolic Signature Classified Using the System.

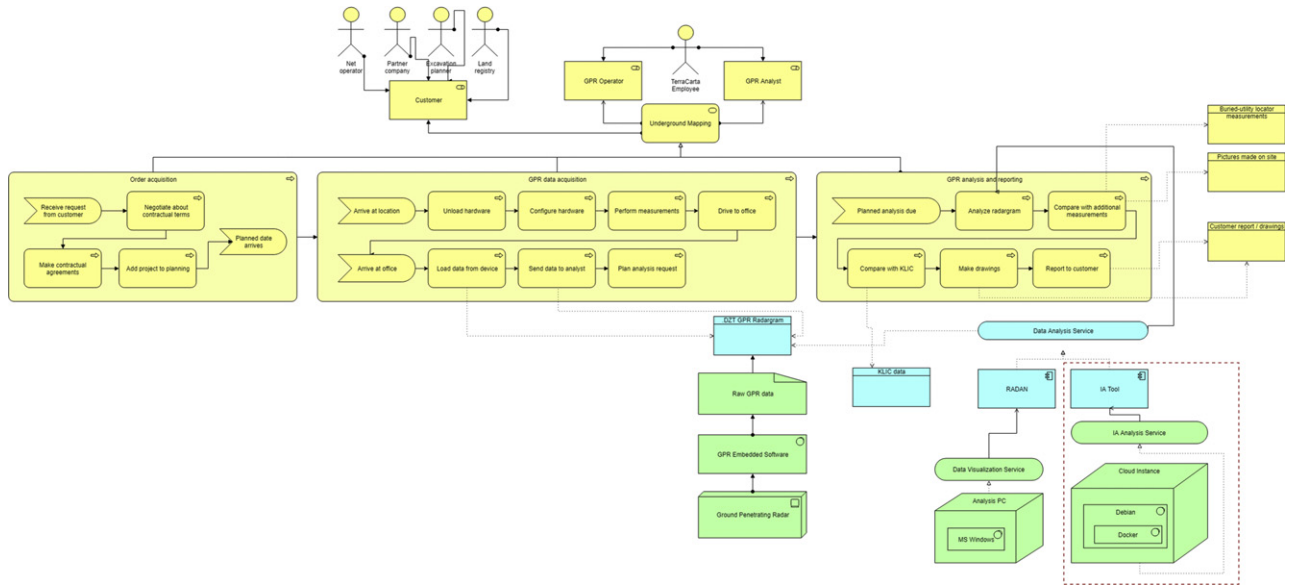


Fig. 6. ArchiMate Model for a Generic GPR Analysis Process.

the GPR image is inspected and preliminary classifications are made based on intuition. Those are compared with additional information such as the KLIC registry or pictures made of trenches dug near utilities. This adds certainty to the analysis. Finally, a drawing is made of the identified utilities. This drawing is then consolidated into a report and sent to the customer. This concludes the analysis and allows the customer to work safely. The intelligent system is part of the ‘data analysis service’ composition and is highlighted with a dashed box. Its business value lies in assisting the user during analysis with respect to material type. This is currently not supported by existing tooling.

#### *Early Validation Comments*

Resulting from utilizing the ADR methodology, the intelligent system was developed in a spirit of co-creation with an industry partner. Consequently, practitioner feedback has been processed into artefact design from the start. Additional feedback was acquired from their upper management and a GPR analyst. It acknowledges the business value provided by the artefact. Specifically, one remark stood out, being that ‘this tool could potentially change entirely the way I do my work’.

## **DISCUSSION**

### *Explaining Performance Differences Between CNNs*

The histogram- and DCT-based CNNs were inspired by previous work harnessing SVMs for material type classification (El-Mahallawy & Hashim, 2013). SVMs can only handle relatively sparse data. GPR data are however anything but sparse, with a  $100 \times 1024$  pixel B-scan slice already yielding approximately 100.000 features. Consequently, scholars were required to reduce input data dimensionality. Histograms and the DCT substantially reduce dimensionality, presumably without data and thus discrimination loss (El-Mahallawy & Hashim, 2013). Precisely this sparsity may in our case result in poor performance when CNNs are applied. In fact, accuracies were only slightly better than random selection while non-sparse feature vectors yielded accuracies averaging 80%. We therefore argue that applying dimensionality reduction to GPR input data for CNNs does indeed deteriorate model performance. We suggest that this behaviour occurs because feature extraction is effectively applied twice. This can be explained through the inner workings of a CNN: the convolution operation applied to the input data effectively allows the network to learn a preconfigured amount of filters itself. We thus suspect that applying feature extraction techniques to reduce input data dimensionality blinds CNN convolution operations to idiosyncrasies in the data, resulting in the relatively poor performance observed. This is in line with the general argument in the DL community to use minimum feature extraction with CNNs (Chollet, 2018).

### *Effectiveness of Variations*

Next, the effectiveness of variations applied to the CNNs is discussed. The discussion primarily focuses on the B-scan CNN variations, since only for this class improvements can be reported. Specifically, the effectiveness of data set expansion, varied activation functions, varied batch size, varied signal gain and combining individual variations is discussed. Based on initial model performance, we hypothesized that our models were underfit given the lack of variety present in our data set. This point of view was confirmed

by expanding the dataset from 770 to 2,426 objects, introducing redundancy and extra random noise. This improved the model by approximately 10–15 percentage points. It is however unclear if it remains underfit. State-of-the-art activation functions like Swish and Leaky ReLU did not lead to substantial performance improvements. For Leaky ReLU, there is a slight chance that this observation emerged from misconfiguring the  $\alpha$  parameter (Versloot, 2019). However, we argue that it is more likely due to the compactness of our CNNs. That is, Swish and Leaky ReLU avoid the death of ReLU powered networks. In those, neurons can die as a result of the vanishing gradients problem, which becomes stronger when networks are deeper (Versloot, 2019). For Swish, improved model performance was observed in very deep networks (Ramachandran, Zoph, & Le, 2017; Versloot, 2019).

The models used in this study were compact with only two or three convolutional blocks and one densely classified block. Therefore, we argue that ReLU suffices for compact CNNs for GPR imagery. The activation function Tanh resulted in model improvements (Versloot, 2019). It is unclear why this behaviour occurs. However, we believe this might be related to the Batch Normalization and/or L2 Regularization techniques applied to the CNNs. Since Tanh activates on the  $[-1, +1]$  range, it might be a more native fit to regularized networks compared to, for example ReLU. Unsurprisingly, increasing batch size improved model performance (Versloot, 2019). This is in line with the mathematical constructs revolving DL model optimization (Chollet, 2018). Neither a surprise are the increasing memory requirements. The DL practitioner should thus always strike an optimal balance between batch size and hardware capabilities before training a DL model. Besides energy gain, we also trained variations without any gain, with strong gain introducing inverted attenuation and linear instead of exponential gain. Fig. 7 demonstrates the effect of those variations on an arbitrary B-scan. The results demonstrate that regular gain performs best, followed by linear gain. Apparently, the main object reflection is considered to be most discriminative for the material type. Although they are not the main discriminator, sub reflections do benefit the discriminative power of the model. This argument is supported by the observation that both stronger and no gain introduce worse performance. Finally, combinations of individual variations were retrained to assess model performance. All three combinations from Table 2 resulted in better model performance, sometimes substantially with respect to observed loss. Why this occurs remains unknown (Versloot, 2019).

### Study Limitations

The study reported in this chapter is limited in multiple ways. The first is how the simulations were generated. We used GprMax 2D for this purpose, which simulates wave

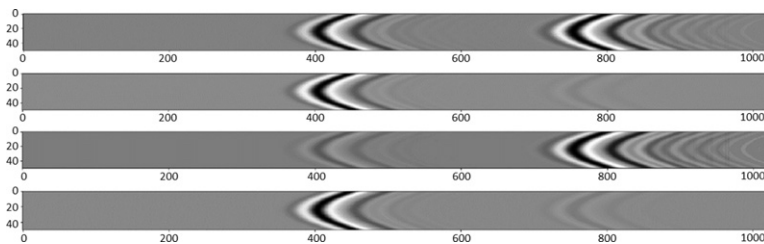


Fig. 7. Regular, No, Strong and Linear Gain Applied; Rotated 90° Counterclockwise.



emission and reception in 2D. Real GPRs however emit and receive them in 3D. This may result in deviations between similar hyperbolae in 2D and 3D imagery. Since no data are mixed, this does not impact the discriminative power of our model (Versloot, 2019). However, the intelligent system should be used with caution. The second limitation is the noise traditionally present in GPR imagery. Although random noise was added in the simulations, it is unknown whether this fully captures the noise levels present in real imagery. Third, during the training process an issue with applying gain was discovered as a result of pre-processing GprMax output data (Versloot, 2019). It is assumed that this issue did not impact ML performance, but it must be corrected should the intelligent system be used with real data. Fourth, the CNNs trained in this work were tuned manually with respect to their hyperparameters. Although this is acceptable practice in the ML community, tooling has emerged which converts finding suitable architectures and hyperparameters into a large search problem (Chollet, 2018; Versloot, 2019). Although the results show that our tuning efforts already lead to plateauing model performance, it may be the case that even better hyperparameters can be identified. Fifth, as illustrated in Section ‘Explaining Performance Differences Between CNNs’, it remains unknown whether the model is still underfit. It may be the case that model performance can be increased by, for example adding similar objects, objects with peculiarities and objects disturbed by the presence of other objects. Finally, validation feedback was only acquired from within one organization, being the industry partner of this work. To derive additional insights like adoption criteria, it must be validated more broadly.

## CONCLUSION

In this work, an intelligent system for predicting utility material types from GPR imagery was designed and developed. It is driven by three classes of CNNs specifically trained for this purpose. Two of them, the histogram-based and DCT-based ones, were inspired by previous research on this problem. The third was inspired by the DL community’s wisdom that data should be as raw as possible when using CNNs. GprMax was used for simulating the utilities.

Initially, model performance for all three algorithms was mediocre, presumably due to underfitting resulting from a lack of variety in the dataset. By training various variations to the initial algorithm, including expanding the dataset to 2,426 utilities, performance of the B-scan model was increased to approximately 80%. The histogram-based and DCT-based models did not improve. The system was embedded into a generic GPR analysis process. Early validation comments were retrieved from our industry partner, confirming its business value. Our work therefore contributes to science and practice in multiple ways:

- First, CNNs can successfully be applied to GPR-based object material recognition. Although previous studies achieved this as well, the models trained in this work predict a more varied set of targets which partially overlap in terms of electromagnetic properties.
- Second, the literature on automating utility mapping by means of ML was introduced to IA.
- Third, the feasibility of this approach was demonstrated by designing and developing an intelligent system that interfaces between the ML models and GPR analysts.
- Fourth, this work has allowed our industry partner to validate novel ideas related to interpreting GPR imagery, possibly optimizing their analysis process by consequence.

Multiple suggestions for future work can be made:

- Primarily, it is suggested that the data set is further expanded to assess whether the CNN is still underfit. Possibly, model performance can be improved even further.
- Second, we suggest assessing automated hyperparameter tuning suitability for GPR analysis.
- Third, the compatibility of GprMax 2D simulations and real-world GPR data could be explored.
- Fourth, the generalizability of training CNNs on data simulated for the GPR used by our industry partner to other GPRs could be investigated.
- Fifth, the intelligent system designed and developed in this work could be validated more thoroughly, acquiring insights in design and adopting criteria for tooling supporting GPR analyses.

## ACKNOWLEDGEMENTS

We acknowledge the efforts of our industry partner TerraCarta B.V., which has substantially contributed to the research results with field expertise and simulated GPR imagery.

## REFERENCES

- Almaimani, M. (2018). *Classifying GPR images using convolutional neural networks*. Master's thesis, The University of Tennessee at Chattanooga, Chattanooga, TN.
- Ashby, W. (1957). *An introduction to cybernetics*. London: Chapman & Hall Ltd.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer Science+Business Media, LLC.
- Britton, C., & Bye, P. (2004). *IT architectures and middleware*. Boston, MA: Addison-Wesley Educational Publishers Inc.
- Cassidy, N. J., & Jol, H. M. (2009). Ground penetrating radar data processing, modelling and analysis. In H. M. Jol (Ed.), *Ground penetrating radar: Theory and applications* (pp. 141–176). Amsterdam: Elsevier Science.
- Chollet, F. (2018). *Deep learning with Python*. New York, NY: Manning Publications Co..
- Cummings, M. (2014). Man versus machine or man + machine? *IEEE Intelligent Systems*, 5, 62–69.
- Diez-Olivan, A., Del Ser, J., Galar, D., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion*, 50, 92–111.
- El-Mahallawy, M. S., & Hashim, M. (2013). Material classification of underground utilities from GPR images using DCT-based SVM approach. *IEEE Geoscience and Remote Sensing Letters*, 10(6), 1542–1546.
- van der Eng, P. (1987). *De Marshall-hulp, Een Perspectief voor Nederland 1947–1953*. Houten: De Haan.
- Fitts, P. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Technical Report. Division of the National Research Council, Oxford.
- Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing. *International Journal of Production Economics*, 15–26.
- Giannopoulos, A. (2005). Modelling ground penetrating radar by GprMax. *Journal of Construction and Building Materials*, 19(10), 755–762.
- Isaacsson, W. (2014). *The innovators: How a group of hackers, geniuses and geeks created the digital revolution*. New York, NY: Simon & Schuster.
- Jordan, M., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 6245, 255–260.
- Kadaster. (2008). Klic-wab: Het registreren van uw belang via Internet.
- Kagermann, H., Wahlster, W., & Helbig, J. (2013). *Recommendations for implementing the strategic initiative INDUSTRIE 4.0. acatech–National Academy of Science and Engineering*. Federal Ministry of Education and Research, Final Report.
- Kumar, S. K. (2017). *On weight initialization in deep neural networks*. Technical Report.
- Lasi, H., Fettke, P., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering*, 4, 239–242.

- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 30(3), 286–297.
- Pasolli, E., Melgani, F., & Donelli, M. (2009a). Automatic analysis of GPR images: A pattern-recognition approach. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 2206–2217.
- Pasolli, E., Melgani, F., & Donelli, M. (2009b). Gaussian process approach to buried object size estimation in GPR images. *IEEE Geoscience and Remote Sensing Letters*, 7(1), 141–145.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 8(3), 45–77.
- Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Swish: A self-gated activation function. Retrieved from <https://arxiv-org.ezproxy2.utwente.nl/abs/1710.05941>
- Rijksoverheid. (2015). *Graafschade aan ondergrondse leidingen en kabels*.
- Rijksoverheid. (2019). *Wet informatie-uitwisseling bovengrondse en ondergrondse netten en netwerken*.
- Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. arXiv preprint arXiv:1609.04747.
- Scheers, B. (2001). *Ultra-wideband ground penetrating radar, with application to the detection of anti personnel landmines*. PhD thesis, Université Catholique de Louvain, Louvain-la-Neuve.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly*, 35, 37–56.
- Shao, W., Bouzerdoum, A., & Phung, S. L. (2013). Sparse Representation of GPR traces with application to signal classification. *IEEE Transactions on Geoscience and Remote Sensing*, 51, 3922–3930.
- Smith, L. N. (2018). *A disciplined approach to neural network hyper-parameters: Part 1 – Learning rate, batch size, momentum, and weight decay*. Technical Report. US Naval Research Laboratory, Washington, DC.
- Versloot, C. W. A. (2019). *Amplifying the analyst: Machine learning approaches for buried utility characterization*. Master's thesis, University of Twente, Enschede.
- Weia, J. S., & Hashimb, M. (2012). Ground penetrating radar backscatter for underground utility assets material recognition. In *33rd Asian conference on remote sensing (ACRS 2012)* (Vol. 3, pp. 2179–2187). Pattaya, Thailand.
- Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Berlin: Springer-Verlag.
- Zhang, Y., Huston, D., & Xia, T. (2016, April). Underground object characterization based on neural networks for ground penetrating radar data. In *Nondestructive characterization and monitoring of advanced materials, aerospace, and civil infrastructure 2016* (Vol. 9804, p. 980403). International Society for Optics and Photonics.