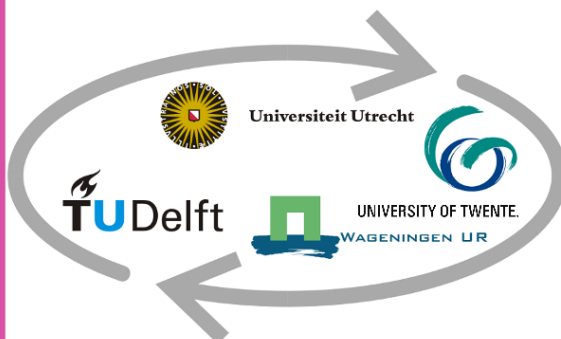


**Integrating Ground Penetrating Radar data in a GIS to detect
unmarked graves via a Convolutional Neural Network framework for
3D point cloud interpretation**

**A case study at the Jewish Cemetery in The Hague,
The Netherlands**

MSc thesis by Christian Riesner

March 04, 2022



Integrating Ground Penetrating Radar data in a GIS to detect unmarked graves via a Convolutional Neural Network framework for 3D point cloud interpretation

-

A case study at the Jewish Cemetery in The Hague, The Netherlands

Author: Christian Riesner
E-mail: c.riesner@students.uu.nl
Student number: 6630529

Master thesis submitted to the joint venture of
Utrecht University, TU Delft, Wageningen University & Research, Twente University,
in partial fulfilment of the requirements for the degree:

Master of Science in Geographic Information Management and Application

Supervisors:

Dr Ir. Dominique Ngan-Tillard (TU Delft)
Ir. Edward Verbree (TU-Delft)

Responsible Professor:
Prof. Dr Ir. P.J.M. van Oosterom (TU Delft)

March 04, 2022

Preface

This thesis, *'Integrating Ground Penetrating Radar data in a GIS to detect unmarked graves via a Convolutional Neural Network framework for 3D point cloud interpretation'*, has been written as the final part of the master's programme Geographical Information Management and Applications, in the period between September 2021 and March 2022. This course has been like the journey to Ithaka. And without the help of some people, I would not have made it.

For this graduation research, I want to thank my two supervisors, Dominique and Edward, for the regular, constructive meetings and feedback that helped bring more insight into two complete unknown domains. Especially Dominique, who introduced me to the field of Ground Penetrating Radar and collected the data together for this project.

Thank you also to David Jonker from *Dotka Data* for providing the aerial images utilized in this work.

For the entire journey, I want to give my gratitude to Kristina. She was always there to keep me on the road to Ithaka. Thank you.

Abstract

The Jewish cemetery in The Hague is a listed national monument of The Netherlands and is a document of the Jewish heritage in the area. The site is occupied by more than 10.000 people buried there. Only 2.860 can be located on the ground as they are marked with a gravestone on their surface top. The spatial positions of the remaining unmarked burial locations are of high interest to cemetery management and the Jewish community. As a physical inventory is not possible, non-invasive methods are preferred. Ground Penetrating Radar (GPR) is a technology that has been successfully applied to detect sub-surface features. A survey with multiple sensors, including GPR, was undertaken on two sample areas of the cemetery. The conventional method to analyse the collected data sets is the interpretation of vertical and horizontal visualisations. This method required expert knowledge.

An alternative approach is to integrate the GPR data into a Geographic Information System (GIS) in the form of a 3D-point cloud. The data assessment was supported by applying the deep learning algorithm, *PointCNN*, that performs an automated classification of the points to segment the burial locations as features of interest. In this research, various classification models were created, applied, and evaluated to their usability. Furthermore, additional data sets were combined with the classified point clouds to incorporate more information to identify the unmarked burials.

The applied method highlights the advantages of a GIS for this spatial analytical task and reveals some vulnerable points in the application of deep learning. Therefore, the workflow in this research project will have to undergo adaptations as recommended at the end of the report.

Contents

Preface	I
Abstract.....	II
List of Figures	V
List of Tables	VIII
List Abbreviations	IX
1 Problem and Context.....	1
1.1 Introduction	1
1.2 Problem description	1
2 Research objectives, questions, and scope	4
2.1 Research objectives	4
2.2 Aim and research questions.....	4
2.3 Scope of research	4
2.4 Hypothesis	5
3 Theoretical background	6
3.1 Deep learning.....	6
3.1.1 Convolutional Neural Networks (CNN)	7
3.2 Point cloud features.....	8
3.2.1 Deep learning applied to 3D data features	8
3.3 Automatic data analysis of GPR data	10
4 Methodology.....	12
4.1 Case study area.....	12
4.2 Data collection, processing, and integration.....	13
4.2.1 Ground Penetrating Radar (GPR) data	14
4.2.2 Global Position System (GPS) data	15
4.2.3 Topographical Surface data	15
4.2.4 Historical Data.....	16
4.3 Research Software	16
4.4 Project Workflow	17
4.5 Semiautomated point cloud classification by CNN model.....	17
4.5.1 Assessment and labelling of GPR point cloud.....	18
4.5.2 Prepare point cloud training data	20
4.6 Train model for point cloud classification	21
4.7 Evaluation and testing of the classification model	24
5 Results.....	26
5.1 Created classification models and validation parameters.....	26
5.2 Visual evaluation of the classification models.....	29
5.3 Quantitative evaluation of the classification models	34
5.4 Topographical data.....	36
5.5 Historical aerial data	36

6	Discussion.....	37
6.1	Implementation of the methodology.....	37
6.2	Usability of methodology.....	37
6.3	Research Limitations	38
7	Conclusion	40
7.1	Research questions	40
7.2	Conclusion	43
7.3	Recommendations for future research	44
8	References	45
	<i>Annex A</i>	49
	<i>Annex B</i>	50
	<i>Annex C</i>	51
	<i>Annex D</i>	53
	<i>Annex E</i>	54
	<i>Annex F</i>	55
	<i>Annex G</i>	56
	<i>Annex H</i>	57
	<i>Annex J</i>	58
	<i>Annex K</i>	60
	<i>Annex L</i>	61
	<i>Annex M</i>	62

List of Figures

Figure 1: Example of a radargram profile across an intact coffin recorded at the White Rock Cemetery in Pineville, Louisiana (Goodman and Piro, 2013)	2
Figure 2: A volumetric voxel model displaying the GPR data of an archaeological site (Mohamed <i>et al.</i> , 2020).....	2
Figure 3: In this example, the isosurfaces shows discrete anomalies indicative of marked/unmarked burial (Goodman and Piro, 2013).	2
Figure 4: Venn diagram representing the hierarchical learning concepts of AI technology (Author).	6
Figure 5: Schematic representation of the layered structure in a neural network design (Sorokina, 2017)	7
Figure 6: Conceptual structure of a convolutional neural network consisting of the input layer, convolutional layer, pooling layer, and fully connected layer (Biswal, 2020)	8
Figure 7: The upper section demonstrates the convolution on structured (gridded) data where each iteration reduces the resolution of the patch and increases the number of channels (visualised by the dot thickness). The below section describes the X-Convolution (transformation) function to 'gain' or 'project' the neighbours' properties to less representing points with supplemented information load (Li <i>et al.</i> , 2018).....	9
Figure 8: LEFT - Multichannel GPR (Channel#N) with the representation of the orthogonal scan planes B, C, and D. MIDDLE and RIGHT - (a) 3D data block and (b) combination of 1 B-scan and 2 C scan images (Kim <i>et al.</i> , 2021).....	11
Figure 9: Location of the research area, the Jewish Cemetery in The Hague (Author).	12
Figure 10: Overview layout displaying the location of the existing tombstones and higher vegetation (Enthoven, Drs. Francine Puttmann <i>et al.</i> , 1992). Indication of research sample areas (blue and red) (Author).....	13
Figure 11: General methodology workflow from data collection to model evaluation (Author)	14
Figure 12: Overview GPR survey set up for sample area AB1 (not to scale) with dimension line width (detail image) (Author).....	14
Figure 13: 2D GPR radargrams (B-scans) of survey Line #0 sample area AB1 of unprocessed raw data (left side) and after the application of filter and topographic correction (right side) (EKKO_Project ject line view, Author).....	15
Figure 14: Detailed workflow diagram starting from the data collection until the interpretation and evaluation in a GIS. (Author).	17
Figure 15: TOP - 2D-Radargram (B-scan) of GPR data line #32. MIDDLE: - Visualisation of vertical profile slice (4cm thickness) of the point cloud in <i>ArcGIS Pro</i> coloured based on the assigned intensity values. BOTTOM: - Manually selected (blue highlighted) points classified as class #1 - <i>burial locations</i> (Table 3). The displayed cross section shows burial with no gravestone present (Author).	19
Figure 16: Orthomosaic of area AB1 with overlayed GPR grid and highlighted line #32 (Author).....	19
Figure 17: Isometric view on sample area AB1 with the partition of training data (red) and validation data (blue) (Author).	20

Figure 18: Training settings overview and epoch parameter (Author).	22
Figure 19: Training- and Validation-Loss curve obtained by a training session (Author).	23
Figure 20: Visual evaluation of one model output (orange points) compared to the reference point cloud (black points). LEFT – Location of assessed part <i>patch 1</i> (red frame) in sample area <i>AB2</i> . CENTER – Orthogonal view on the reference points (black) and the predicted (red) points. RIGHT – Oblique view on the reference points (black) without and with the predicted points (red) (Author).	25
Figure 21: LEFT – Schematic structure of confusion matrix (Draelos, 2019). RIGHT – Confusion matrix of model A displaying the TPs and FPs between classes #0, #1, and #2 (see 5.3) (Author).	25
Figure 22: Seven horizontal time slices of area <i>AB2</i> demonstrating the interpretation option in the GPR processing software <i>EKKO_Project</i> . The slice width and depth are indicated in the top part of the strips. Areas between the survey lines are the results of an interpolation process performed in <i>EKKO_Project</i> . The colourisation is based on the signal intensity values (Author).	29
Figure 23: Visual evaluation of the sample area <i>AB2</i> with the reference data (black) representing class #1 and the corresponding model predictions (coloured) (Image 1 of 2) (Author).	30
Figure 24: Visual evaluation of the sample area <i>AB2</i> with the reference data (black) representing class #1 and the corresponding model predictions (coloured) (Image 2 of 2) (Author).	31
Figure 25: Orthogonal view on the visual evaluation patch <i>P1</i> in sample area <i>AB2</i> displaying the reference data set (black) and the model prediction (individual colour) of points class #1 representing the underneath located burial (coffin) (Author).	32
Figure 26: Oblique view on the visual evaluation patch <i>P1</i> in sample area <i>AB2</i> displaying the reference data set (black) and the model prediction (individual colour) of points class #1 representing the underneath located burial (coffin) (Author).	33
Figure 27: A cut-out from aerial images displaying the research area, including the two sample areas (blue frames). TOP LEFT – 18.03.1945 (DotKa, 2022); TOP RIGHT – 01.03.1945 (DotKa, 2022); and BOTTOM LEFT – 28.10.2021 (Author).	36
Figure 28: Vertical Line (B-scan) view in <i>EKKO_Project</i> software tool on scan line #19 of area <i>AB2</i> . The unmarked burial locations are the hyperbolas reaching their angular point in a depth range 0.4 – 0.8m below the surface. At the distance mark of 16m, the signal disturbances are notably caused by the only burial marked with a gravestone (Author).	37
Figure 29: Result of manual vertical adjustment of single top traces, displaying the first layer of points representing the surface top with multiple misalignments visible between the scan lines (Author).	39
Figure 30: Photocopy of a historic map dating 1832 displaying the dune area between The Hague and the village of Scheveningen (Enthoven, Drs. Francine Puttmann <i>et al.</i> , 1992). The red frame indicates the site of the Jewish Cemetery.	49
Figure 31: Position and dimension of the two sample areas <i>AB1</i> and <i>AB2</i> . Note that the detail image of <i>AB2</i> is rotated 90° counter-clockwise. The red boundary box representing the patch <i>P1</i> for the visual evaluation. (Author).	50
Figure 32: Scan lines (B-scans) Line 0 – Line 11 of sample area <i>AB2</i> (Author).	51
Figure 33: Scan lines (B-scans) Line 12 – Line 23 of sample area <i>AB2</i> (Author).	52

Figure 34: Coversheet (left) and example of detail layout plan (plattegrond) of the cemetery with individual gravestone numbers for each plan number (Author).....	54
Figure 35: Aerial imagery of the R.A.F. dated from 07/04/1945 covering the area of interest with a blow-up of the cemetery (haagsgemeentearchief).....	55
Figure 36: Block-shaped point cloud area <i>AB1</i> with attribute based colouring depending on the assigned intensity value. TOP to BOTTOM – with orthomosaic superimposed; point cloud displaying all three classes; point cloud displaying the classes #1 and #2; point cloud displaying class #1 (Author).....	56
Figure 37: Three profile views showing the variety of the underground properties of the sample area <i>AB1</i> . In the top window the disturbances caused by the gravestones (light green square shapes at the surface) are prominent. The centre image displays the difference between burial locations marked and unmarked by a gravestone. The bottom image is a slice of undisturbed soil property where intensity values mark natural sediment changes (Author).57	
Figure 38: Histograms displaying the Block and Point-distribution of the different training- (left) and validation- (right) data sets (Author).....	59
Figure 39: Loss graphs of the nine model training sessions (A – J) (Author).	60
Figure 40: Confusion matrices by model evaluation on area <i>AB2</i> (Model B – J) (Author).	61
Figure 41: Classified point cloud of area <i>AB2</i> superimposed with historical aerial imagery (TOP) (DotKa, 2022) and layout plans (BOTTOM) (Enthoven, drs. Francine Puttmann <i>et al.</i> , 1992).....	62

List of Tables

Table 1: \mathcal{X} -Convolution operation parameter and description (Li et al., 2018).....	10
Table 2: Historical aerial imagery properties (Dotka).	16
Table 3: Classification assignment and selection criteria for point cloud labelling. (Author)	18
Table 4: Parameter settings of the training data preparation (Author).....	21
Table 5: List displaying the training performance parameter for each epoch (8) and classes (2) (Author).....	24
Table 6: Training settings and resulting validation parameters for nine models (A-J) obtained from sample area <i>AB1</i> (Author).	28
Table 7: Evaluation parameters for each class prediction for the nine models A – J (Author).	34
Table 8: Confusion matrices based on the model evaluation on data set <i>AB2</i> (Author).....	35
Table 9: Characteristics of the resulting point cloud from photogrammetric process areas <i>AB1</i> and <i>AB2</i> (Author).....	36
Table 10: Coordinates of reference points with error residuals (reference system RD new [Hz] and NAP [V] (Author).....	53
Table 11: Extraction (20 entries) of the database of the cemetery register with names (Naam), date of death (Overlijdensdatum), Gravestonenummer (Zerknummer), parents, church register number (formulier) (Francine Puttmann).....	54

List Abbreviations

AI	Artificial Intelligence
ASPRS	American Society for Photogrammetry and Remote Sensing
CC	Cloud Compare
CNN	Convolutional Neural Network
CSV	Comma-Separated Values
CSVM	Convolutional Support Vector Machine
DIM	Dense Image Matching
DL	Deep Learning
DSM	Digital Surface Model
FME	Feature Manipulation Engine
GIMA	Geographic Information and Management Application
GIS	Geographic Information System
GPR	Ground Penetrating Radar
GPS	Global Positioning System
GPU	Graphic Processing Unit
HDF	Hierarchical Data Format
HMM	Hidden Markov Model
LIDAR	Light Detection and Ranging
ML	Machine Learning
NAP	Normaal Amsterdams Peil
RDBMS	Relational Database Management System
RGB	Red Green Blue
RMS	Root Mean Square Error
SEC	Spreading and Exponential Compensation
SVM	Support Vector Machine
TU	University of Technology
UAV	Unmanned Aerial Vehicle

1 Problem and Context

1.1 Introduction

Land as property has always been a valuable entity. However, the fact that the terrestrial surface is limited, and the vast increasing world population is pilgriming towards the urban areas does require a reformation on the actual land management policies. Where horizontal growth is no longer an option, the third dimension becomes key to optimising land resources for economic and social development in metropolitan areas (Ng and Choy, 2018). Within this situation, cemeteries regularly come under pressure. Because surrounding development projects usually enclosed the graveyards, preventing any extension and creating an immense increase of property value (Goodman and Piro, 2013). The growth of the population is causing additional stress with a higher demand for the use of the burial grounds. The organisation of those sites needs to adapt to that circumstance. If the funeral rites and geological properties allow, creating underground catacombs can be a potential solution. Such an approach has been already realised in the city of Jerusalem, reviving the practice of underground tombs (Nos.nl, 2018).

In addition, it can also be a knowledge gap that is hindering the management of a cemetery from using the available space to its total capacity. Older graveyards are regularly lacking information on their spatial occupation. Historical burial sites are often missing headstones or site markers. One reason can be vandalism or the simple diminishing over time (Goodman and Piro, 2013). Another cause can be the missing financial background of the bereaved to establish a permanent site marker. Efforts have been made to close this gap by detecting and documenting those unmarked burial locations without infringing any religious policies. As a result, remote sensing technology has been the primary choice as they prevent the labour-intensive and ethically delicate search by excavation. Multiple researches showed that Ground Penetrating Radar (GPR) is the most effective archaeological application in finding and mapping unmarked graves via remote sensing (Doolittle and Bellantoni, 2010). Other studies relativise this statement by concluding that the GPR application does not provide a reasonable success rate of grave detection due to the dependency of soil parameters (King, Bevan and Hurry, 1993). However, the exploitation of the underground data by GIS capabilities is at its early stages.

GPR is a geophysical technology to explore the subsurface in shallow depths in a high resolution that has been under constant development since the 1970s (Unumandakh, Amarsaikhan and Sato, 2007). It can detect underground anomalies or disturbance signatures in a minimum of time while avoiding any disturbance of the assessed soil. Followed by data post-processing techniques, it is possible to interpret the results to discover potential burials. (Bevan, 1989). Until today, the GPR data made tremendous development in their quantity and quality. Starting from vertical single 2D radargram slices presenting the results of one applied wavelength up to the 3D volumes visualisation derived from stepped frequency measurements. Those developments improve the usability and reliability of the technology and open the door to new options for data usage. (Bevan, 1989)

1.2 Problem description

The detection of unmarked burial sites with a ground penetrating radar is not as simple as it sounds to someone without a background in this topic. However, variables like the dielectric properties of the buried objects and their surrounding soils determine the chances of successfully exploring the underground structure (Doolittle and Bellantoni, 2010). In addition, developed data post-processing steps (e.g., parameter approximation, filter application, topographic correction) simplify the interpretation to a certain degree.

Software applications, provided mainly by the sensor manufacturers, enable the examination of the data in 2D vertical radargram slices (Figure 1) or voxel volume visualisation derived by interpolation between the radargrams. The latter requires narrowly spaced profiles to display horizontal amplitude slices (Figure 2) (Mohamed *et al.*, 2020). Grasmueck and Green highlight the advantages of 3D-GPR compared to the individual 2D approach due to the increased resolution and details of the underground features (Grasmueck and Green, 1996).

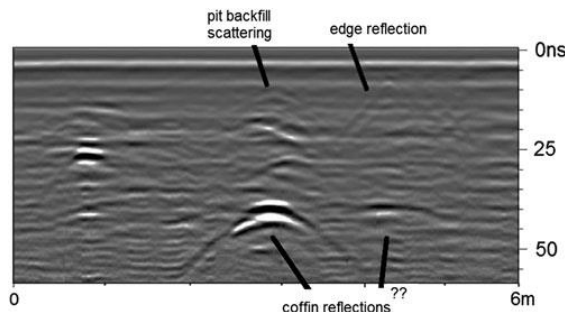


Figure 1: Example of a radargram profile across an intact coffin recorded at the White Rock Cemetery in Pineville, Louisiana (Goodman and Piro, 2013)

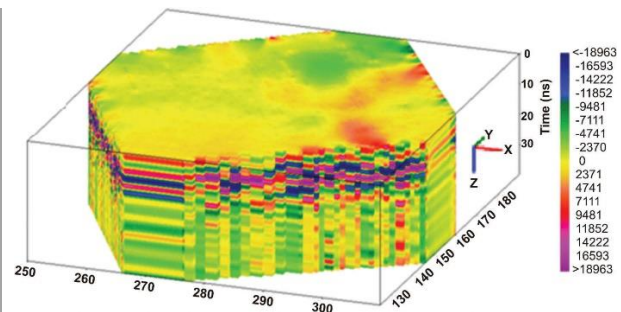


Figure 2: A volumetric voxel model displaying the GPR data of an archaeological site (Mohamed *et al.*, 2020).

To this stage, few studies have been conducted to assess GPR data in a 3D point cloud data representation. The successful combination of 2D or 3D GPR data and 3D terrestrial data has been done in several projects. They concentrate primarily on the (sub-)surface documentation of service utilities in urban environments (Tabarro *et al.*, 2017; Wolf *et al.*, 2018; Merkle, Frey and Reiterer, 2021). The partition of the 3D point cloud in horizontal cross-sections is necessary as the visualisation of the content as a whole is not practical for any analysis (Merkle, Frey and Reiterer, 2021). Several efforts have been recently made to overcome this step backwards from a dimensional perspective. One is the extraction of GPR anomalies in a Geographic Information System (GIS) by transforming them into a vector format feature. However, even though the improvement of interpreting the GPR data is significant, it needs to be mentioned that the mapping process is complex and labour-intensive (Poscetti, Zotti and Neubauer, 2015).

Goodman and Piro present a research project to extract required volumetric areas based on their reflection signal value. With a set threshold value, it is possible to render isosurfaces representing discrete anomalies in three dimensions (Figure 3) (Goodman and Piro, 2013).

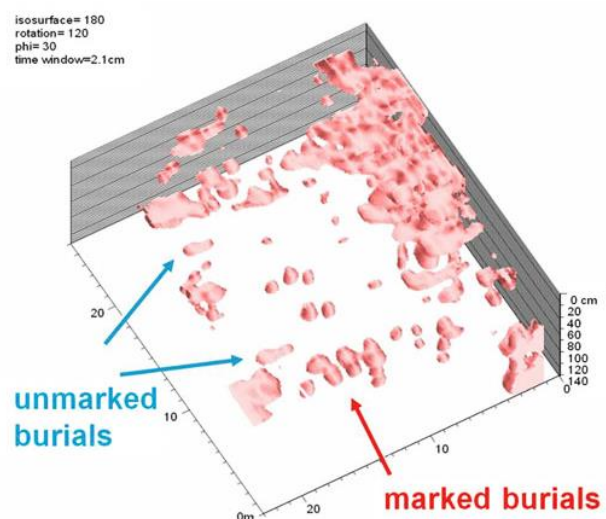


Figure 3: In this example, the isosurfaces shows discrete anomalies indicative of marked/unmarked burial (Goodman and Piro, 2013).

The two above mentioned methodologies, using GIS capabilities for visualisation and the classification via isosurfaces, demonstrate the potential of 3D GPR data. Tischler et al. are predicting a new era of GPR data analysis by integrating them with other georeferenced data and using the functionalities provided by a GIS (Tischler, Collins and Grunwald, 2002). Moreover, combining additional data (e.g. topographic data) will increase the analysis's efficiency and build integrity to its results (Merkle, Frey and Reiterer, 2021).

In recent years machine learning, artificial intelligence, and deep learning has been prominent keywords in data sciences. Deep learning is also used by the geospatial industry, especially for raster data examination. Various research projects have been conducted to exam numerous types of applications. Most of them use pattern recognition to select and interpret features of interest (e.g., traffic sign detection, road condition monitoring, floral species determination). The application on spatial three-dimensional data sets has not been investigated in-depth yet, but the first results were promising. For example, the segmentation of 3D point clouds based on pre-trained models. Those training scenarios are derived by terrestrial or aerial LIDAR technology (Remondino and Golkar, 2019). The results have potential for future applications but underline at the same time their dependency on the complexity of the used training data sets (Malinverni *et al.*, 2019).

This research will apply these methodologies to detect sub-surface anomalies of GPR data collected on a graveyard located in The Hague /The Netherlands. The data acquisition will correspondingly cover the terrain topography and the subsurface of the cemetery using aerial photography and GPR. Both data sets will undergo post-processing (filter, corrections, transformation, etc.) in separate software packages before the integration into the GIS will be carried out. The primary objective is the development of a deep learning algorithm to classify the GPR data in a 3D point cloud representation. Subsequently, a similar method will be applied to different forms of data representation like raster and voxel visualisation. Finally, the results will be evaluated.

In addition, historical data will be processed and included in the GIS.

Combining the above-listed information will substantiate the visual analysis to detect unmarked burial sites in the study area.

The following section will present the research objectives more in detail. Secondly, the theoretical background is listed. The applied research method and the results of its application are elaborated before the conclusion of this research is drawn.

2 Research objectives, questions, and scope

2.1 Research objectives

The objective of this research is to gain information out of three-dimensional subsurface data. It will explore the data processing possibilities within a GIS. The focus is set on developing an automated data classification to segment data representing underground anomalies. Therefore, this method will help the visual interpretation of subsurface properties. Moreover, a workflow will be created for 1) remotely detecting and identifying subsurface objects and properties, 2) integrating multiple data types (layers), and 3) the application in other fields of interest.

2.2 Aim and research questions

The target of this research is the classification and visualisation of GPR data in a GIS. For this, spatial data collection at the Jewish Cemetery in The Hague took place. Their interpretation will aim at the detection of unmarked burials at the site. As a supplement, topographic data will be generated and integrated into the system. Automated classification via a deep learning framework will be explored on the GPR data for the segmentation part.

Based on this, the main research question is:

“To what level is the classification by deep learning and visual interpretation of Ground Penetrating Radar data possible in a GIS?”

Four sub-questions are generated to help in answering this central question:

- What GPR data representation format (3D point cloud, raster, voxel) is suitable for a GIS interpretation?
- Can the generated deep learning results (e.g., pattern/value detection and extraction) enhance the interpretation of subsurface data to detect unmarked burials?
- How can topographical survey data be beneficial for the interpretation process of subsurface data?
- To what extent do historical data sets of the research area support data interpretation of GPR data to detect unmarked graves?

2.3 Scope of research

The focus of this project is the development of the automated classification of 3D data sets in a GIS. For this purpose, GPR data were collected from multiple sample areas on the Jewish cemetery in The Hague. The entire dimension of the graveyard covers about 6,2ha. The acquisition of the GPR data and their post-processing requires expert knowledge. The faculty of Civil Engineering and Geosciences of the TU Delft is providing the equipment and knowledge transfer. In addition, topographic surface data sets and historical data will be collected to complement the GPR data. The topographic data will be limited to the corresponding sample areas of the site. The research on historical data sources lead to the Foundation of the Jewish Cemetery in The Hague (*Stichting Instandhouding Joodse begraafplaats 's-Gravenhage*) and the author of the of the book *‘De joodse begraafplaats an de Scheveningseweg in Den Haag – Geschiedenes en Restauratieverslag’* (Enthoven, drs. Francine Puttmann *et al.*, 1992). The historical archive of the municipality Den Haag (*Haags Gemeentearchief*), and the provider of geospatial data sets *Dotka*, also contributes information in the form of aerial images documenting the area’s historical development.

A classification of the topographic surface properties would enhance the adjustments during the post-processing phase of the GPR data, but this will not be a priority of this project due to time restrictions.

2.4 Hypothesis

The hypothesis of this research is that GPR and topographic data provide valuable information to remotely detect unmarked burial sites. Deep learning methods applied in a GIS will enhance the identification of the burial locations by classifying the GPR data. The process will be based on the GPR data properties. It is expected that the quality of the results strongly depends on the spatial resolution and the intensity sensitivity of the GPR data. Another parameter to consider is that the deep learning methodology relies on the training data's quality. This fact is one critical part of the project and will be essential for a successful implementation.

The research hypothesis will be tested at the Jewish Cemetery in The Hague. Gained knowledge of the underground layout will support the cemetery's management for future activities and help cross-reference it with existing records in the archives.

Furthermore, a successful development of the workflow will allow an application in different applications and scales.

3 Theoretical background

3.1 Deep learning

Thinking machines have stimulated people's imagination even before the first programmable computer was developed (Menabrea and Lovelace, 1842). In present times the terms *artificial intelligence (AI)*, *machine learning (ML)*, and *deep learning (DL)* are omnipresent as those systems are in use by many daily life applications. As the use fields appear to be infinite, research on this matter will be active (Bengio, Goodfellow and Courville, 2017). However, deep learning is only one subcategory in AI technology and represents one single style embedded in multiple learning concepts (Figure 4).

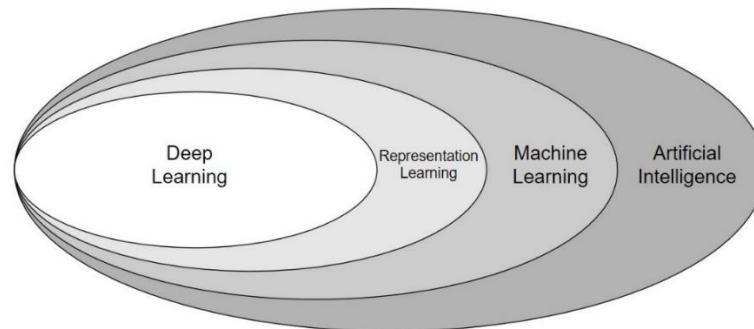


Figure 4: Venn diagram representing the hierarchical learning concepts of AI technology (Author).

According to Chang and Lin, deep learning is a machine learning sector that automatically selects and extracts features from data sets (Chang and Lin, 2011). Even though it emerges as a relatively new technology, DL has a long development history, starting in the '40s of the last century. The hypothesis did undergo multiple phases starting from *cybernetics* (1940-1960'), over *connectionism* (1980-1995), to the renaissance as *deep learning* (2006-present) (Bengio, Goodfellow and Courville, 2017). Two significant developments triggered the stated rebirth. First, there is the remarkable improvement of the computing power of hard- and software components. This fact allowed the models to expand in size and complexity. And secondly, the availability of a suitable amount of training data is a central foundation of the *DL* concept. Since the advent of *Big Data*, numerous data sets are existing to train models and apply them. Machine learning can be separated into two learning theories, the supervised and the unsupervised models. *Deep learning* requires labelled training data to teach the algorithm and is therefore learning in a supervised manner. The setup of *DL* references the structure and the functionality of a brain's neural network, the learning process is based on the analysis of the provided data (Thakur *et al.*, 2019). One definition of *DL* is given by (Patterson and Gibson, 2017) in, 'Deep learning as a neural network with a large number of parameters and layers in one of four fundamental network architectures.'. Those four architectures are:

- Unsupervised pre-trained networks
- Convolutional neural networks
- Recurrent neural networks
- Recursive neural networks

The automatic extraction of features is one area where *DL* unfolds its full capability and advantage compared with conventional machine learning. In this process, the network selects properties of a dataset as parameters to classify the data consistently. This development made it obsolete to label manually large feature sets for data classification and exceeded the traditional processes in terms of speed and accuracy.

Convolutional neural networks and recurrent neural networks are *DL* architectures in use in multiple fields. Examples of such areas are recognising faces in imagery, bioinformatic data classification, and sound analytics. The examination of written words as discrete textual units is yet another *DL* application facilitating our daily lives (Patterson and Gibson, 2017).

The neural network architecture is described as a network constructed by several simple connected processors named ‘neurons’ (Schmidhuber, 2015). The figure below sketches an overview of the relationship between the three main components of such a network: the *input layer*, the *hidden layers*, and the *output layer* (Figure 5).

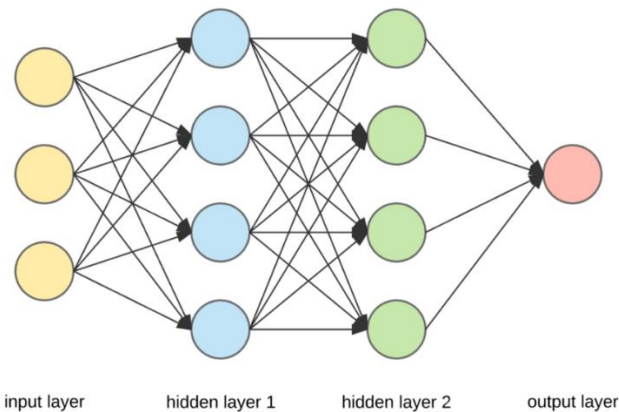


Figure 5: Schematic representation of the layered structure in a neural network design (Sorokina, 2017)

3.1.1 Convolutional Neural Networks (CNN)

Convolutional neural networks are designed to learn higher-order features in the dataset through convolutions. The successful implementation in computer vision (e.g., image recognition) is one example where its potential got acknowledged by the public (Patterson and Gibson, 2017). A standard *CNN* is a composition of an input layer, convolutional layer, pooling layer, and fully connected layers.

The convolutional layer is always the first hidden layer in a CNN and acts as a filtering entity to determine output features. Compared to the input layer, the filter is of smaller spatial size and extracts compressed information of the input (Ke *et al.*, 2018).

The pooling layer succeeds the convoluted layer and reduces the spatial size. This downsampling filter enhances the data processing steps on one side and supports the detection of characteristic features in the input layer on the other side. The latter is also maintaining the training process of the model (Dasgupta *et al.*, 2018). The pooling layer does compute a fixed function of the input layer and thus does not involve any parameters. Commonly, the pooling layer uses a max operation (max pooling) and extracts the maximum value in the predefined filter matrix. Another function used in a pooling layer is computing the average value of the filter region as an output feature.

The finishing part of the network is the fully connected layer. It serves as a class scoring unit used for the output layer at the end of the network. The layer's dimension is defined by $[1 \times 1 \times N]$, where N is the number of evaluated output classes. Each neuron of this layer is connected with each neuron of the previous layer. Fully connected layers transform the input data based on the activations of the input volume and parameters (e.g., weights, biases etc.) (Patterson and Gibson, 2017).

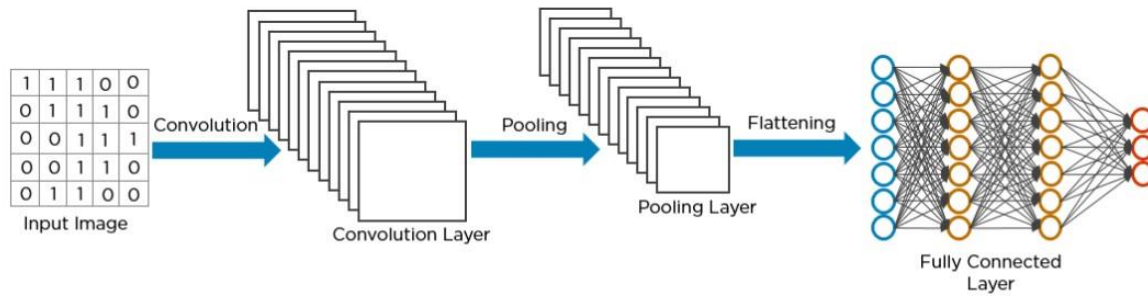


Figure 6: Conceptual structure of a convolutional neural network consisting of the input layer, convolutional layer, pooling layer, and fully connected layer (Biswal, 2020)

The efficacy of *CNN* is based on the regular grid structure of the assessed data sets (e.g. images, sound) (Malinverni *et al.*, 2019). The focus on information on the grid through a small filter allows the detection of variations like edges and corners for a robust classification (Winiwarter *et al.*, 2019). On the contrary, tabular data sets from a relational database management system (RDBMS) that do not have any spatial relationship can be examined via a *CNN*. The columnar data are exported with a fixed materialised view and provide the required structure (Patterson and Gibson, 2017).

3.2 Point cloud features

Point clouds in their present representation emerged in the early stage of the 2000s' with the introduction of automated active light-sensing technology (light detection and ranging [LIDAR]). The acquisition of 3D points in sub-millimetre precision, in large amounts and at a meagre cost of time, created an enthusiasm of a new era where traditional survey methodologies (e.g., photogrammetry) seemed to be redundant and fated to vanish. The disillusion followed promptly with realising that processing the voluminous data sets with their indiscrete points does not only require high-performance hardware. The processing software also did not have the capacities (algorithms) to automatic obtain valuable information. In the preceding two decades, several developments in the field of point clouds improved the acquisition and processing procedures. This includes the renaissance of the photogrammetric method through dense image matching (DMI) algorithms and the detection of primitive features (e.g., plane, cylinder, etc.). However, the automated interpretation of point cloud data is disadvantaged compared with the production developments in quantity and quality.

Point cloud data have a solid semantic representation and hold statistical properties unaffected by certain transformations. The data can be classified into two conventional types, intrinsic and extrinsic features. A simple classification with human interaction involved can provide a solution for elementary problems. However, it is considered not suitable for more complex scenarios (Chen *et al.*, 2020).

3.2.1 Deep learning applied to 3D data features

With the afore in mind, using a *CNN* to segment point cloud features seems contradictory due to their architectural characteristics. An enormous number of points, distributed in a three-dimensional space in an irregular, non-gridded manner, and that with varying density is not suitable to be assessed by a supervised *CNN*. According to Malinverni *et al.*, the status quo of processing point clouds is the labour-intensive creation of 3D models by domain experts. The general workflow includes the following steps (Malinverni *et al.*, 2019):

1. Identifying architectural elements of interest via visual assessment.
2. A representation as a parametric 3D object for each element
3. The accurate placement of the 3D object into a 3D scene.
4. Classify the placed element according to the requirements of the domain and annotate additional information of various types.

To accelerate this process and to avoid the quality loss due to generalisation, an automated workflow is needed to enable a direct conversion from point cloud data into a profound output.

Zhang et al. is stating that deep learning algorithms can learn how to segment feature representations from 3D point clouds (Zhang *et al.*, 2018).

One key element is to introduce a type of spatial structure to the data. Several research projects tried to solve this issue by transforming the point cloud into various 2D regularised representations. For example, Hu and Yuan (2016) examined the option to label every single point of the cloud based on its rasterised neighbourhood. Subsequently, the accumulated images were used for digit recognition and classification via a *CNN* (Hu and Yuan, 2016).

The similar idea is to create textured 3D models from the point cloud as a source for 2D segmentation. Supervised machine learning on such data was applied successfully to classify 3D cultural heritage models (Grilli, Özdemir and Remondino, 2019).

An alternative approach to structure the data is the transformation into voxel format. A standard technique is assigning the average value of the points to the enclosing single grid cell block. Then, the resulting 3D grid is labelled via a 3D *CNN* through upscaling techniques (Badrinarayanan, Kendall and Cipolla, 2017).

So far, the stated procedures have the method in common to translate the irregular point cloud into a suitable isotropy data structure. However, recent research has shown that a *CNN* can directly be applied to the unordered 3D features of a point cloud. Qi et al. introduced *PointNet*, where symmetric functions are employed to the input data to achieve stability (Qi, Su, *et al.*, 2017). The enhanced version, *PointNet++*, describes the local structures amongst the point using the *PointNet* function in a hierarchical manner (Qi, Yi, *et al.*, 2017). *DeepSets* (Zaheer *et al.*, 2017) and *SO-Net* (Li, Chen and Lee, 2018) are other systems following a similar approach but do not gain the output quality of *PointNet++* (Chen *et al.*, 2020). Li et al. promote feature learning with an X-transformation from the input point clouds to, 1.) assign weights based on the input features and 2.) permute the points in an order with latent and canonical properties (Figure 7). *PointCNN* uses this technique based on the help of graph structures to facilitate the learning process (Li *et al.*, 2018).

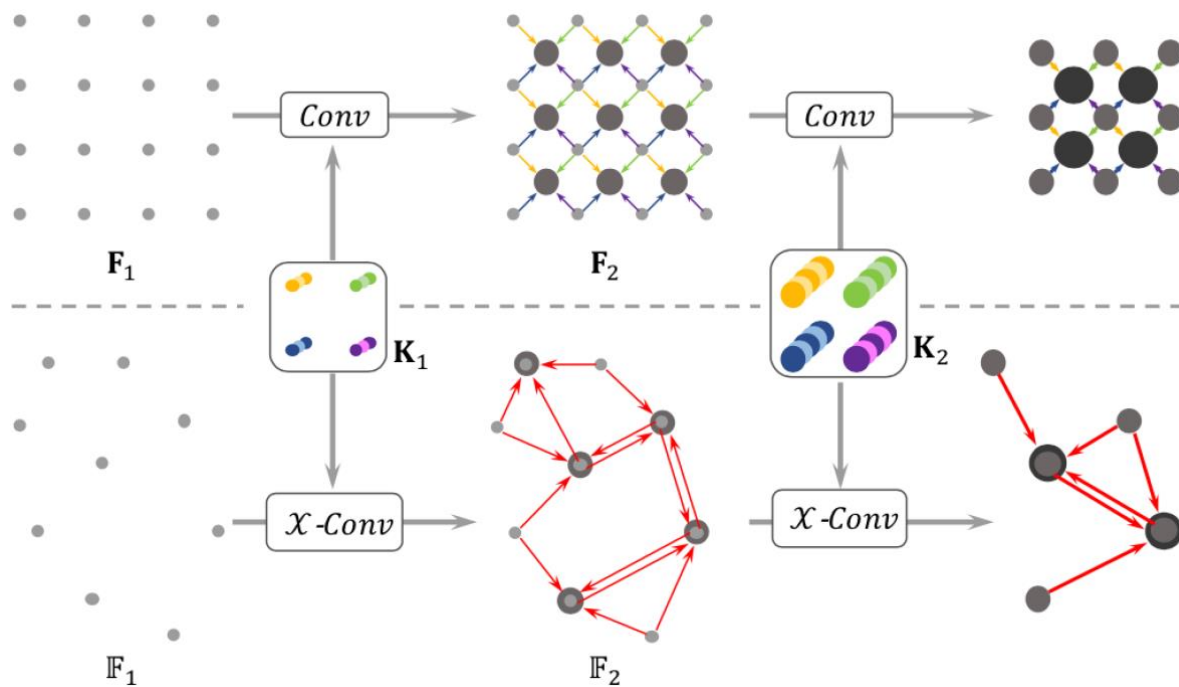


Figure 7: The upper section demonstrates the convolution on structured (gridded) data where each iteration reduces the resolution of the patch and increases the number of channels (visualised by the dot thickness). The below section describes the X-Convolution (transformation) function to 'gain' or 'project' the neighbours' properties to less representing points with supplemented information load (Li *et al.*, 2018).

One characteristic behaviour of a convolution layer is the dependency of the output on the inputs point order. That is the main barrier for a convolution operation on points from a point cloud. The transformation function assigns a random order to the input points to achieve an order-independent feature. Points within the point cloud are functioning as representative points that define the local neighbourhood. Those points need to be suitable for either the ‘projection’ (classification) or ‘aggregation’ (segmentation) of the information. For this, a random-down sampling or corresponding farthest-point function is applied to the input data. The \mathcal{X} -Convolution transformation matrix operates in the created local environments to collect the neighbouring points together with their associated features for convolution (Li *et al.*, 2018). The algorithm is described in the equation below:

$$F_p = \mathcal{X}\text{-Conv}(\mathbf{K}, p, \mathbf{P}, \mathbf{F}) = \text{Conv}(\mathbf{K}, \text{MLP}(\mathcal{X} - p) \times [\text{MLP}_\delta(\mathcal{X} - p)\mathbf{F}])$$

$\mathbf{P} = (p_1, p_2, \dots, p_k)$ are the neighbouring points with the corresponding features $\mathbf{F} = (f_1, f_2, \dots, f_k)$. The trainable kernels are represented by the variable \mathbf{K} . Based on these parameters, the ‘projected’ or ‘aggregated’ features F_p will be convoluted on the representative point p . Table 1 defines the parameters and processing steps in more detail.

ALGORITHM 1: \mathcal{X} -Conv Operator

Input : $\mathbf{K}, p, \mathbf{P}, \mathbf{F}$
Output : \mathbf{F}_p

1: $\mathbf{P}' \leftarrow \mathbf{P} - p$

2: $\mathbf{F}_\delta \leftarrow \text{MLP}_\delta(\mathbf{P}')$

3: $\mathbf{F}_* \leftarrow [\mathbf{F}_\delta, \mathbf{F}]$

4: $\mathcal{X} \leftarrow \text{MLP}(\mathbf{P}')$

5: $\mathbf{F}_\mathcal{X} \leftarrow \mathcal{X} \times \mathbf{F}_*$

6: $\mathbf{F}_p \leftarrow \text{Conv}(\mathbf{K}, \mathbf{F}_\mathcal{X})$

▷ Features “projected”, or “aggregated”, into representative point p

▷ Move \mathbf{P} to local coordinate system of p

▷ **Individually** lift each point into C_δ dimensional space

▷ Concatenate \mathbf{F}_δ and \mathbf{F} , \mathbf{F}_* is a $K \times (C_\delta + C_1)$ matrix

▷ Learn the $K \times K$ \mathcal{X} -transformation matrix

▷ Weight and permute \mathbf{F}_* with the learnt \mathcal{X}

▷ Finally, typical convolution between \mathbf{K} and $\mathbf{F}_\mathcal{X}$

Table 1: \mathcal{X} -Convolution operation parameter and description (Li *et al.*, 2018).

Chen *et al.* evaluated *PointCNN* in comparison with other segmentation and classification procedures. It appeared to be robust in the application and achieved a high efficiency. Furthermore, the necessary training effort for the algorithm was to a lesser extent than for the other assessed systems (Chen *et al.*, 2020).

3.3 Automatic data analysis of GPR data

The automated processing of GPR data to detect and classify distinctive features, like subsurface objects or soil properties, has been investigated in multiple studies. Most research concentrate on the analysis of the 2D-radargrams (B-scans) (Figure 1) and (Figure 8). Those two-dimensional raster presentations are part of the conventional, manual interpretation of subsurface data. Next to data quality, the accuracy of the result also depends on the examiner’s knowledge and experience. The computerised detection of buried objects (e.g., landmines), appearing as hyperbolas in the GPR images, has been explored by Gader *et al.* by applying a rule-based feature extraction and adaptive whitening process. A Hidden Markov Model (HMM) facilitated noise reduction in the radargram images (Gader, Lee and Wilson, 2004). Subsequently, the system was improved by adding an adaptive pre-processing unit to an HMM sensor, followed by a corrective training component and background model identifier. As a result, the recognition performance could be increased to 50% (Frigui, Ho and Gader, 2005). An alternative workflow was created by Pasolli *et al.* as a combination of pre-processing, segmentation, object (pattern) recognition and material recognition. To classify the materials of the detected objects, a Support Vector Machine (SVM) was utilised. A performance increase in object detection was discovered by using synthetic images, where the background noise has been removed, as input data and the hyperbolas vectorised (Pasolli *et al.*, 2009).

A real-time GPR application to detect land mines was introduced by Kovalenko et al. that operates on one dimension. The method solely uses the single wave signals (A-scan) detected by the GPR receiver. It performs its analysis based on waveforms representing different classes of buried objects (Kovalenko, Yarovoy and Lighthart, 2007).

CNN's emergence in image processing applications was consequently used on GPR B-scan raster images (Figure 1) for their interpretation enhancement. Ozkaya et al. present a tailored solution named *Convolutional Support Vector Machine (CSVM)*, like *CNN*, a series of convolution and pooling layers, but with the difference that an *SVM* filter is used to generate input feature maps. Several *CSVM* models were tested and showed the ability to automatically calculate the shape of the buried object and the soil type (Ozkaya et al., 2020). One step further from a dimensional perspective is the method combining multiple B-scans from a multichannel GPR to form intersectional C-scans and D-scans with individual layer orientation (Figure 8). The combination of the layers creates a 3D GPR data representation. Based on this, a feature can be reviewed at various vertical and horizontal slice images. Finally, collating all those representations in one standardised 2D-image format serves as input data to a *CNN* (Figure 8). The introduction of more perspectives decreased false predictions compared with the processing of single framed 2D data sets and, therefore, strengthened the system's robustness (Kim et al., 2021).

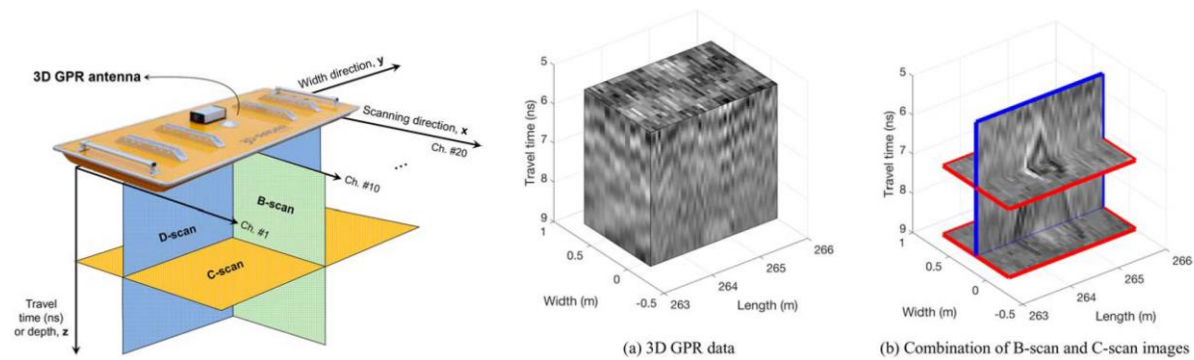


Figure 8: LEFT - Multichannel GPR (Channel#N) with the representation of the orthogonal scan planes B, C, and D. MIDDLE and RIGHT – (a) 3D data block and (b) combination of 1 B-scan and 2 C scan images (Kim et al., 2021).

As stated at the beginning of this chapter and based on the literature review, most systems are founded on classifying one or multiple 2D GPR images. Also, the integration in a GIS is part of only a few research projects. GIS has been mainly used to present and disseminate results, but not for the actual processing (Tabarro et al., 2017). An approach based on 3D point cloud as input data for a *CNN* was not investigated yet and is therefore appropriate for this research as a novum.

4 Methodology

This section will introduce the methodologies that were applied during this research project. First, data sourced by multiple sensor systems and formats will be collected, transformed, and integrated into one system. Then, data processes will be developed to explore their functionality for identifying unmarked burial sites. Those last steps will be taken for different input data formats, namely 3D point cloud-, and raster- format. An evaluation of the results will follow by reference data sets.

The sub-questions will be separately examined and evaluated.

This section begins with the presentation of the study area. The available data sets and the workflow with applied software will be presented subsequently.

4.1 Case study area

The Jewish Cemetery in The Hague is placed right opposite the Peace Palace and thus belongs to the historic part of the city (Figure 9). Founded in 1694, it was located on the north-western edge of the existing settlement area on one side and adjoining the dune area on the other (Annex A, Figure 30). The cemetery is a national monument of the Netherlands managed and maintained by the foundation for the Preservation of the Jewish Cemetery in The Hague (*Stichting Instandhouding Joodse begraafplaats' s-Gravenhage*).

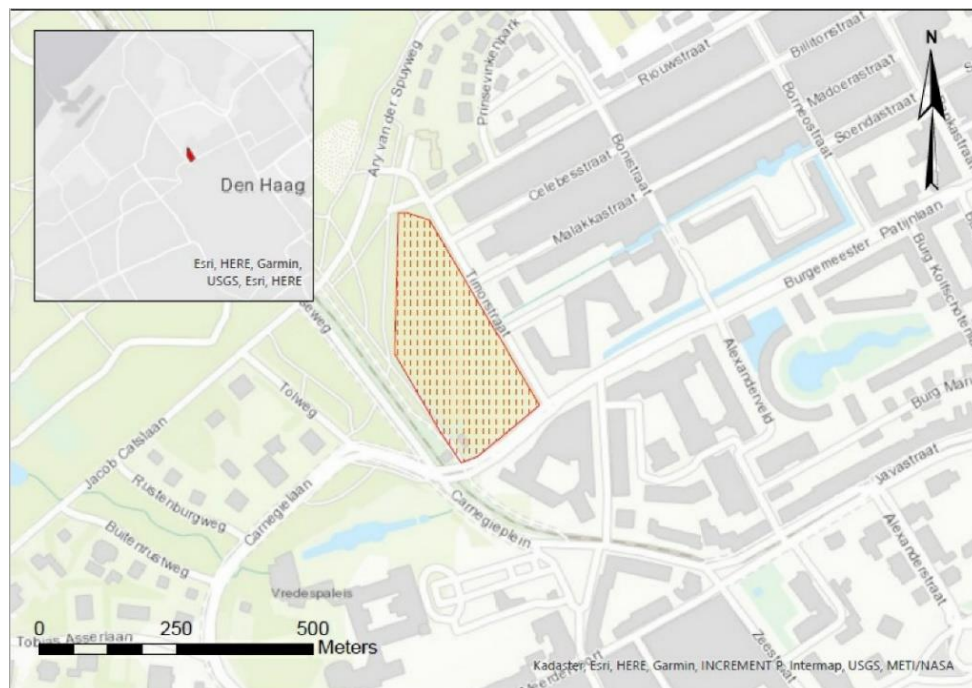


Figure 9: Location of the research area, the Jewish Cemetery in The Hague (Author).

According to the records, there are more than 10,000 people buried. However, only 2,860 tombstones indicate the position and provide further details of the deceased. Several reasons caused this discrepancy. One is that tombstones were not affordable to most of the Jewish community. Another cause is the natural disappearance over the last centuries. Many markers got lost during the second world war as the German army used the stones for trench construction and as base plates for military equipment (e.g., ant-aircraft guns). Moreover, at the end of the German occupation, a British airstrike in this area also decimated their amount (Foundation Jewish Cemetery, 2021).

Since the middle of the 1980s, comprehensive restoration activities were initiated by the foundation established in 1984. A throughout documentation campaign (Figure 10) of the remaining tombstones occurred between 1986 and 1988, resulting in a register accessible to the public (Foundation Jewish Cemetery, 2021).

Today the cemetery is only used exceptionally since its capacity reached its limit at the beginning of the last century. Given the Jewish tradition that no graves or tombstones can be removed, a second area serves as a graveyard for the Jewish community.

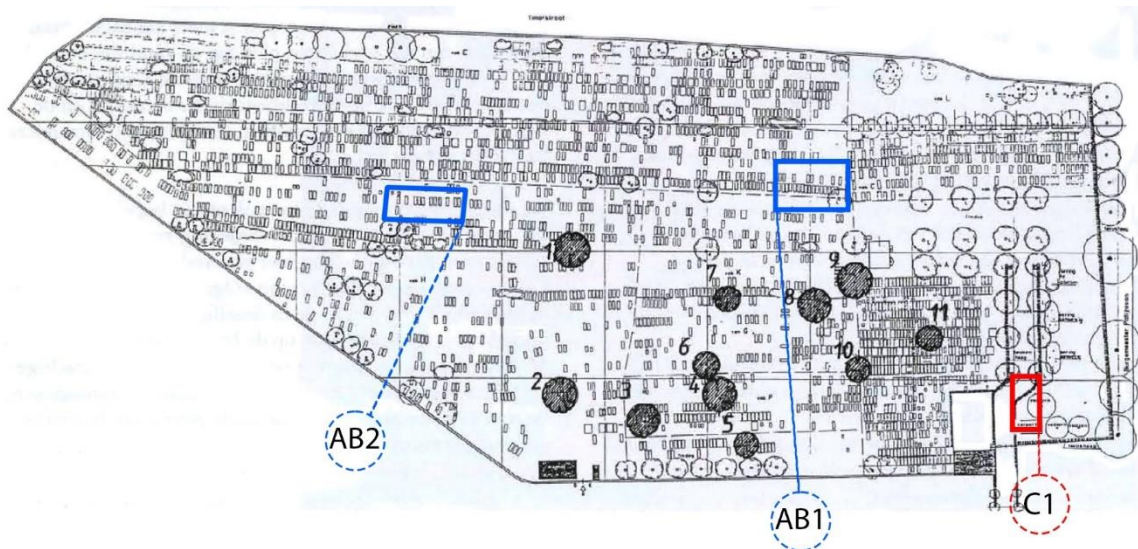


Figure 10: Overview layout displaying the location of the existing tombstones and higher vegetation (Enthoven, Drs. Francine Puttmann et al., 1992). Indication of research sample areas (blue and red) (Author).

As noted in the *scope*, due to its size, the total area of the cemetery does not allow an examination in this project.

The objective to use a deep learning algorithm to classify data requires a training data set to start. For this purpose, two sample areas (*Area AB*) are needed to cover both types of sub-terrain, known and unknown. For example, that can be along a line of known (marked) burial sites and a (partially) empty patch or along a known line with an alleged open area (e.g., access path) adjacent. Such a hybrid style will be beneficial to 1) train the algorithm based on the set classification [*Area AB1*] and 2) evaluate the results on a similar segment [*Area AB2*]. Images of the two patches with their spatial location and dimensions are in Figure 31 (Annex B).

Area C1 contain complete unknown properties of the underground occupancy and will be assessed by the model. According to local information the plot was a former gardening patch and is located southwest of the entrance. All three sample areas are indicated in Figure 10.

4.2 Data collection, processing, and integration

Different data sources are used for this research, as listed below. The majority is accumulated on the site by surveys with various sensor systems. Additionally, historical data are added to the collected information to enhance interpretation. A general overview of the workflow is displayed in Figure 11 below.

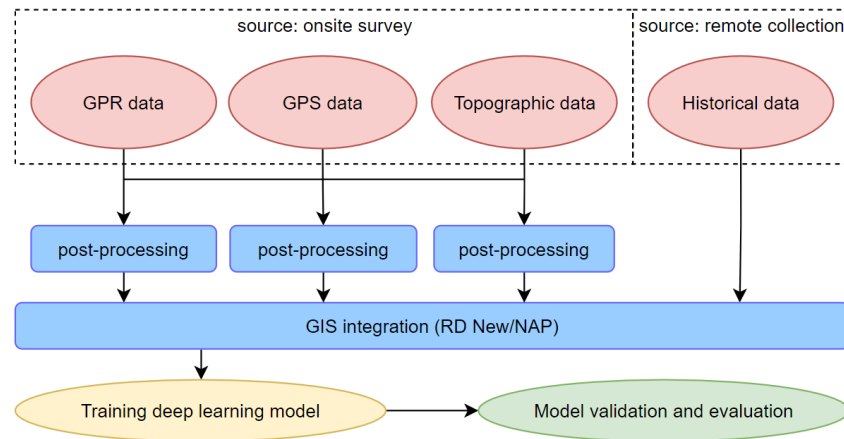


Figure 11: General methodology workflow from data collection to model evaluation (Author)

4.2.1 Ground Penetrating Radar (GPR) data

Remote sensing technology is employed at the TU Delft Faculty of Civil Engineering and Geosciences research projects. They use the GPR sensor *pulseEKKO GPR* by *Sensors & Software Inc.* The applied frequency of the sensor in this project is 500 MHz. According to the literature and the fabricators directive, this wavelength is suitable for forensic and archaeological investigation. It is a compromise between the data resolution and ground penetration depth. The pre-selected areas (AB1 and AB2) data acquisition was executed in a grid spaced approach covering the corresponding area. The survey resulted in 40 and 24 single scan lines at a 0,25m spacing. For efficiency reasons, the lines were collected in a 'zigzag'-pattern, meaning that every second line was measured towards the grids baseline. A graphical overview of the survey setup AB1 is located below in Figure 12.

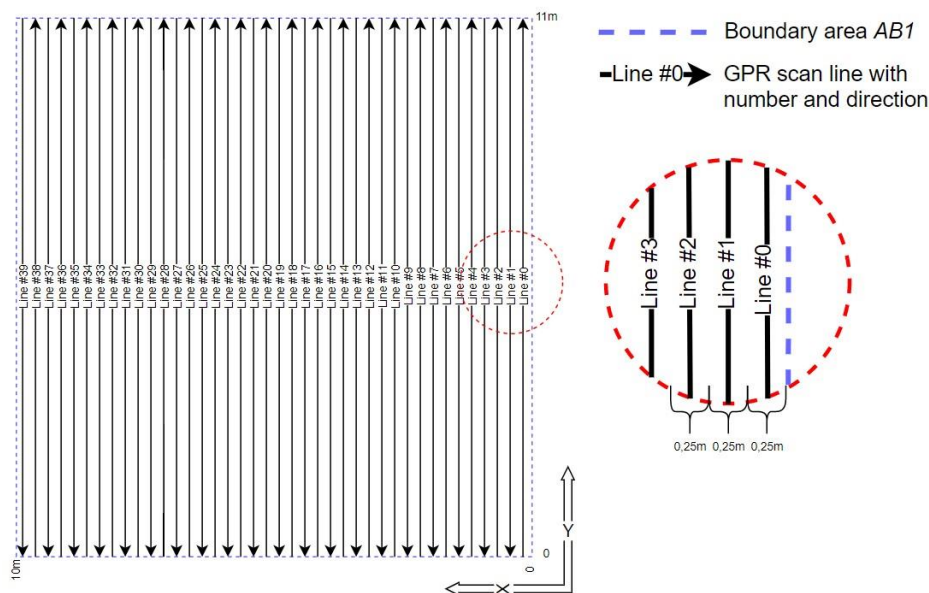


Figure 12: Overview GPR survey set up for sample area AB1 (not to scale) with dimension line width (detail image) (Author).

The post-processing of the raw data was performed with the software *EKKO_Project (Version 6)*. The package is equipped with various filter options to increase data quality. As a result, negative impacts caused by effects like attenuation and environment noise (e.g., *Nyquist* noise) could be minimised. The settings of the applied filter parameter are also helpful to highlight the signals of interest (e.g., reflection hyperbolas) in the 2D-radargram (B-scan) Figure 13. The B-Scans (24 lines) of the survey area AB2 are displayed in Annex A (Figure 32 and Figure 33).

The GPR data sets need to undergo a three-dimensional spatial correction as their collection occurred without accompanying documentation of the spatial position and elevation of the sensor. The 3-parameter transformation (2 translation and 1 rotation) for the horizontal adjustment is based on the GPS measurements on each corner of the two grids and individual tie-points (see 4.2.2).

A correction of the vertical dimension is applied on a line-by-line basis (Figure 13). The reason for this is that the GPR data are defined along a local horizontal surface line, where neither topographic properties (height differences) nor datum-based elevations are embedded. The matching vertical profile obtains the topographic parameters for the line adjustment. Those values are, in turn, extracted from the georeferenced topographic surface data (see 4.2.3).

Additional processing steps include the application of various filter to emphasize the data signal information. Such filters remove unwanted low frequencies from the GPR trace (*Dewow*) or apply functions of increasing gain (*SEC2*) to counteract the attenuation effect on the signal. A signal velocity of 0,100m/ns is used to determine the vertical distance (depth) of the reflections and transmissions.

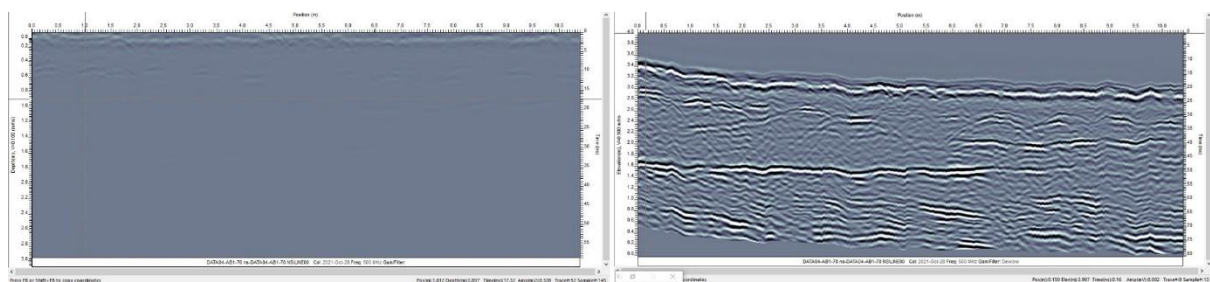


Figure 13: 2D GPR radargrams (B-scans) of survey Line #0 sample area AB1 of unprocessed raw data (left side) and after the application of filter and topographic correction (right side) (EKKO_Project ject line view, Author)

The processed GPR data grids can be provided in various spatial formats to export for further usage. At the beginning of this project, it was intended to obtain a topographically corrected point cloud as an input data set for the subsequent classification process via deep learning. This was not possible due to technical problems which will be elaborated later in section 6.3. As a result of those issues a dataset was used instead where a list (CSV - data format) included the point cloud data with the assigned attributes. The points are a result of an interpolation between all scan-lines (B-scans) creating a block-shaped point cloud with an average density of 125.000 points per cubic meter (Annex G, Figure 36).

4.2.2 Global Position System (GPS) data

Integrating of all different data sets implies a transformation into a coordinate system of a higher order. Therefore, the national reference system of The Netherlands, *Amersfoort RD New*, and the vertical datum, *NAP (Normaal Amsterdams Peil)*, were used as they were deemed the best fit for the study area. Reference points were obtained with GPS fast static surveys in each sample area (*AB1*, *AB2*, *C1*). The post-processed coordinates are listed in Annex D.

4.2.3 Topographical Surface data

There are three main reasons why the topographic data of the site surface will be relevant for this project. One is the georeferencing of the GPR data to integrate them into a horizontal and vertical reference system. Secondly, there is the topographic correction of the GPR data needed. And lastly, the data will supplement the analysis in the GIS.

An Unmanned Aerial Vehicle (UAV) was utilised for the data collection. The aerial images were processed by the software *Pix4Dmapper* by applying photogrammetric methodologies. The resulting data sets (3D point cloud, mesh, ortho-photo, etc.) can be used for further examination and will enhance, together with the other collected data, the overall interpretation.

4.2.4 Historical Data

There are additional historical data sources next to the documents of the foundation. A query in the archives of the municipality The Hague (*Gemeentearchief*) resulted in the finding of detailed layout plans of the cemetery dating 1990. They are documenting the locations of the single grave with individual numbering, which corresponds to the index numbers in the database of the Jewish foundation (Annex E, Figure 34, Table 11).

Aerial observations by the British Royal Air Force occurred in the region during the second world war monitoring the activities of the occupying German forces. This imagery reveals significant actions in the cemetery (Annex F, Figure 35). An overlay with the present situation gives valuable information about the actual layout. Frequent aerial imaging flights by the Dutch government document the topographical and topological developments in the post-war era. A review of public and commercial archives led to the imagery that will complement the collection to fill the gap until the cemetery's restoration (Table 2).

Year	Description	Size (pixel)	Data format
1983	1983KAARTBLAD30_057, 's-Gravenhage	11375x11862	TIFF
1982	1982_3105_DEN HAAG_3303, 's-Gravenhage	9394x11345	JP2
1976	1976KAARTBLAD30_088, 's-Gravenhage	8775x9560	TIFF
1971	1971KAARTBLAD30_092, 's-Gravenhage	8462x9231	TIFF
1966	1966KAARTBLAD30_089, 's-Gravenhage	8400x9114	TIFF
1954	1954KAARTBLAD30_124, 's-Gravenhage	8371x8349	TIFF
1945	106G-4570_3043, 's-Gravenhage	9394x11345	JP2
1945	526_01_5043, 's-Gravenhage	6644x6570	JP2

Table 2: Historical aerial imagery properties (Dotka).

4.3 Research Software

Several software packages are utilised to perform this research. One milestone to fulfil is the integration of the different data sets in a GIS, namely *ArcGIS Pro (Version 2.9.0)*. Once achieved, the processing tools in a GIS will be explored to interpret the data by detecting unmarked grave locations. The focus is set on the deep learning functionality in *ArcGIS Pro* to train and apply a *CNN* model for the automated 3D data classification. This functionality uses the *PointCNN* algorithm presented in section 3.2.1.

Before this stage, several pre-processing procedures are necessary for individual data sets.

In preparing the GPR data for the import into the GIS, processes like filter, adjustments, and corrections are applied in the software *EKKO_Project (Version 6)*. The program facilitates the export of 3D point cloud data in different formats (e.g., ASCII, CSV, LAS). Furthermore, it also provides tools for anomaly detection and interpretation. Those insights are used for verification of the results produced in the GIS. The UAV raster data set will be processed by the application *Pix4D_mapper (Version 4.5.6)* using the Dense Image Matching (DIM) method. The gained 3D-point clouds for each sample area (*AB1* and *AB2*) give details of the surface properties like undulations, colours, and elevation. An open-source tool, *Cloud Compare (Version 2.12)*, was used to extract topographic data for further processing purposes. The details of the entire workflow are presented in section 4.4 and in Figure 14

4.4 Project Workflow

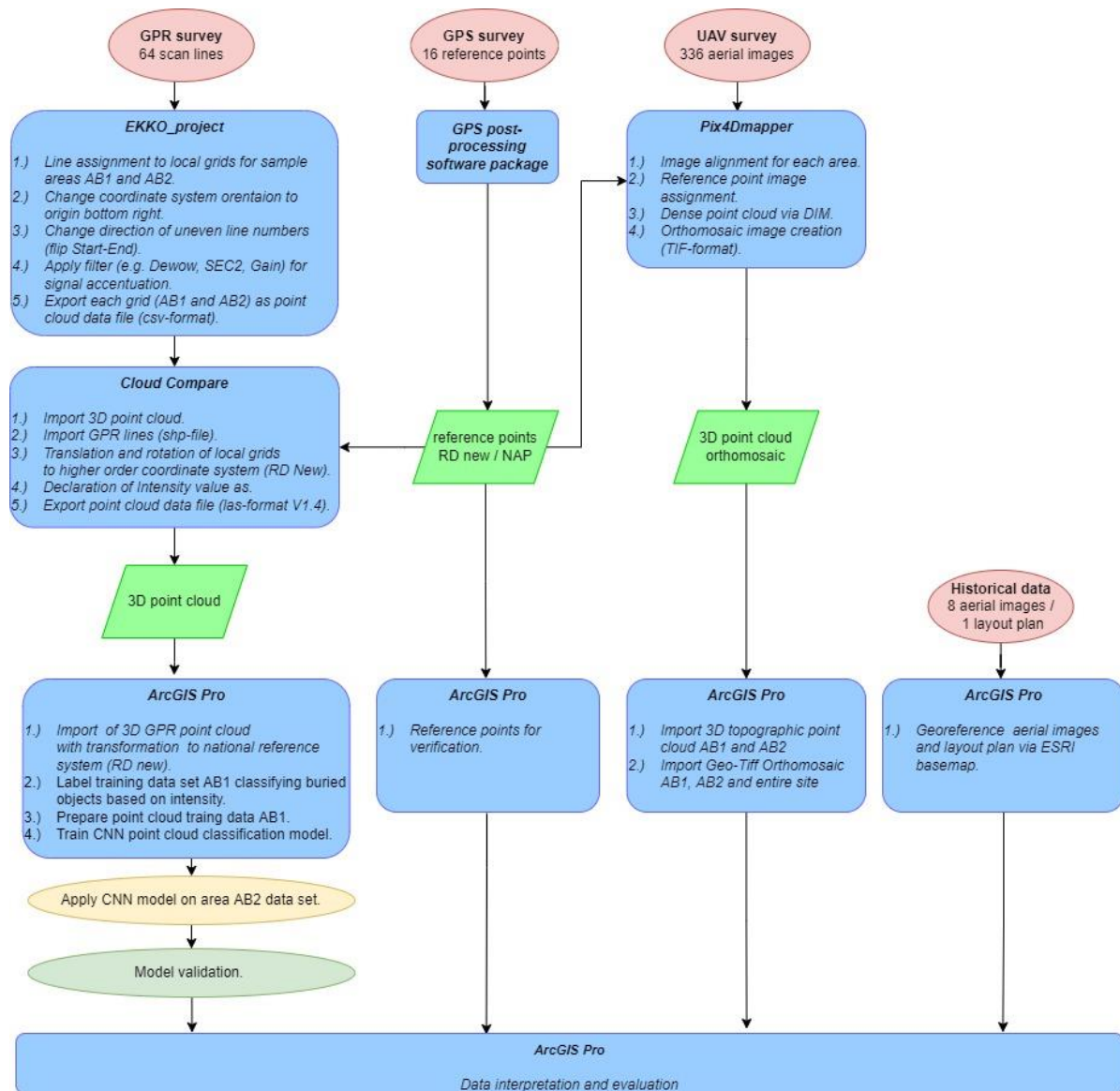


Figure 14: Detailed workflow diagram starting from the data collection until the interpretation and evaluation in a GIS. (Author).

4.5 Semiautomated point cloud classification by CNN model

Several prerequisites and processes must be set up and undertaken before applying a model to a data set. Those phases can be structured into the following sequence:

1. Assessment of the provided training data to their quality. If necessary, manual data labelling and cleaning (e.g., noise filtering) need to be commenced.
2. Preparation of the point cloud data generates the data that will serve to train and validate the CNN model. Various parameters are defined that impact the subsequent training step at this stage.
3. Train and validate the point cloud classification model by implementing the *PointCNN* architecture as a deep learning framework.
4. Classifying a point cloud using the trained *PointCNN* model.

Each phase for this project is elaborated on in the following sub-sections.

4.5.1 Assessment and labelling of GPR point cloud

A general rule states that the quality of a *CNN* model strongly depends on the quality of its training data. Therefore, for automated classification, it is necessary that the features of interest are clearly identifiable, free from artefacts (e.g., signal noise, random errors, etc.) and correctly labelled.

The designated data set is the representation of the sample area *AB1*, which is available in a rectangular-shaped block (Length / Width / Depth – 10,5m / 10m / 1,7m) containing 23.053.608 points (Annex G). Next to the location information, each point inherits a value between 0-138, representing its intensity.

A pre-labelled data set is necessary to obtain a *CNN* model for automated point cloud classification as training and validating input. For this purpose, the data set was classified into three categories using the GIS predefined classes of the *LAS-Format 1.4* of the *ASPRS* (*ASPRS*, 2002).

The classification assignment for the project purpose and the applied selection criteria are listed in Table 3 below.

Class #	Classification <i>LAS-Format 1.4</i>	Project Classification	Selection criteria	Intensity values/ Depth (m)
0	Created, Never Classified	Natural soil layer and changes of the sediment properties (e.g., clay layer)	Subsurface area with low-intensity values assigned in the first layer and all other data below burial depth.	0 – 30 / 0 - 1m
				0 – 138 / 1 - 1.73m
1	Unclassified	Burial locations	Subsurface area with high-intensity values in burial depth.	15 – 138 / 0.2 - 1m
2	Ground	Gravestone	Surface and near subsurface area with distinctive intensity values assigned.	5 – 15 / 0 – 0.2m

Table 3: Classification assignment and selection criteria for point cloud labelling. (Author)

The manual labelling process was undertaken in *ArcGIS Pro* using the profile tool that enables the user to create vertical slices of the point cloud. For the sample area *AB1*, 4cm thick slices were assessed to define the desired classes. According to information received by the management of the cemetery, the standard burial depth is about 110cm. Consequently, the buried coffins are occupying areas between ca. 0.6 and 1.1m depth. Considering this, only points with significant intensity values located in this vertical range were selected and classified as a burial location (class #1).

The quantitative distribution of manually classified points in *AB1* resulted in the following numbers:

- Class 0: 21.502.219 points (93,27%)
- Class 1: 1.304.629 points (5,66%)
- Class 2: 246.760 points (1,07%)

Figure 15 demonstrates the differences of the B-scan radargrams and the vertical sections of the 3D point cloud that is coloured by the intensity values. Also, the manual classification process is displayed by the points selected as burial locations (highlighted by blue colour).

The successive Figure 16 points at the location of the demonstrated scan line #32 of sample area *AB1*. The orthomosaic of the area is overlaid with the gridlines (red) of the GPR survey (Figure 12). The explored vertical cross section is highlighted yellow where no crossing of any gravestones on the surface is visible.

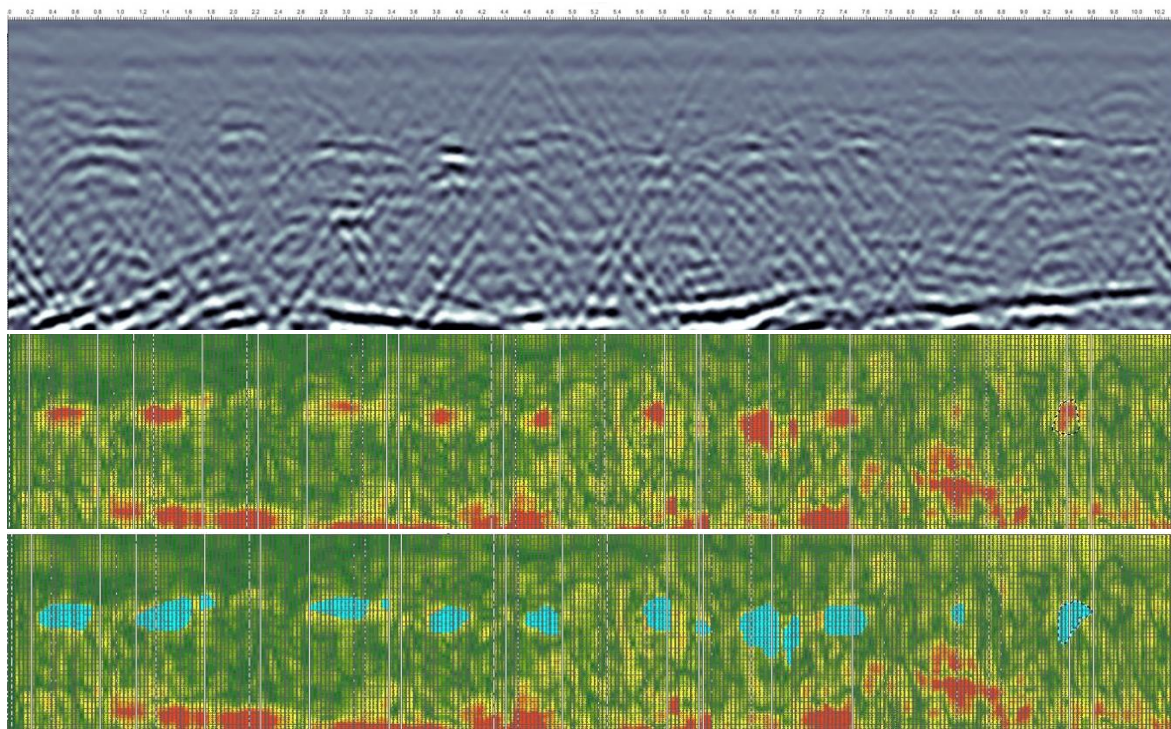


Figure 15: TOP - 2D-Radargram (B-scan) of GPR data line #32. MIDDLE: - Visualisation of vertical profile slice (4cm thickness) of the point cloud in *ArcGIS Pro* coloured based on the assigned intensity values. BOTTOM: - Manually selected (blue highlighted) points classified as class #1 - burial locations (Table 3). The displayed cross section shows burial with no gravestone present (Author).



Figure 16: Orthomosaic of area AB1 with overlaid GPR grid and highlighted line #32 (Author).

The imbalance of the class proportions is not favourable for the subsequent training phase and will be discussed in section 5.3 . A common approach to resolving this issue is introducing more classes to the data set. Unfortunately, the content of the subsurface data set does not provide many additional options (e.g., confining layers) to reduce the volume of the class #0.

Another issue discovered during the labelling process is that the selection of the burial locations was complex in areas where gravestones and other objects caused signal disturbances (Annex H).

4.5.2 Prepare point cloud training data

After the manual labelling process, the data set met the qualitative standards to create a *CNN* model. For this process, a training and validation part is necessary. According to Viswambharan and Singh, partitioning the sample area *AB1* with the proportion 2/3 and 1/3 was done parallel to the x-axis, ensuring a distribution of the different scenarios (marked, unmarked graves and undisturbed soil) (Figure 17) (Viswambharan and Singh, 2020).

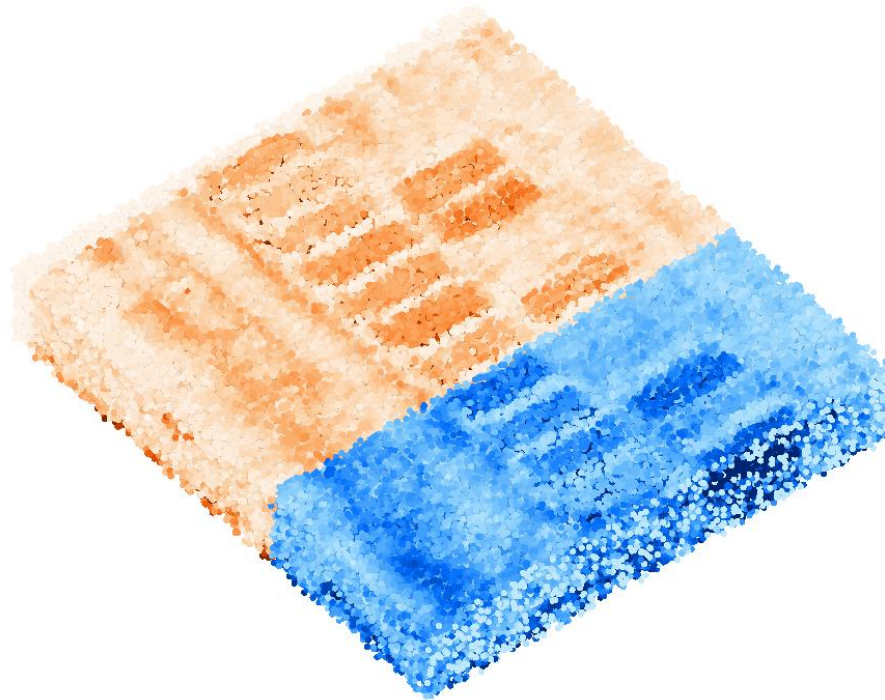


Figure 17: Isometric view on sample area *AB1* with the partition of training data (red) and validation data (blue) (Author).

Before the training phase can be initiated several parameter settings will need to be defined for the input data. The dimensions of the sample blocks and their maximum number of contained points need to be defined.

The cubical subdivision, block, will be applied over the entire data set, each representing a single sample. That is the smallest entity that can be fed to the learning algorithm. The block size with the resulting volume describes the room of content (points) that can be addressed in this entity. In addition, it does consequently set the total number of samples (blocks) over the data set. Previous research suggests choosing a block size that sufficiently captures parts of the objects of classification and their surrounding context. That means it is only necessary to capture object feature characteristics by one block for identification. It is therefore not needed to envelop an object in its entirety. So, the burials require a certain proportion of applicable points in the block to get identified and classified.

The maximum point number is a threshold for computational reasons. If exceeded, a subdivision into smaller blocks in this area ensures the usage of all the available data for the training.

The preparation process results in a Hierarchical Data Format (HDF5-format) designed to support large, heterogeneous, and complex data sets (Group, 2015). The block size and the max. point number set the properties of the HDF5-tiles that will be applied to the training and validation data set.

In this project, the main features of interest, burial locations, are not distinctive in their feature appearance in shape or dimension. Therefore, multiple scenarios are created by applying various pre-set parameters. The employed values are listed in Table 4. For every prepared data set, the number of *Blocks* and their encased points is visualised in two bar charts, separately for the training and validation data sets. The diagrams are in Annex J.

# prep. Data	Class Codes of Interest	Block Size (cm)	Block Point Limit	Output file (pcdt-format)
1	1, 2	50	8192	trainingdataAB1_50_8192
2	1;2	30	12000	trainingdataAB1_30_12000
3	1	20	1000	trainingdataAB1_20_1000
4	1	50	20000	trainingdataAB1_50_20000_1
5	1	100	10000	trainingdataAB1_100_10000_1

Table 4: Parameter settings of the training data preparation (Author).

4.6 Train model for point cloud classification

The *PointCNN* learning framework will be employed on the prepared point cloud data set during this phase. Next to the standard input parameter (e.g., *input point cloud* and *model output location*), several additional variables can be used to influence the training result. The selection of the attribute(-s) assigned to each point strongly affects the training results. The *Intensity* attribute is selected in this research as it is born from the signal value of the GPR measurement. Other possible options (e.g., *Return number*, *Number of returns*, *RGB values*, etc.) are not possible to use. That is due to the unique properties of the sub-surface data and the difference in their acquisition compared with the standard task of using data obtained by LIDAR sensors.

Another practical parameter is the minimum threshold of points present in one block. The number allows ignoring blocks with a low amount of information included. As a result, those blocks will be neglected for the training and accelerate the overall process. About the different input training sets with their block sizes and maximum point numbers, the were multiple training sessions with various minimum point number values. Their influence is described in the following section 5.1. Considering the quadratic block shape of the data set (area *AB1*) and the equal density of points within, this parameter is expected not significantly to impact the results.

As a default, the *PointCNN* framework will assign equal weights to each class when commencing the model training. However, the weights are allocated by designating one or multiple class codes with a higher-grade priority. The remaining classes are declared as *Background Class Code*. For example, in the project case, the class of interest is *Class # 1 – Unclassified*. Therefore, the weight is solely set in this while the other two classes are not considered separately. Two training sessions also involved *Class #2 – Ground* to detect the gravestones. The results are discussed in the next section, 5.3.

Two more vital parameters to mention are the *Number of Epochs* and the *Batch size*. The number of epochs defines how often the learning algorithm will go through the total training data set.

The size of the batches, on the other hand, determines the number of samples (blocks) the training data set is divided. At the end of each batch (group of blocks), the predictions are compared to the reference variables and an error is calculated. That error is the base for improving the algorithm and updating the model. The division can be done in one or multiple batches. (Brownlee, 2018). A general rule states that a smaller batch size achieves the best training stability but at a higher cost of computational power as a trade-off due to the high number of iterations for each epoch. Former experiments obtained the best results with a batch size of 32 or smaller. The batch size is a tuning factor between the stability of the learning process and the pace of the learning procedure (Brownlee,

2020). During this research, batch sizes 2 and 4 are used for the training. Any attempt to raise the number significantly (e.g., 32 and higher) overburdened the applied hardware components. The cause is that a smaller batch size makes it more feasible to fit one batch of training data in a memory (8.0GB of GPU memory available). A remark needs be made that the used samples (blocks) are containing already a large amount of data (points) what makes a group (batch) considerably hard to process at once. A comparison of the two different sizes and their impact are listed in Table 6 and Chapter 5.1. Other options are possible to control the training process (e.g., Class remapping, pre-trained model refinement, learning rate, *etc.*). But since they are not relevant to the purpose of this project, they will not be described in detail.

During and after the training, several statistical values monitor and evaluate the process on its efficiency and quality (Figure 18).

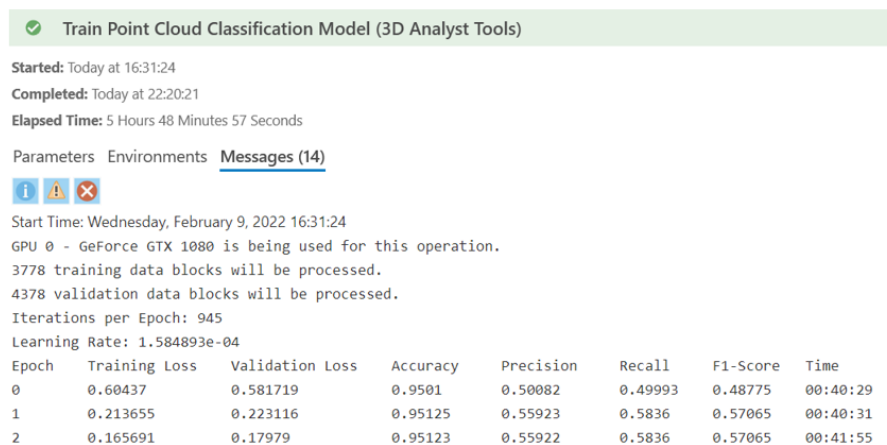


Figure 18: Training settings overview and epoch parameter (Author).

As displayed in the figure above, seven values are listed for each epoch in a model training session. Whereas the time amount mirrors the duration for each epoch, the remaining variables indicate the model development's dynamic progress.

A description of the parameters is given below:

- **Training Loss**

The general rule applies for loss-values, "The lower the better.". The value reflects how well the model learned during this epoch. It also shows if a model can learn on the given training data set. A model that is unable to learn or needs more training data is considered an *underfitting* model (Figure 19).

- **Validation Loss**

The validation is here on how good the model applies to other data. For example, a model with an excellent (low) training loss, but a high or unstable validation loss is a well-trained model that is not useful for other data sets. Such a constellation results in an *overfitting* model (Figure 19).

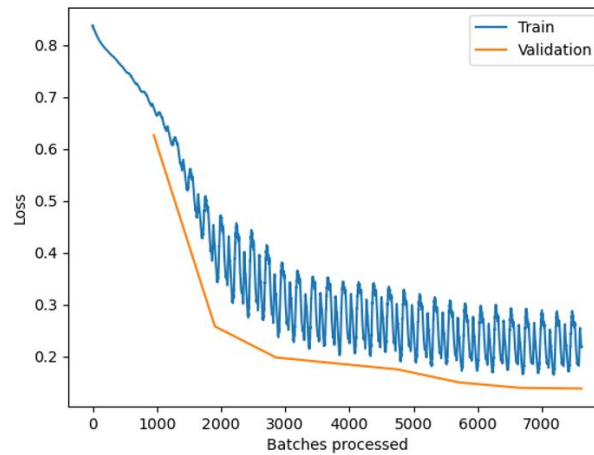


Figure 19: Training- and Validation-Loss curve obtained by a training session (Author).

- **Accuracy**

Accuracy refers to the relationship between the correct predicted points and the total amount of assessed points. The value is read as a percentage indicator (Korstanje, 2021).

$$Accuracy = \frac{\# \text{ correct predictions}}{\# \text{ total predictions}}$$

It is essential to acknowledge that this metric is only valuable if all classes are equally distributed. The higher an imbalance appears, the less representation on the model performance on the minor classes occur. The training set for this project is highly imbalanced (see 4.5.1). Therefore, this value is interpreted carefully.

- **Precision**

This metric connects to the positive predictions of the model. It determines the fraction of the correct positive predictions (True Positives [*TP*]) within the total amount of positive predictions (True Positives [*TP*] and False Positives [*FP*]).

$$Precision = \frac{TP}{TP + FP}$$

The resulting fraction allows discovering how precise the model detects the True Positives. A noisy or unprecise model might make many positive predictions, but the selection method also includes many positives that are negatives in the real world. On the other hand, a precise model also referred to as a 'pure model', can only make true positive predictions. But it does not mean that it can find all of them on a data set. (Korstanje, 2021).

- **Recall**

Like the *Precision* parameter, the *Recall* variable considers the *TP* cases. But with the difference, it sees them relative to the actual total number of positive entities in the training set. The number of False Negatives (*FN*) is introduced to the equation.

$$Recall = \frac{TP}{TP + FN}$$

A higher value implies that the model can find all positive entities in the data with the possibility of a few *FP* included. On the other hand, a low *Recall* number proves that the model cannot select the majority or none of the positive entities.

- **F1-Score**

The *F1-Score* defines the ‘harmonic mean’ of the *Precision* and the *Recall* value. So by that description, it is the average of both values with an equal allocation of their weights.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Aiming for a high *F1-Score*, the two input values must be high. An average score means that one of the variables contains a low number. If both, *Precision* and *Recall*, are at a low level, the logical consequence will be a low *F1-Score*.

All the described parameters provide valuable information on the training progress and the performance of the created model. One important part is that those numbers are based on the results of all classes of the input data and therefore provide only an overall (average) overview of the model calculated during this epoch. A more detailed list of the model performance on every class for every epoch is provided after the entire training process (Table 5).

This list allows a model validation in reference to the classes of interest.

EPOCH	CLASS CODE	CLASS DESCRIPTION	PRECISION	RECALL	F1_SCORE
0	0	background	1	0.951248	0.973581
0	1	Unclassified	0	0	0
1	0	background	0.988735	0.976999	0.982705
1	1	Unclassified	0.376601	0.594365	0.423036
2	0	background	0.991951	0.968170	0.979473
2	1	Unclassified	0.281337	0.602922	0.343744
3	0	background	0.984933	0.976880	0.980685
3	1	Unclassified	0.419117	0.584009	0.441467
4	0	background	0.977365	0.986359	0.981627
4	1	Unclassified	0.537314	0.558298	0.503505
5	0	background	0.987388	0.980115	0.983618
5	1	Unclassified	0.418363	0.621385	0.458401
6	0	background	0.982251	0.986159	0.984024
6	1	Unclassified	0.488741	0.600064	0.491868
7	0	background	0.982359	0.986390	0.984195
7	1	Unclassified	0.495740	0.605361	0.497486

Table 5: List displaying the training performance parameter for each epoch (8) and classes (2) (Author).

Based on the values it is possible to choose an individual model from a specific epoch where the class of interest is detected the best. All training sessions in this project are focused on achieving a high *Recall* value (default software setting). Therefore, their values are usually higher compared to the *Precision* value.

4.7 Evaluation and testing of the classification model

A comparison evaluates the created model with the reference data set from sample area *AB2*. Those reference data are generated by filtering and labelling the raw data points in the software *Cloud Compare* by applying intensity- and elevation- filters. Due to time restrictions, this method has been chosen instead of the manual labelling process as applied on area *AB1*. As a result, the reference class will have a to some extent different quality than the training data. However, it is sufficient to compare the various models and their results.

Next to the visual evaluation (Figure 20), a geoprocessing tool in *ArcGIS Pro* allows to assess the created model by re-classifying the reference data set and compare it to the ground truth. The same method has been used in the validation process during the training session. As a result, the parameters *Precision*, *Recall* and *F1-Score* indicate the model's performance. Furthermore, a confusion matrix (Figure 21) provides an overview of the distribution of *TP*, *FP*, *TN*, and *FN* predictions.

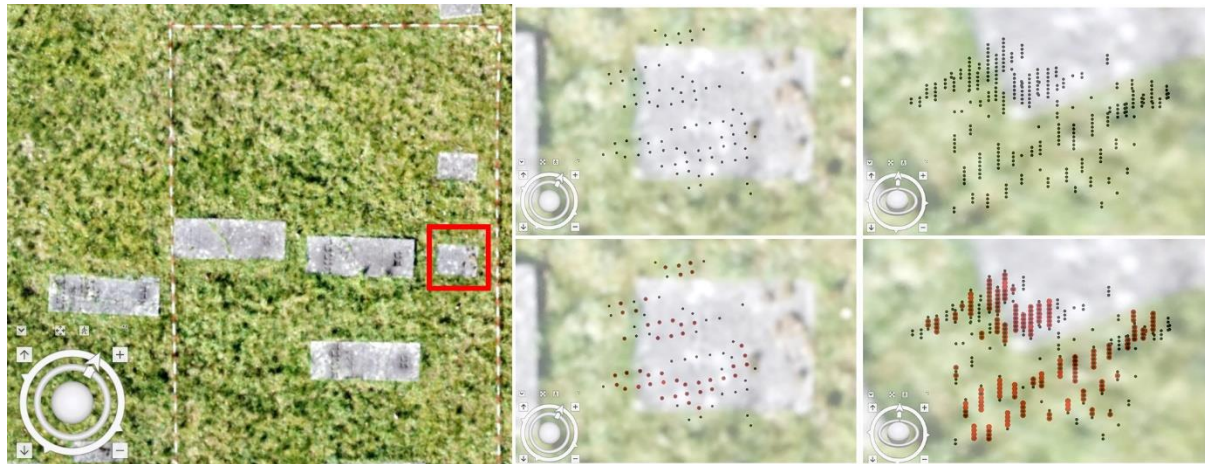


Figure 20: Visual evaluation of one model output (orange points) compared to the reference point cloud (black points). LEFT – Location of assessed part *patch 1* (red frame) in sample area *AB2*. CENTER – Orthogonal view on the reference points (black) and the predicted (red) points. RIGHT - Oblique view on the reference points (black) without and with the predicted points (red) (Author).

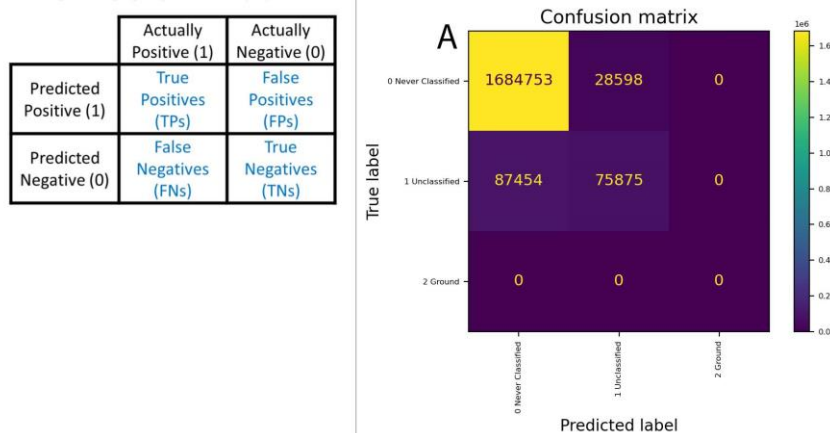


Figure 21: LEFT – Schematic structure of confusion matrix (Draelos, 2019). RIGHT – Confusion matrix of model A displaying the TPs and FPs between classes #0, #1, and #2 (see 5.3) (Author).

Confusion matrices employ the prediction of a model to a reference test dataset. On the one hand, it serves to highlight the strengths and weaknesses of a model, and on the other hand, it can help to compare multiple models with each other. Based on this outcome, the user can choose the best model that predicts the class of interest.

5 Results

This chapter presents the results of the project. The first part is dedicated to the *deep learning* approach via the *PointCNN* framework applied on the ground surface data from the sample area *AB1*. Then, based on the achieved parameters, an evaluation and comparison of the produced models are held. A visual comparison of the sample area *AB2* and quantitative analysis via confusion matrices will assess the achieved models.

The subsequent parts present the data derived from the UAV imagery and the collection of historical data in the form of multiple areal images overlayed to the subsurface data.

5.1 Created classification models and validation parameters

Nine different classification models were created out of the five prepared training data sets (Table 6: Training settings and resulting validation parameters for nine models (A-J) obtained from sample area *AB1* (Author).Table 6). Those are analysed by comparing the overall model's validation parameters and the performance on detecting the *classification #1 – unassigned – burial location*.

The parameters reveal a significant difference between the overall model validation and the numbers refer to the detection on *class #1*. This situation is caused by the imbalance between the training and validation data set classes. However, by comparing the values directly, a difference in the *Precision (P)*, *Recall (R)*, and *F1-Score (F1)* numbers allow the conclusion that there is a strong dependency on the prepared input data (e.g., block size) and the setup (e.g., batch size) of the training session itself.

Session **A** marked the start of the training. There the default settings of the *ArcGIS* deep learning package have been used. While the overall model values seem to be average compared to the other models, it can be said that the overall quality of the model is sufficient as (*P*) and (*R*) both exceed a threshold of 60%. Furthermore, predicting the class of interest in the data set is one of the best models with (*P*) and (*R*) around 65%.

Significant settings for session **A** are the block size of 50cm, a maximal point number of 8.192 and an equal weighting on all classes. The default setup of 25 epochs resulted in a processing duration of 23,5 hours. However, the subsequent models were trained using eight epochs for efficiency reasons.

Training sessions **B** and **C** derive the weakest values of all models. The relatively high *Training Loss* (in **B** - 0,1454), the far below average (*P*) value (in **C** - 0.1804), and the corresponding other numbers are making those two models the least suitable to apply.

Equal parameters of the two sets are the block size of 30cm and a maximum point number of 12.000. The differences are in the weighting of the classes for the training process. While session **B** uses an even distribution (equal to session **A**), the setting in session **C** concentrates on class #1. Therefore, it is surprising that the validation parameters of the class code appear to be the weakest.

In addition, the sizes of the batches differ between the sets (**B**=2 and **C**=4). This change decreases processing time due to the maximum usage of the available computational power.

The following two models, **D** and **E**, are almost equal in their process settings, with the only difference in the batch sizes (**D**=2 and **E**=4). On one side there was again a reduction (ca. 30%) of the training time. However, this temporal gain is traded in exchange for the model's quality. While model **D** got the best overall validation values (except the training loss) of all the nine models, doubling the batch size in the training process for model **E** causes a quality loss of the values below the average. The same effect occurs correspondingly to the validation parameters of class #1. However, both models do not perform well in detecting the burial locations.

One unique characteristic of the input data is that the block size of 20cm is the smallest tested during this project. Subsequently, the maximum number of contained points is the lowest, with 1000 points per block. This dimensional property can be the reason for the light performance on the class prediction. Whereas the background class #0 is not dependable on the size of the blocks because it does not

include any features and is therefore considered as 'white noise' around the elements of interest. The characteristics of the buried features need to be enclosed by one box and its number of containing points. In this case, it appears that the smaller block sizes (20cm and 30cm) are already close or below the threshold of enclosing the distinctive features of a buried object. Therefore, models **B**, **C**, **D**, and **E** are less efficient on the class #1 prediction.

The models **F** and **G** analysis is challenging due to their controversial outcome. Both are based on the same training input data (block size = 100cm; max. points no. = 10.000) and training setup (min. points no. = 8000, epochs = 8). With the only difference in the size of the batches (**F**=2 and **G**=4). The results vary in their quality, especially in the class validation parameters. While both models can achieve the best values for the class #1 *Recall* variable (ca. 70%), model **F**'s Precision value differs significantly from model **G** (0.3901 vs 0.528). The same phenomenon is present when comparing the overall model validation parameter. The model with the bigger batch size is reaching better results in the direct comparison. This observation contradicts the prior conclusion made in this project with models **D** and **E**. On the other side, it aligns with various research in the area of *CNNs*. The influence of the batch size parameter has been tested multiple times in the field of image recognition. The result is that the accuracy increased with the number of training samples (batches) (Radiuk, 2017; Takase, 2021). Confirming additional training sessions during this project could not be performed due to mentioned computational limitations.

Lastly, the models **H** and **J** will be examined. Again, the only difference between them is the size of the batches (**H**=2 and **J**=4). The quality gain caused by this increment is also apparent. Furthermore, both models perform well according to the validation parameters. The input data settings are 50cm of block size and a maximum point number of 20.000 per block. The latter value is the highest of all the training data in this research and therefore can be one reason for the good results. Another aspect of this can be that the block size is suitable for discovering the features of interest.

Like for all models, it would need additional sessions with varying parameters (e.g., number of epochs, max. number of points/block, batch size, etc.) to corroborate the findings and refine one of the models to the most fitting for classifying the subsurface features of interest.

The loss curves visualising the training process for each session are presented in (Annex K, Figure 39).

Training Classification Model													
Training Session	input # prep. Data	Block size (cm)	min. Point/Block	Training data blocks	Validation data blocks	Iterations per epoch	Class Code of Interest	Backgrd Class code	Class Descr.	Model Selection Criteria	Epochs (E)	batch size	duration (hrs)
A	1	50	4000	5498	4634	2742	NA	NA	0;1;2	Recall	25	2	23.5
B	2	30	10000	3968	3789	1946	NA	NA	0;1;2	Recall	8	2	9.8
C	2	30	10000	3968	3789	952	1	0	0;1	Recall	8	4	6.3
D	3	20	500	3778	4378	1884	1	0	0;1	Recall	8	2	8.3
E	3	20	500	3778	4378	945	1	0	0;1	Recall	8	4	5.8
F	5	100	8000	5333	3613	2665	1	0	0;1	Recall	8	2	9.5
G	5	1000	8000	5333	3613	1333	1	0	0;1	Recall	8	4	6.2
H	4	50	5000	4885	3968	2435	1	0	0;1	Recall	8	2	9.3
J	4	50	5000	4885	3968	1218	1	0	0;1	Recall	8	4	5.5

Overall Model Validation Parameter											Class Code 1 Validation Parameter						
Training Session	min. Training Loss	E	min. Valid. Loss	E	max. Precision	E	max. Recall	E	max. F1-Score	E	max. Precision	E	max. Recall	E	max. F1-Score	E	
A	0.038	6	0.1256	23	0.7155	4	0.7196	18	0.6744	14	0.6748	8	0.6512	4	0.5373	24	
B	0.1454	5	0.1754	6	0.5592	4	0.5835	8	0.5706	6	0.5373	5	0.6213	6	0.5035	5	
C	0.0495	8	0.1379	8	0.8076	5	0.5932	8	0.6231	8	0.1804	8	0.6059	5	0.2567	8	
D	0.0831	3	0.0661	8	0.8538	6	0.8147	5	0.8109	8	0.5373	5	0.6213	6	0.5035	5	
E	0.115	6	0.122	8	0.8085	5	0.7371	7	0.7178	7	0.4975	7	0.6244	5	0.4539	7	
F	0.0594	8	0.0715	7	0.8378	8	0.6959	8	0.7289	8	0.3901	3	0.7004	2	0.464	8	
G	0.0515	6	0.0751	8	0.833	3	0.7567	7	0.7629	7	0.528	7	0.6961	3	0.5377	7	
H	0.0351	8	0.0675	7	0.8386	8	0.787	8	0.7916	8	0.5504	3	0.6601	2	0.5278	8	
J	0.0571	6	0.1074	6	0.8406	6	0.7886	8	0.7576	7	0.5984	8	0.675	6	0.5131	7	

Table 6: Training settings and resulting validation parameters for nine models (A-J) obtained from sample area AB1 (Author).

5.2 Visual evaluation of the classification models

As described in section 4.7, a visual evaluation of the nine obtained models helps to envision the performance of predicting the features representing burial locations. As reference serves the sample area *AB2*, which was labelled via intensity thresholds.

Next to the entire sample area *AB2*, the patch **P1** (red bounding box Annex B, Figure 20) was selected in the reference dataset to make a visual comparison. The assessed cluster represents the anomalies caused by a smaller burial location that is marked with a gravestone.

The image below displays the visualisation of *AB2* by multiple horizontal time slices in the software *EKKO_Project* (Figure 22). Another overview of the model performance on the entire area *AB2* is presented in Figure 23 and **Figure 24** with the reference point cloud (black colour) presented on the beginning.

The figures, Figure 25 and Figure 26, provide two separate perspectives (orthogonal and oblique view) on the sample patch **P1**. The top row shows again the reference points (black colour) based on the intensity classification. The images are the predictions made by the nine individual models (Model **A** – **J**).

The visual inspection provides insights where the predictions correspond to the ground truth (True Positives) and where a mismatch (False Positives).

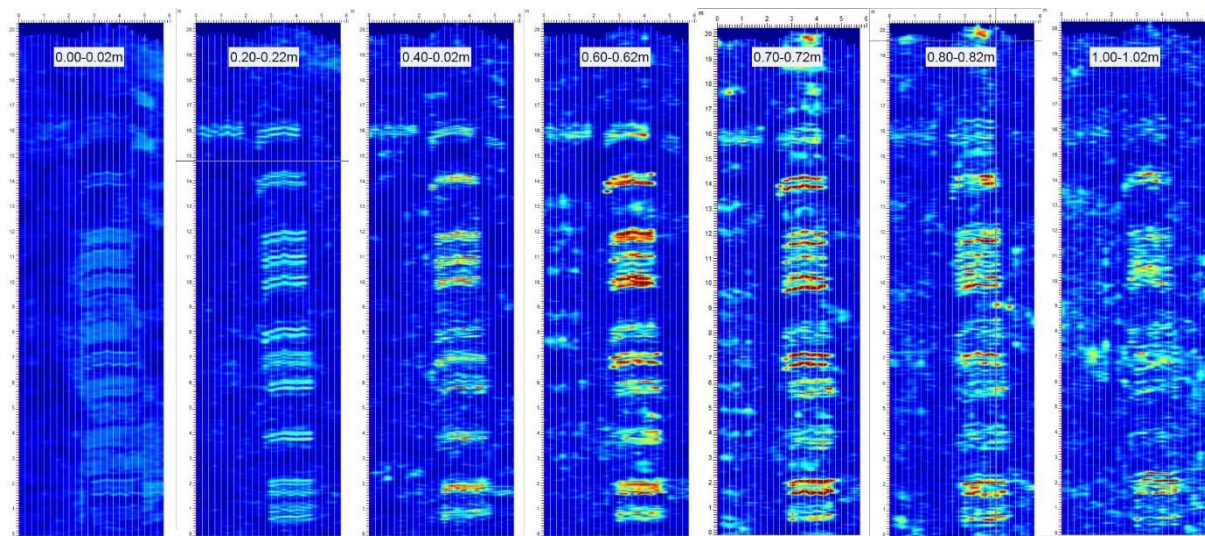


Figure 22: Seven horizontal time slices of area *AB2* demonstrating the interpretation option in the GPR processing software *EKKO_Project*. The slice width and depth are indicated in the top part of the strips. Areas between the survey lines are the results of an interpolation process performed in *EKKO_Project*. The colourisation is based on the signal intensity values (Author).

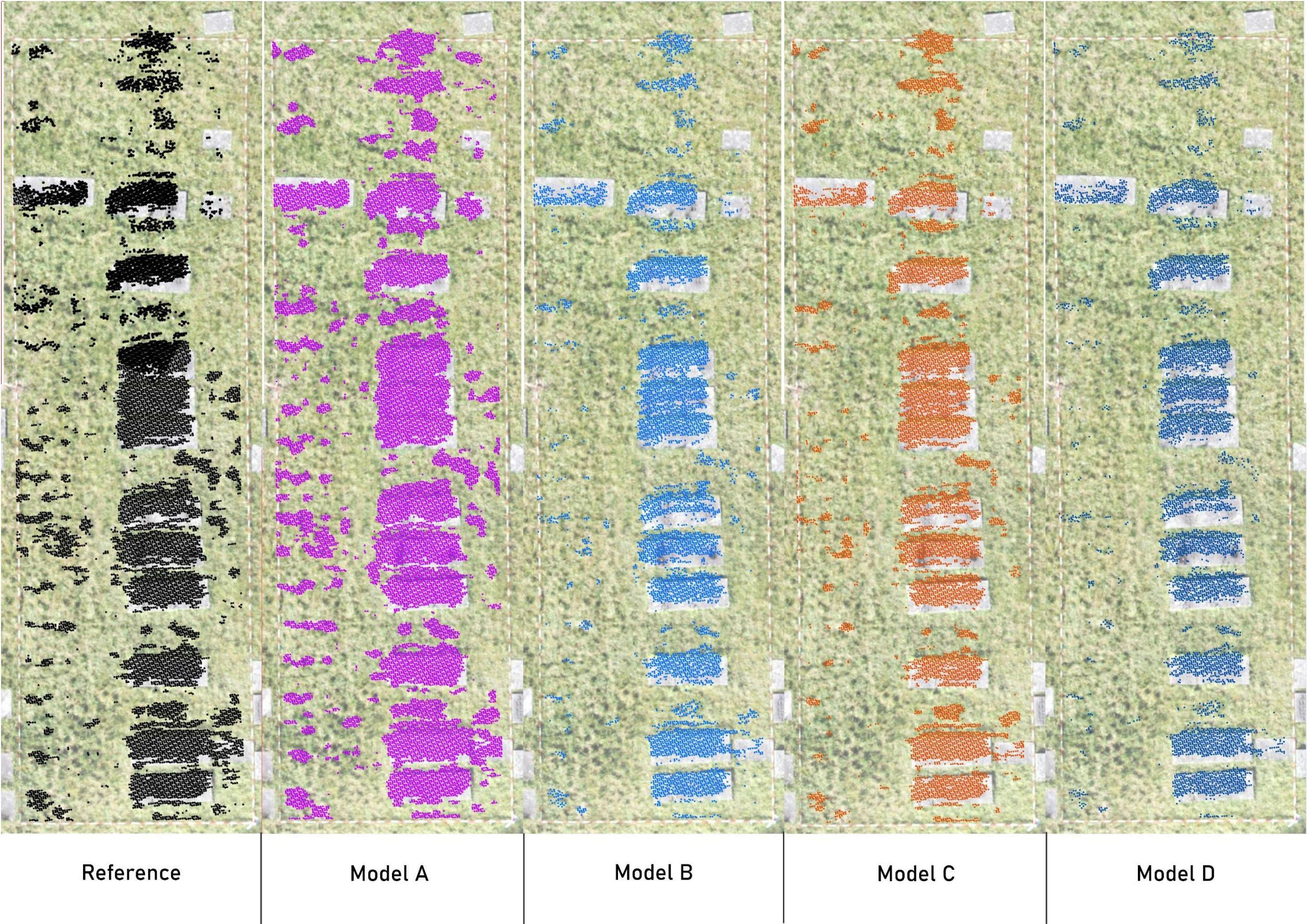
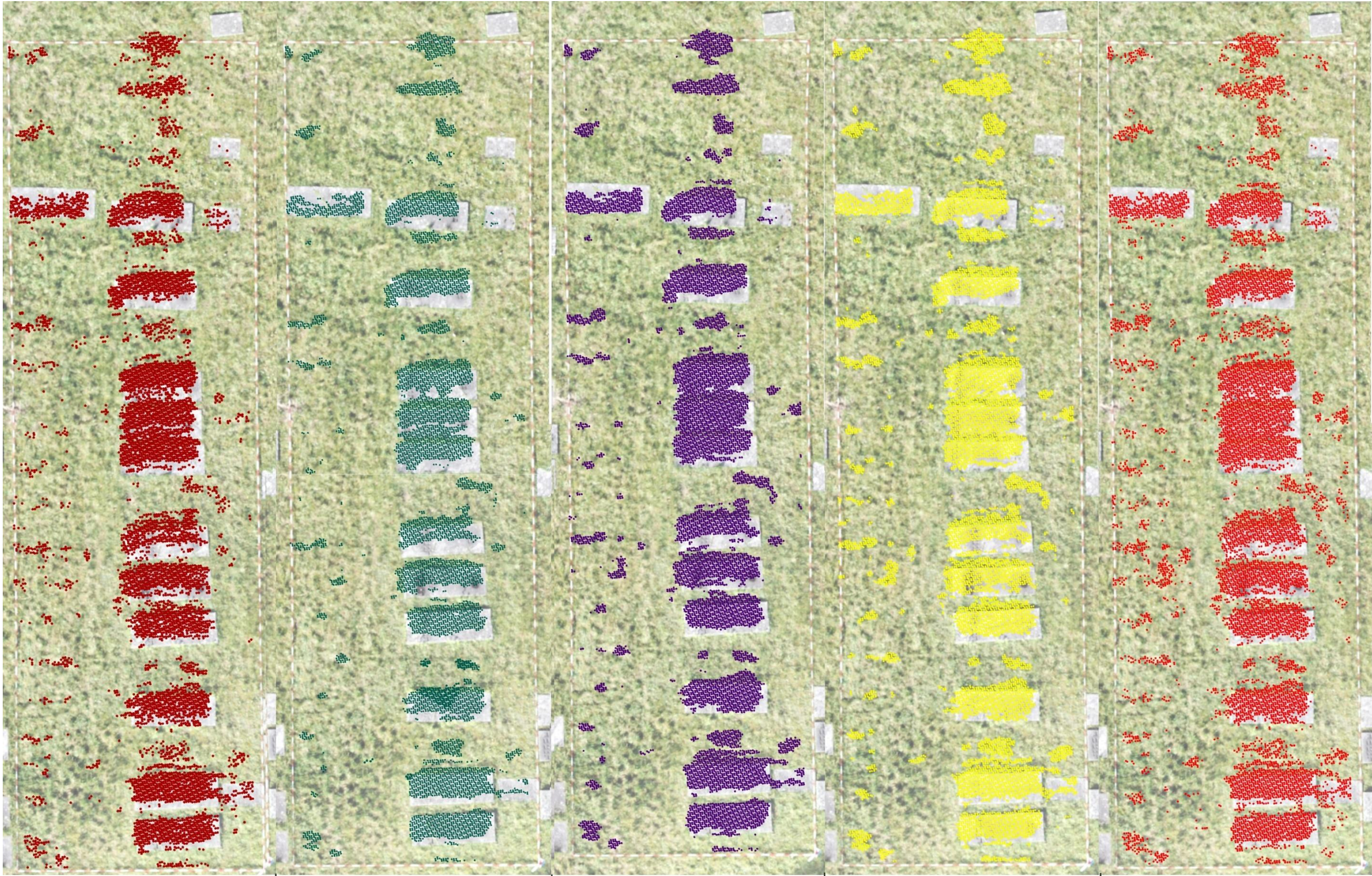


Figure 23: Visual evaluation of the sample area AB2 with the reference data (black) representing class #1 and the corresponding model predictions (coloured) (Image 1 of 2) (Author).



Model E

Model F

Model G

Model H

Model J

Figure 24: Visual evaluation of the sample area AB2 with the reference data (black) representing class #1 and the corresponding model predictions (coloured) (Image 2 of 2) (Author).

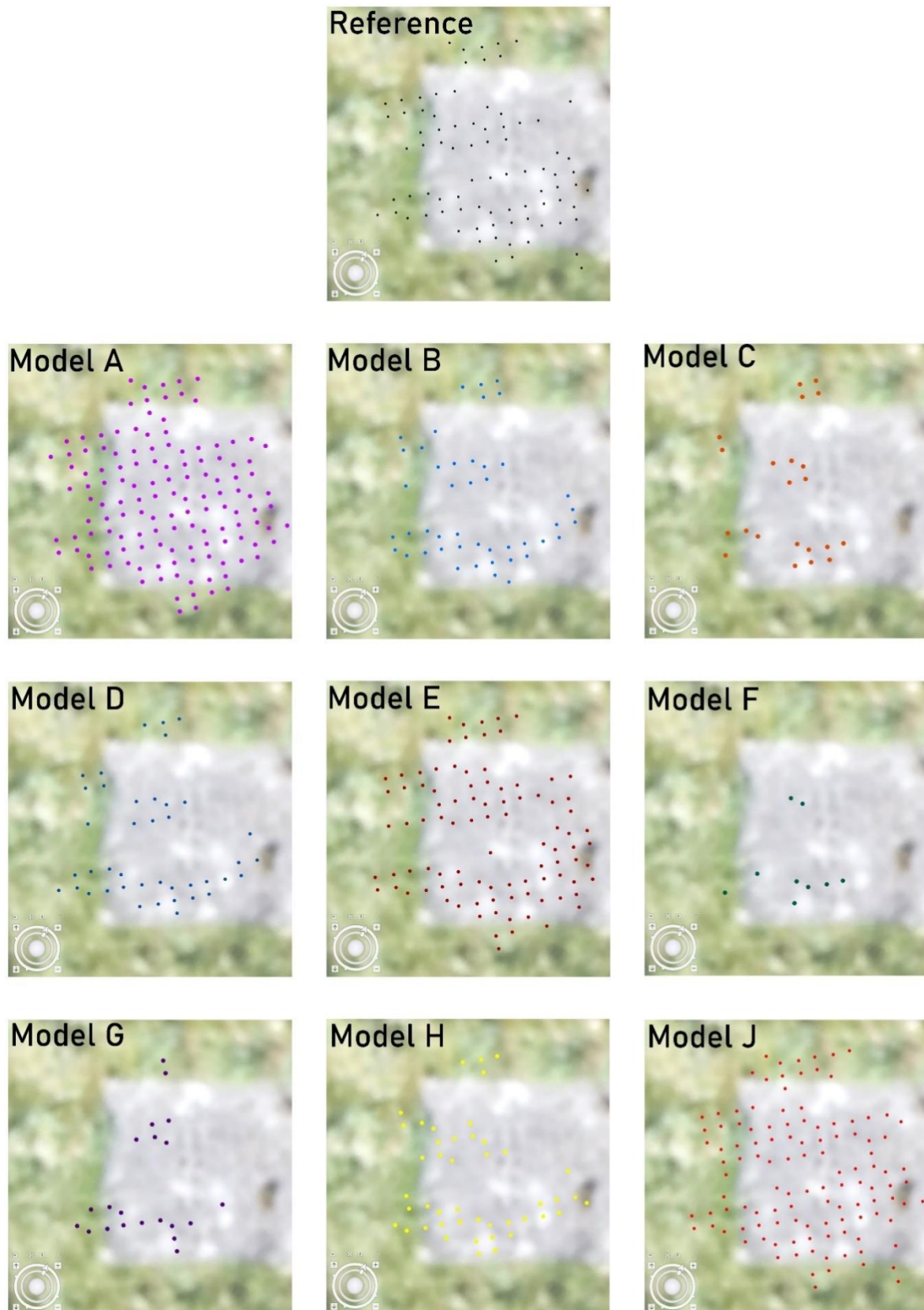


Figure 25: Orthogonal view on the visual evaluation patch $P1$ in sample area $AB2$ displaying the reference data set (black) and the model prediction (individual colour) of points class #1 representing the underneath located burial (coffin) (Author).

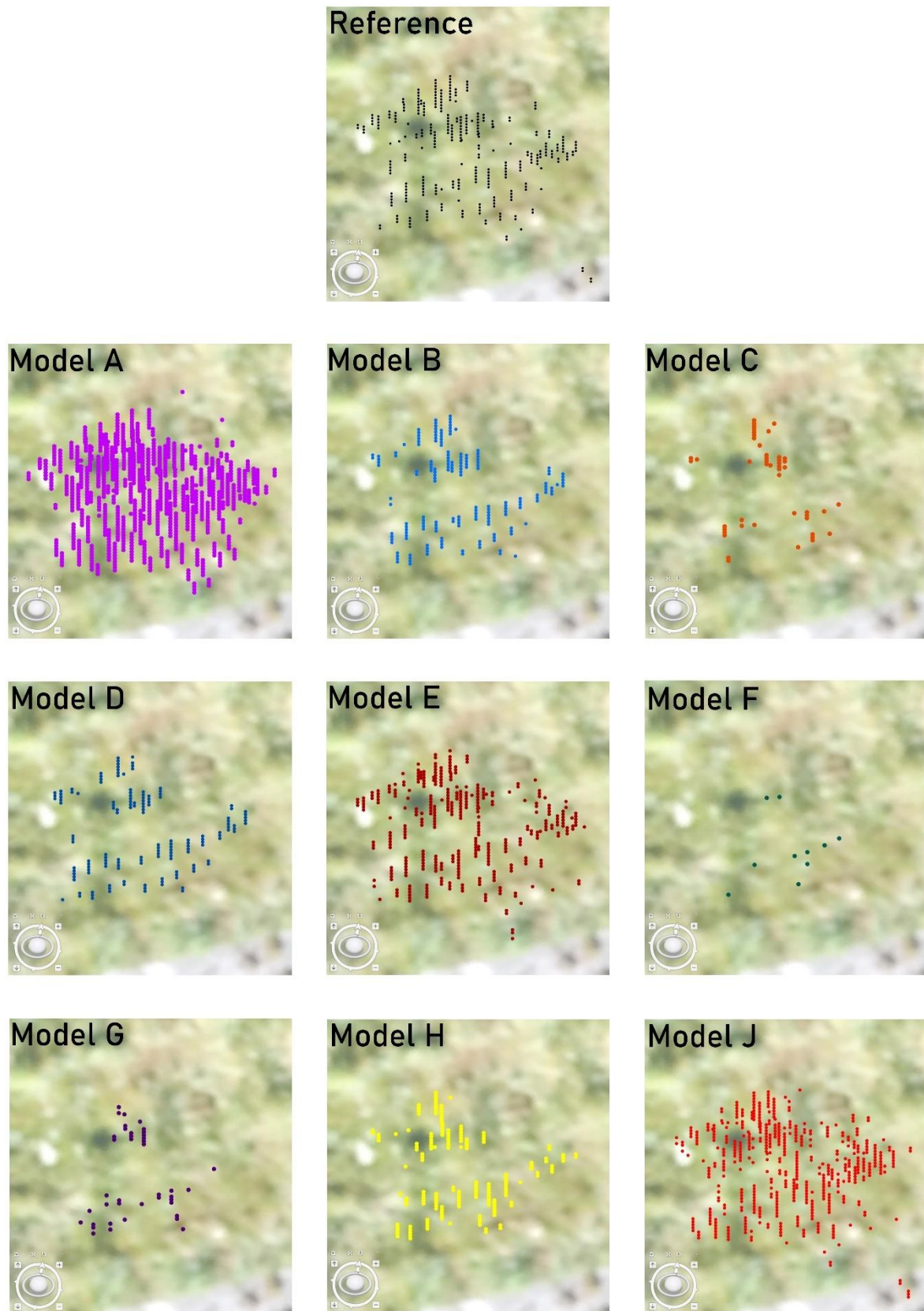


Figure 26: Oblique view on the visual evaluation patch *P1* in sample area *AB2* displaying the reference data set (black) and the model prediction (individual colour) of points class #1 representing the underneath located burial (coffin) (Author).

5.3 Quantitative evaluation of the classification models

For a number-based classification assessment, each model is evaluated via an independent data set. The applied GIS software package (*ArcGIS Pro* Version 2.9) enables the analysis of each model based on the classified sample area *AB2*. Such as, in the validation process during the training session, the parameters *Precision*, *Recall* and *F1-Score* provide valuable information. In addition, a confusion matrix quantifies the amount of correct and false predictions. For example, in Figure 21, the resulting confusion matrix of model **A** displays the distribution of TPs and FPs between the three classes *#0-Never Classified*, *#1-Unassigned*, and *#2-Ground*. While the first two classes share the predictions, the class *#2-Ground* does not get any detections. The cause is that the class represents the gravestones and is greatly underrepresented in the data blocks of *AB1* and *AB2*. The validation of training session **A** already revealed a disregard of the class. The same results apply to model **B**. For this reason, the class is ignored for all subsequent models.

The evaluation parameters and confusion matrices values are listed below in Table 7 and Table 8. The graphs of the confusion matrices from model **B** – **J** are in Annex L, Figure 40.

Evaluation Parameters					
Model	Class Code	Class Name	Precision	Recall	F1-Score
A	0	Never Classified	0.9507	0.9833	0.9667
A	1	Unclassified	0.7263	0.4646	0.5667
A	2	Ground	NA	NA	NA
B	0	Never Classified	0.9638	0.9900	0.9768
B	1	Unclassified	0.8539	0.6102	0.7118
B	2	Ground	NA	NA	NA
C	0	background	0.9571	1.0000	0.9781
C	1	Unclassified	0.9992	0.5302	0.6928
D	0	background	0.9463	0.9894	0.9674
D	1	Unclassified	0.7873	0.4110	0.5401
E	0	background	0.9550	0.9967	0.9754
E	1	Unclassified	0.9368	0.5078	0.6586
F	0	background	0.9431	0.9984	0.9700
F	1	Unclassified	0.9571	0.3675	0.5311
G	0	background	0.9509	0.9932	0.9716
G	1	Unclassified	0.8667	0.4615	0.6023
H	0	background	0.9558	0.9962	0.9756
H	1	Unclassified	0.9291	0.5169	0.6642
J	0	background	0.9715	0.9845	0.9780
J	1	Unclassified	0.8108	0.6973	0.7498

Table 7: Evaluation parameters for each class prediction for the nine models **A** – **J** (Author).

Confusion Matrices Evaluation Models A - J									
MODEL	TRUE	PREDICTED	COUNT		MODEL	TRUE	PREDICTED	COUNT	
A	0	0	1684753		E	0	0	1707752	
A	0	1	28598		E	0	1	5599	
A	0	2	0		E	1	0	80391	
A	1	0	87454		E	1	1	82938	
A	1	1	75875						
A	1	2	0		F	0	0	1710661	
A	2	0	0		F	0	1	2690	
A	2	1	0		F	1	0	103300	
A	2	2	0		F	1	1	60029	
B	0	0	1696296		G	0	0	1701760	
B	0	1	17055		G	0	1	11591	
B	0	2	0		G	1	0	87956	
B	1	0	63658		G	1	1	75373	
B	1	1	99671						
B	1	2	0		H	0	0	1706906	
B	2	0	0		H	0	1	6445	
B	2	1	0		H	1	0	78911	
B	2	2	0		H	1	1	84418	
C	0	0	1713279		J	0	0	1686769	
C	0	1	72		J	0	1	26582	
C	1	0	76725		J	1	0	49440	
C	1	1	86604		J	1	1	113889	
D	0	0	1695216						
D	0	1	18135						
D	1	0	96193						
D	1	1	67136						

Table 8: Confusion matrices based on the model evaluation on data set AB2 (Author).

Both the parameter values and the confusion matrices give insights into the various performances of the models on predicting class #1. While the *Precision* is almost always higher than 80%, the *Recall* varies between 37% and 70%. There are corroborating and contradicting outcomes in the validation parameter obtained by the model training sessions. While there is the confirmation that model **J** is classifying the burial locations well, model **H** is deriving lower *Recall* results than it was expected (66% validation vs 52% evaluation). The opposite applies to the two models **B** and **C**. The first results after the validation step were expecting a weak outcome for the followed evaluation part. However, the application on the reference set created above average numbers with a high *Precision* value in model **C**.

5.4 Topographical data

Section 4.2.3 describes the acquisition and the analysis of the photogrammetric data (336 aerial images). As a result of this process, the following two products got embedded in the GIS environment.

1. Topographic surface data as 3D point cloud format

For both sample areas (AB1 and AB2), a dense 3D-point cloud was created via *DIM*. Each site contains six tie points (individual markers) surveyed via GPS measurements. The resulting processing parameters and error deviations are in Table 9.

Area	Points	pts/m ³	RMS (X) cm	RMS (Y) cm	RMS (Z) cm
AB1	7817127	80505	1,9	1,7	1,1
AB2	8596573	60299	3,6	2,6	2,5

Table 9: Characteristics of the resulting point cloud from photogrammetric process areas AB1 and AB2 (Author).

2. Orthomosaic

As a 'side product' of the photogrammetric processing, orthographic imagery was obtained for the sample areas and the cemetery in total. As a result, the sample areas could achieve an average ground resolution of 0,4cm/pixel and 4,6cm/pixel for the site's overview (Annex B).

5.5 Historical aerial data

Section 4.2.4 elaborates the characteristics of the various historical data sets collected during this research. Even though the aerial images taken in 1945 are the oldest data found, they are suitable for being georeferenced in a GIS. Also, they provide the most valuable information due to their high spatial resolution. The cause for this is the low altitude during the image capturing. Figure 27 provides a line-up of the two available data from 1945 and the orthomosaic raster created from the survey dated 2021.

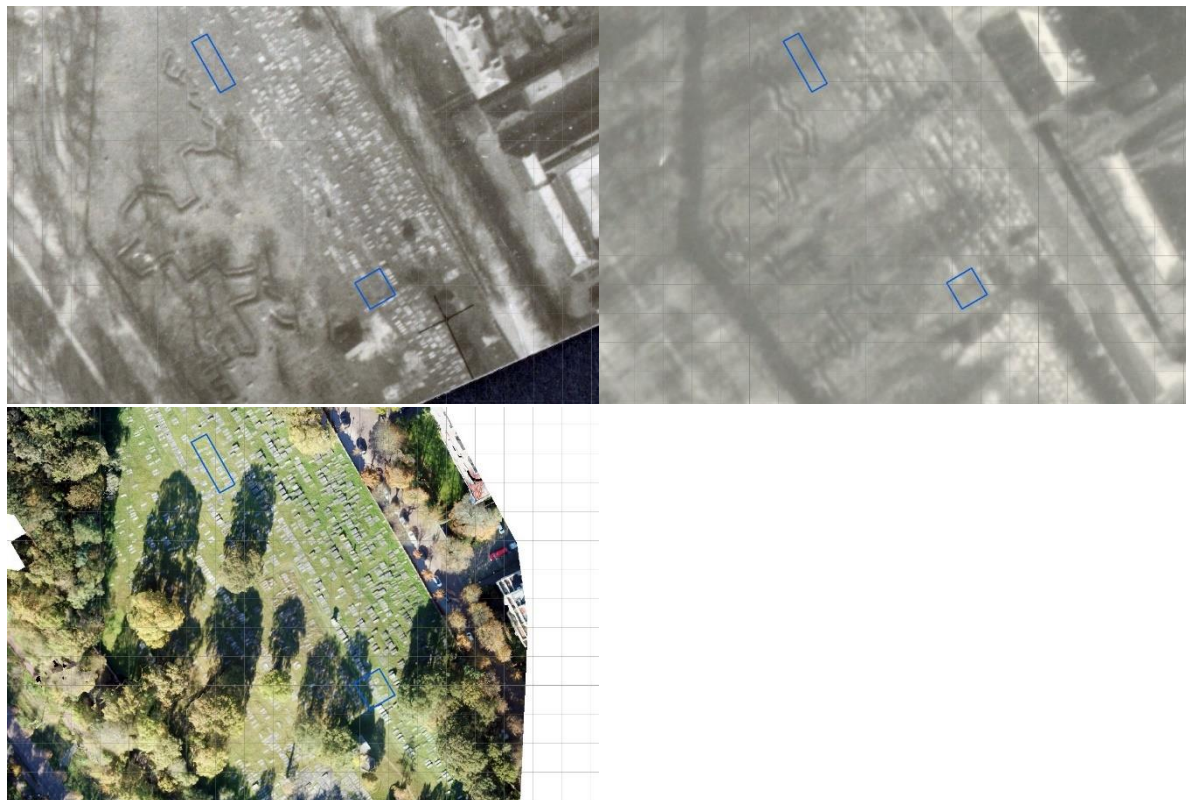


Figure 27: A cut-out from aerial images displaying the research area, including the two sample areas (blue frames). TOP LEFT – 18.03.1945 (DotKa, 2022); TOP RIGHT – 01.03.1945 (DotKa, 2022); and BOTTOM LEFT – 28.10.2021 (Author).

6 Discussion

This research project aims to extract information from GPR survey data to detect unmarked burial locations at the Jewish cemetery in The Hague. The data processing was held in a GIS environment where the possibilities of automated data classification via a *CNN* got explored. The outcomes got combined with other information to assist the interpretation and detection.

This section first refers to the prior set research objectives and discusses the implementation of the used methodology and the derived results. Subsequently, the usability will be reviewed by highlighting the differences between the research approach and the existing established workflow. Lastly, the caveats will be addressed that were encountered during this project.

6.1 Implementation of the methodology

The established workflow of this project enables the remote detection of burial locations that are not identifiable on the surface area. Moreover, the *CNN* based classification of the GPR data sets in within a GIS is the first attempt to utilize such a technique on data sets of this make. The evaluation of the individual results showed that multiple variables are influencing the outcome. However, the produced models and their predictions demonstrate the potential of processing large numbers (areas) of point data as a part of an automated workflow. This standard workflow, starting from the data acquisition until the analysis, still needs to be refined to overcome a series of regular processing (e.g., filtering, transformation, etc.) steps and software specific issues.

One crucial fact to consider is that for the *PointCNN* training data, both types of burial locations (marked and unmarked) have been considered for classification. It is expected that a partition of those two groups would result in a low number of data representing the unmarked burials and so not be suitable for the training purpose. An example of this effect is presented by the failed detection of the point data representing the gravestones.

The GIS provides the necessary 3D environment to thoroughly inspect the classified point clouds (Figure 23 to Figure 26). In addition, the GIS forms a platform that enables the compilation of multiple data sets of various formats. For example, several raster data sets like aerial imagery and layout plans have been superimposed with the classification results (Annex M, Figure 41).

6.2 Usability of methodology

The system promoted in this research project is an attempt to employ the advantages of the deep learning technology together with the capabilities of a Geographic Information System to analyse data obtained by a GPR survey.

Due to their unique format, the GPR data will always have to undergo post-processing steps after the collection in a specialised software package. For interpretation purposes, the tool used in this project, *EKKO_Project*, allows the analysis of vertical and horizontal data slices (Figure 22, Figure 28).

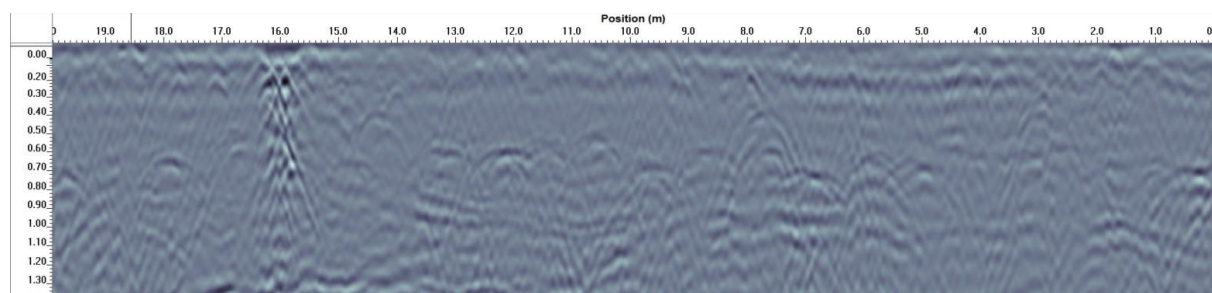


Figure 28: Vertical Line (B-scan) view in *EKKO_Project* software tool on scan line #19 of area AB2. The unmarked burial locations are the hyperbolas reaching their angular point in a depth range 0.4 - 0.8m below the surface. At the distance mark of 16m, the signal disturbances are notably caused by the only burial marked with a gravestone (Author).

While the above-displayed example is relatively simple to analyse by selecting the hyperbolas representing the reflection of the coffins, it requires expert knowledge to interpret B-scans where multiple signals interfere with each other (Annex C). Combining the B- and C-Scans enables only experienced users to analyse the GPR data sufficiently. For detecting unmarked burial locations, a trained examiner is necessary.

Section 3.3 describes attempts to use deep learning frameworks to facilitate the manual selection process in the different scan layers. Most research projects focus on recognising infrastructure features in urban areas by applying a *CNN* model on the GPR data in a raster format. Promising results were achieved by Kim et al. on objects with regular feature properties (e.g. maintenance holes) (Kim *et al.*, 2021).

Previous studies looked at the segmentation of 3D-point clouds via *CNN* frameworks. However, all those studies based on LIDAR point clouds as input data, which are different from GPR data. On the contrary, the employment of *PointCNN* on a point cloud data set describing a volumetric sub-surface area hasn't been part of any study yet.

The results of this research show that a suitable classification model can correctly extract the buried objects of interest. Based on the variables influencing the classification outcome, the argument can be made that a model's effectiveness is strongly locally dependent on the training area and its features. Consequently, it is not appropriate for application in other locations and therefore of limited usability. On the other hand, a well-defined model can process extensive areas after being trained on one or two sample patches. In the research area of the *Jewish cemetery in The Hague*, it would be helpful to apply a model on a data set of the entire site as a primary filter of burial locations. Furthermore, it is suggested to explore the possibilities to separate this intermediate product once more into a group of marked and unmarked burial locations. An additional *CNN* model could be used for the detection and classification of point clusters, or a clustering algorithm could serve this task.

Regardless of the potential of automated point cloud classification, the evaluation of the model results confirmed that human validation is still necessary to verify if a prediction is mirroring the truth. Concerning the useability of the used method, it is required to establish a standard workflow as it is recommended in the prior section. Also, some technical issues will have to be resolved in this process.

Without a doubt, the advantage of combining the GPR data with other data sets and the visualisation options of the GIS platform is evident. Moreover, this domain also allows less experienced users to read parts of the GPR data and make conclusions with the help of the supporting data sets.

6.3 Research Limitations

Various types of problems emerged during this research project. While most of the issues could be resolved, some of the obstacles were unviable to overcome and led to compromises and amendments of the initial project's workflow.

One major problem was the blocked export functionality of the GPR data as a standard 3D point cloud format from the *EKKO_Project* software for further processing in *ArcGIS Pro*. The option to export the data was not given due to missing GPS data. Such data need to be assigned to the single GPR traces. An RTK-GPS antenna attached to the GPR sensor system usually collects the position and elevation information simultaneously during the survey. The individual character and the system's particular GPS data format made them impossible to reproduce for an assignment later. An alternative approach was the manual orientation of the GPR survey grids to the national system (RD New). In addition, a topographic correction was applied for each grid line based on the corresponding topographical profile, which was gained from the photogrammetric DSM. But even with the complete post-processed, spatial

transformed (3 parameters) and topographical corrected data existing in the *EKKO_Project* software environment, it was impossible to transfer those data into another system. A final attempt was made by vertically adjusting the single top traces in a large CSV- file. The software manufacturer extracted the file on request. Unfortunately, the manual adjustment did not produce a satisfactory result (Figure 29), and this method was also neglected.

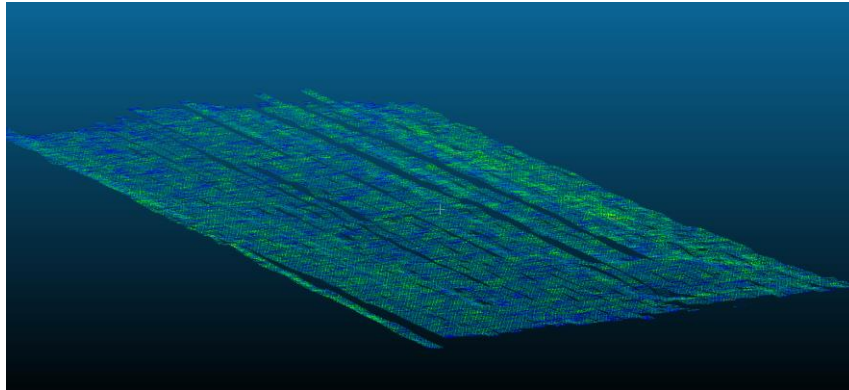


Figure 29: Result of manual vertical adjustment of single top traces, displaying the first layer of points representing the surface top with multiple misalignments visible between the scan lines (Author).

This issue resulted in using a data set represented in a local coordinate system without topographical information.

Before this data set could be imported into ArcGIS Pro, several processing steps were taken in Cloud Compare and EditPad Light V8 software packages.

Two potential solutions are suggested to avoid this rather cumbersome post-processing workflow for transformation steps. First, the simultaneous GPS data recording at the GPR survey would resolve most issues. Unfortunately, such equipment was not available at the time of the data collection. An alternative can be the development of an integrated data workflow via *Feature Manipulation Engine* (FME) to automatize the transformation process into a standard data format. This option is also favourable as the engine is designed to manage large spatial data.

As already mentioned, high demand for computational power was experienced. In addition, the manual labelling of the training point cloud claimed a large part of the graphical capacities for displaying and selecting subsets of the 3D-point cloud in the profile view. Consequently, the already labour-intensive progress was delayed in execution by the extensive loading times.

The training procedure for each model was so computationally intensive that a migration of the project and the assigned files to a computer with higher performance was necessary. Still, the average duration for training and validation was circa 14 hours for 8 epochs. Therefore, to create and refine the models via *PointCNN*, it is recommended to use an arrangement of multiple graphic cards with high specifications. The same suggestion applies to the subsequent classification process.

Both above-stated issues stressed the project's time limitation, so the development of solutions had to be discarded to fulfil the primary goal of the research. However, such development will be part of further research.

One fundamental issue of the applied method was the fact that no verified reference data were available to train and validate the classification models. While it is relatively simple to create such a reference on surface data where the objects features are distinctive, it is more complex to do this on sub-surface data. The manual labelling of the data set from area *AB1* was challenging as influenced by subjective selection processes that can't get reproduced. Consequently, there is an uncertainty assigned to the training data and to the resulting models.

7 Conclusion

7.1 Research questions

In section 2.2 are four research sub-questions formulated that will be individually answered in this chapter before the gained insights will be used to respond to the central question of this thesis.

SQ1: What GPR data representation format is (3D point cloud, raster, voxel) is suitable for an interpretation in a GIS?

This question was made before the GPR data collection for this project and before integrating a GIS. The focus of this research was the process and assessment of GPR data in the data format of a 3D-point cloud. The volumetric representation by single points was created by interpolating the single GPR scan lines for each surveyed grid (sample area).

This format has a significant advantage in the volumetric three-dimensional representation of the sub-surface data. In particular, it is possible to segment the parts of interest and visualize them due to the achieved classification. Furthermore, a sufficient resolution (density) of the point cloud enables the display of characteristic object features. Those characteristics are beneficial for the model training process. However, the feature properties of burial objects (coffins) in this project were not of regular three-dimensional shape and so not potentially useful as a classification criterion.

The benefit of the high-resolution 3-dimensional environment comes with a cost of computational power and data accessibility. Throughout the project, the experience was made that the display of the point clouds in their total density prevents seamless navigation and decreases the overall experience of the visual analysis. Therefore, all 3D data sets were transferred from an external server location to the local data storage of the processing unit.

Even it has not been tested in this project, former research demonstrated that the integration of GPR data in a raster format is feasible and does provide valuable information when combined with supplementing data sets (e.g., vector data) (Cazzaniga *et al.*, 2013; Tabarro *et al.*, 2017).

Not investigated was exploring the GPR data in a voxel format. The main reason is that a transfer from point data into voxel data would lead to a decrease of the level of detail caused by the size dependent resolution of the voxels.

SQ2: Can the generated deep learning results (e.g., pattern/value detection and extraction) enhance the interpretation of GPR subsurface data to detect unmarked burials?

As a general answer, it does facilitate the interpretation process, but does not replace the necessary human interaction involvement. The assessed *CNN* models in this project derived results considering the point's spatial position and intensity value. The visual evaluation mirrored the patterns visible in the B- and C-scans of the GPR data. By applying the model as a 'first filter', it focuses on data that are considered objects of interest. As mentioned in 6.2, the recommendation is to refine the selection by additional filtering or classification methods.

A final answer to this question is that the deep learning approach does enhance the interpretation process in terms of quantity and, to a certain extent, quality. Large data sets can be automatic processed and relevant data extracted. The extraction does represent the ground truth to a specific deviation depending on the model's effectiveness. Considering this, a model verification by an expert in the field of GPR is necessary.

SQ3: How can topographical survey data be beneficial for the interpretation process of subsurface data?

Considering that no topographic correction could be applied in the GPR-data processing (see 6.3), this question will not be answered in the intended completeness. Due to missing survey data and an unsuccessful attempt to introduce those missing information from a secondary data source, it was impossible to adapt the GPR data block to the topographical properties of the surveyed terrain. This circumstance created a missed opportunity to join the GPR with the UAV point clouds. It is anticipated that this would enrich the data by adding more attributes like elevation and colour values, which could boost the options for the deep learning method. In addition, the surface data could help to distinct the classification of marked and unmarked burial locations in a unified data set.

Nevertheless, the option of detecting the gravestones on the surface data sourced from the photogrammetric method is there. An individual process of those data followed by the overlay with the results of the sub-surface data assessment would advance the detection of the non-marked burial locations. The data for this individual process could either be the 3D-point cloud or the raster data (orthomosaic) from the surface. The development of this process is a recommendation for further studies.

The topographic data gathered and processed in this project were integrated into the GIS and complemented the data collection. A separate classification to foster the interpretation progress can be part of continuing research in this field.

SQ4: To what extent do historical data sets of the research area support data interpretation of GPR data to detect unmarked graves?

Multiple historical data sets were collected for the research area. Different formats like raster or tables provide information of the cemetery covering up to 250 years.

For the integration into the GIS, eight historical aerial images (1945 – 1983) were overlayed with the GPR and UAV data. The followed assessment showed that the images captured by Dutch authorities in their frequent observation flights are less valuable due to their lower spatial resolution. That is a result of the altitude at which the images were captured. Remarkable is the resolution of the two oldest images that the British RAF sources from 1945. Caused by the lower flight attitude, details of the site are high enough to recognize visible single gravestones and disturbing activities that occurred during wartime. Both insights are beneficial when examining survey data of the entire cemetery together with the archived register lists from the Foundation of the Jewish Cemetery in The Hague. This list of buried persons and the corresponding layout plan can be linked with the survey data sets and enable the mapping of persons to a spatial location in the cemetery. The fact that there was no systematic order (e.g., chronological) the burials were placed is preventing an assignment of an unmarked burial location to a person in the register.

For this research, only the foundation's general layout plan (Annex M, Figure 41) was embedded in the GIS to corroborate the survey data processing results. The integration of the list containing the register of the cemetery's occupation and their linked detail plans, including the gravestone numbering (Annex E), has not been done because of their lack of accessibility in digital format.

However, it can be concluded that considering all historical information and the survey results enhances detecting the unmarked burial locations.

Leaving to answer the central question of this research project:

“To what level is the classification via deep learning and visual interpretation of Ground Penetrating Radar data possible in a GIS?”

The Ground Penetrating Radar sub-surface data were successfully integrated into a GIS. The data, represented in a 3D-point cloud data format, were utilized to create, and apply multiple classification models based on a Convolutional Neural Network framework. Based on the results, it can be concluded that deep learning models can achieve a classification on volumetric point cloud data. However, the model's efficiency and usability depend on the input training data quality and quantity. Furthermore, a quantitative evaluation confirmed that a fully automated classification process is not feasible yet, and therefore human interaction is still necessary to assess the predictions. However, the potential of the automated classification of sub-surface point cloud data is evident (e.g., pattern detection) and requires more research for further development.

The visual interpretation of the GPR data in the form of a 3D-point cloud in a GIS environment is certainly simplified compared with a conventional software package displaying the data in B-and C-scans. However, multiple perspectives and advanced 3D navigation possibilities facilitate data exploration and object recognition. Moreover, it is the functionality of combining the classified point clouds with other spatial and non-spatial data that upgrades the level of information for an efficient interpretation process.

7.2 Conclusion

This research project investigated the possibility of detecting unmarked burial locations based on ground penetrating radar data that are integrated into a geographic information system and processed via a deep learning framework. Additional data sets of the research area complement the results to enhance the exploitation. The research area was the Jewish Cemetery in The Hague, where marked and unmarked gravesites are present. Two representative sample areas were selected and surveyed by multiple sensors, including GPR and aerial imaging.

The GPR survey data had to be processed in a sensor-oriented software tool before integrating a GIS as a 3D-point cloud. Classifying those subsurface data by a deep learning method was the primary objective. That goal was accomplished with the *PointCNN* module, which functionality is based on a convolutional neural network algorithm. Several classification models have been trained, applied, and evaluated. The evaluation was done in a visual and in a quantifying approach. Both methods confirmed that patterns of point clusters are recognizable. Based on the cemetery's arrangement and layout plan, those clusters could be assigned to potential burial locations that are not indicated on the surface.

The study also revealed that the classifying models' effectiveness depends on the input data and the settings of their training process. An improvement of the model's performance could be achieved by refining the training parameter values (block size, epochs, weighted classes, etc.). A manual labelling process via cross-sections created the reference data for the training and validation purpose. This selection process will need to be critically reviewed as it influences the model's outcome and, therefore, the overall methodology's usability.

As a reflection, the 3D-point cloud integration in a GIS and the automated classification via *PointCNN* could be successfully applied. But at that stage, no significant benefit was noticeable when comparing the applied method with the conventional approach by interpreting B- and C-scans of the GPR data. A change occurs when the advantages of the GIS are unfolded. For example, supplementing data sets from secondary sources and the three-dimensional visualisation of the data provide additional insights that enable user interpretation even without expert knowledge.

Historical aerial imagery was also introduced and overlayed with the obtained GPR data sets. Even when not all images appeared to be of additional value, there were some photographs documenting major impacts in the graveyard's infrastructure that is not visible in the present status of the site.

Altogether, the main conclusion can be drawn that this research project confirmed that GPR-data in the form of a 3D-point cloud could be integrated into a GIS and classified by a deep learning framework. The automated classification supports the detection of underground patterns indicating the location of unmarked burials. However, the promoted method is too complex for the application in smaller areas and does not gain advantages compared to the conventional analysis of GPR data. But the created models help to classify large amounts of point cloud data and facilitate the detection process. For large scale areas, like the research area, it will achieve a major benefit for the progress in processing time and interpretation results.

The GIS establishes a platform to integrate different data of the cemetery and so complies with various information in a multi-perspective environment that supports the detection of the unmarked graves. Finally, the options for conveying the achieved information via a GIS need to be remarked. The collected data and gained knowledge can be presented and distributed to all stakeholders in a clear and interactive format.

7.3 Recommendations for future research

The benefits mentioned earlier of processing and utilizing GPR data in a GIS make it worth investing more research on this topic.

A first approach should be the seamless progress of data integration into the GIS starting from the survey with coupled sensors over to an automated post-processing pipeline and standardized transfer processes.

The deep learning method can be further explored by 1.) creating and refining classification models based on high-quality reference data; 2.) creating a repository of various classification models and assessing their efficiency on multiple scenarios, and 3.) creating an arrangement of *CNN* models and other tools (e.g., cluster data algorithm) for process automatization. These components would enable developing a system that ties on the prior mentioned integration part. That would build a complete workflow.

Future studies should also look at the options to embed and process different data sets into the GIS. For example, existing vector data (e.g., infrastructure records) or other databases with a non-spatial character can complement the current information. Additional information can also be acquired by adding and assessing the 2D GPR data (B- and C-scans) in the same environment. Another CNN framework could be applied and inserted in the workflow for this raster assessment.

8 References

- ASPRS, A. S. for P. & R. (2002) *LAS Specification 1.4-R15 Release Information*.
- Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) ‘Segnet: A deep convolutional encoder-decoder architecture for image segmentation’, *IEEE transactions on pattern analysis and machine intelligence*. IEEE, 39(12), pp. 2481–2495.
- Bengio, Y., Goodfellow, I. and Courville, A. (2017) *Deep learning*. MIT press Massachusetts, USA:
- Bevan, B. W. (1989) ‘Search for graves’, *1989 SEG Annual Meeting*, (June), pp. 280–284. doi: 10.1190/1.1889567.
- Biswal, A. (2020) ‘Convolutional Neural Network Tutorial’. Available at: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/convolutional-neural-network> (Accessed: 24 November 2021).
- Brownlee (2018) *Difference Between a Batch and an Epoch in a Neural Network, Machine Learning Mastery*. Available at: <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/> (Accessed: 31 January 2022).
- Brownlee (2020) ‘How to Control the Stability of Training Neural Networks With the Batch Size’, *Machine Learning Mastery*, pp. 1–27. Available at: <https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/> (Accessed: 31 January 2022).
- Cazzaniga, N. E. *et al.* (2013) ‘a Shared Database of Underground Utility Lines for 3D Mapping and Gis Applications’, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-4/W1, pp. 105–108. doi: 10.5194/isprsarchives-xl-4-w1-105-2013.
- Chang, C.-C. and Lin, C.-J. (2011) ‘LIBSVM: a library for support vector machines’, *ACM transactions on intelligent systems and technology (TIST)*. Acm New York, NY, USA, 2(3), pp. 1–27.
- Chen, Z. *et al.* (2020) *SCIPoC : Semantic Classification of Indoor Point Cloud*.
- Dasgupta, T. *et al.* (2018) ‘Augmenting textual qualitative features in deep convolution recurrent neural network for automatic essay scoring’, in *Proceedings of the 5th Workshop on Natural Language Processing Techniques for Educational Applications*, pp. 93–102.
- Doolittle, J. A. and Bellantoni, N. F. (2010) ‘The search for graves with ground-penetrating radar in Connecticut’, *Journal of Archaeological Science*. Elsevier Ltd, 37(5), pp. 941–949. doi: 10.1016/j.jas.2009.11.027.
- DotKa (2022) *Business Home - Dotka Data luchtfoto’s en kaarten*. Available at: <https://www.dotkadata.com/en/> (Accessed: 20 February 2022).
- Draelos, R. (2019) ‘Measuring Performance: The Confusion Matrix’, *Glass Box*, (1), pp. 2019–2021. Available at: <https://glassboxmedicine.com/2019/02/17/measuring-performance-the-confusion-matrix/> (Accessed: 9 February 2022).
- Enthoven, drs. Francine Puttmann, drs. P. H. *et al.* (1992) *De joodse begraafplaats aan de Scheveningseweg in Den Haag*. Den Haag: Gemeente Den Haag, Dienst Ruimtelijke en Economische Ontwikkeling, afdeling Monumentenzorg.
- Foundation Jewish Cemetery, T. H. (2021) *website foundation*. Available at: <https://joodsebegraafplaats.nl/geschiedenis/> (Accessed: 5 October 2021).
- Frigui, H., Ho, K. C. and Gader, P. (2005) ‘Real-time landmine detection with ground-penetrating radar using discriminative and adaptive hidden Markov models’, *EURASIP Journal on Advances in Signal Processing*. SpringerOpen, 2005(12), pp. 1–19.

- Gader, P., Lee, W.-H. and Wilson, J. N. (2004) 'Detecting landmines with ground-penetrating radar using feature-based rules, order statistics, and adaptive whitening', *IEEE Transactions on Geoscience and Remote Sensing*, 42(11), pp. 2522–2534. doi: 10.1109/TGRS.2004.837333.
- Gemeentearchief (no date) *haagsgemeentearchief*.
- Goodman, D. and Piro, S. (2013) *GPR remote sensing in archaeology, GPR Remote Sensing in Archaeology*. doi: 10.1007/978-3-642-31857-3.
- Grasmueck, M. and Green, A. G. (1996) '3-D georadar mapping: Looking into the subsurface', *Environmental & Engineering Geoscience*. EEGS, 2(2), pp. 195–200.
- Grilli, E., Özdemir, E. and Remondino, F. (2019) 'APPLICATION OF MACHINE AND DEEP LEARNING STRATEGIES FOR THE CLASSIFICATION OF HERITAGE POINT CLOUDS.', *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.
- Group, T. H. (2015) *Hierarchical Data Formats - What is HDF5? / NSF NEON / Open Data to Understand our Ecosystems*. Available at: <https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5> (Accessed: 29 January 2022).
- Hu, X. and Yuan, Y. (2016) 'Deep-learning-based classification for DTM extraction from ALS point cloud', *Remote sensing*. Multidisciplinary Digital Publishing Institute, 8(9), p. 730.
- Ke, P. *et al.* (2018) 'A novel face recognition algorithm based on the combination of LBP and CNN', in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 539–543. doi: 10.1109/ICSP.2018.8652477.
- Kim, N. *et al.* (2021) 'A novel 3D GPR image arrangement for deep learning-based underground object classification', *International Journal of Pavement Engineering*. Taylor & Francis, 22(6), pp. 740–751. doi: 10.1080/10298436.2019.1645846.
- King, J. A., Bevan, B. W. and Hurry, R. J. (1993) 'The reliability of geophysical surveys at historic-period cemeteries: An example from the plains cemetery, mechanicsville, Maryland', *Historical Archaeology*, 27(3), pp. 4–16. doi: 10.1007/bf03373565.
- Korstanje, J. (2021) *The F1 score / Towards Data Science*. Available at: <https://towardsdatascience.com/the-f1-score-bec2bbc38aa6> (Accessed: 31 January 2022).
- Kovalenko, V., Yarovoy, A. G. and Ligthart, L. P. (2007) 'A novel clutter suppression algorithm for landmine detection with GPR', *IEEE Transactions on Geoscience and Remote Sensing*. IEEE, 45(11), pp. 3740–3751.
- Li, J., Chen, B. M. and Lee, G. H. (2018) 'So-net: Self-organizing network for point cloud analysis', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9397–9406.
- Li, Y. *et al.* (2018) 'Pointcnn: Convolution on x-transformed points', *Advances in neural information processing systems*, 31, pp. 820–830.
- Malinverni, E. S. *et al.* (2019) 'DEEP LEARNING for SEMANTIC SEGMENTATION of 3D POINT CLOUD', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2/W15), pp. 735–742. doi: 10.5194/isprs-archives-XLII-2-W15-735-2019.
- Menabrea, L. F. and Lovelace, A. (1842) 'Sketch of the analytical engine invented by Charles Babbage'.
- Merkle, D., Frey, C. and Reiterer, A. (2021) 'Fusion of ground penetrating radar and laser scanning for infrastructure mapping', *Journal of Applied Geodesy*, 15(1), pp. 31–45. doi: 10.1515/jag-2020-0004.
- Mohamed, A.-M. S. *et al.* (2020) 'Mapping the archaeological relics of catacombs at Northeast Saqqara using GPR data, Egypt', *NRIAG Journal of Astronomy and Geophysics*. Taylor & Francis, 9(1), pp. 362–374. doi: 10.1080/20909977.2020.1746891.

- Ng, M. and Choy, C. P. (2018) *Land Framework of Singapore: Building a Sound Land Administration and Management System*.
- Nos.nl (2018) *Een nieuwe 'ondergrondse stad voor de doden' in Jeruzalem*. Available at: <https://nos.nl/artikel/2211691-een-nieuwe-ondergrondse-stad-voor-de-doden-in-jeruzalem>.
- Ozkaya, U. *et al.* (2020) 'GPR B scan image analysis with deep learning methods', *Measurement*. Elsevier Ltd, 165, p. 107770. doi: 10.1016/j.measurement.2020.107770.
- Pasolli, E. *et al.* (2009) 'Automatic Analysis of GPR Images : A Pattern-Recognition Approach', 47(7), pp. 2206–2217.
- Patterson, J. and Gibson, A. (2017) *Deep learning: A practitioner's approach*. ' O'Reilly Media, Inc.'
- Poscetti, V., Zotti, G. and Neubauer, W. (2015) 'Archaeologia Polona vol. 53 (2015), Valeria Poscetti; Georg Zotti; Wolfgang Neubauer', 53, pp. 603–607.
- Qi, C. R., Su, H., *et al.* (2017) 'Pointnet: Deep learning on point sets for 3d classification and segmentation', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660.
- Qi, C. R., Yi, L., *et al.* (2017) 'Pointnet++: Deep hierarchical feature learning on point sets in a metric space', *arXiv preprint arXiv:1706.02413*.
- Radiuk, P. M. (2017) 'Impact of training set batch size on the performance of convolutional neural networks for diverse datasets'. *Information Technology and Management Science*.
- Remondino, F. and Golkar, A. (2019) 'Aerial Point Cloud Classification', XLII(October), pp. 12–14.
- Schmidhuber, J. (2015) 'Deep learning in neural networks: An overview', *Neural networks*. Elsevier, 61, pp. 85–117.
- Sorokina (2017) *Image Classification with Convolutional Neural Networks* / by Ksenia Sorokina / Medium. Available at: <https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8> (Accessed: 24 November 2021).
- Tabarro, P. G. *et al.* (2017) 'A WebGIS to support GPR 3D data acquisition: A first step for the integration of underground utility networks in 3D city models', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W7), pp. 43–48. doi: 10.5194/isprs-archives-XLII-4-W7-43-2017.
- Takase, T. (2021) 'Dynamic batch size tuning based on stopping criterion for neural network training', *Neurocomputing*. Elsevier, 429, pp. 1–11.
- Thakur, S. P. *et al.* (2019) 'Skull-Stripping of Glioblastoma MRI Scans Using 3D Deep Learning', in *International MICCAI Brainlesion Workshop*. Springer, pp. 57–68.
- Tischler, M., Collins, M. E. and Grunwald, S. (2002) 'Integration of ground-penetrating radar data, global positioning systems, and geographic information systems to create three-dimensional soil models', in *Ninth International Conference on Ground Penetrating Radar*. International Society for Optics and Photonics, pp. 313–316.
- Unumandakh, N., Amarsaikhan, D. and Sato, M. (2007) 'Integrated Study of GPR , ER and GIS'.
- Viswambharan, V. and Singh, R. (2020) *Introducing pre-trained geospatial deep learning models*, *ArcGIS Blog*. Available at: <https://www.esri.com/arcgis-blog/products/arcgis/imagery/introducing-ready-to-use-deep-learning-models/> (Accessed: 22 February 2022).
- Winiwarter, L. *et al.* (2019) 'Classification of ALS Point Clouds Using End-to-End Deep Learning', *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 87(3), pp. 75–90. doi:

10.1007/s41064-019-00073-0.

Wolf, J. *et al.* (2018) ‘Combined visual exploration of 2D ground radar and 3D point cloud data for road environments’, *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4/W10), pp. 231–236. doi: 10.5194/isprs-archives-XLII-4-W10-231-2018.

Zaheer, M. *et al.* (2017) ‘Deep sets’, *arXiv preprint arXiv:1703.06114*.

Zhang, L. *et al.* (2018) ‘Large-scale urban point cloud labeling and reconstruction’, *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier, 138, pp. 86–100.

Annex A



1 Het duingebied tussen Scheveningen en Den Haag.
Aan de onderzijde is naast het tolhek duidelijk de joodse
begraafplaats herkenbaar.
Plattegrond door L. Schweickhardt, 1832.



VOM REEKS 1992-5

Figure 30: Photocopy of a historic map dating 1832 displaying the dune area between The Hague and the village of Scheveningen (Enthoven, Drs. Francine Puttmann et al., 1992). The red frame indicates the site of the Jewish Cemetery.

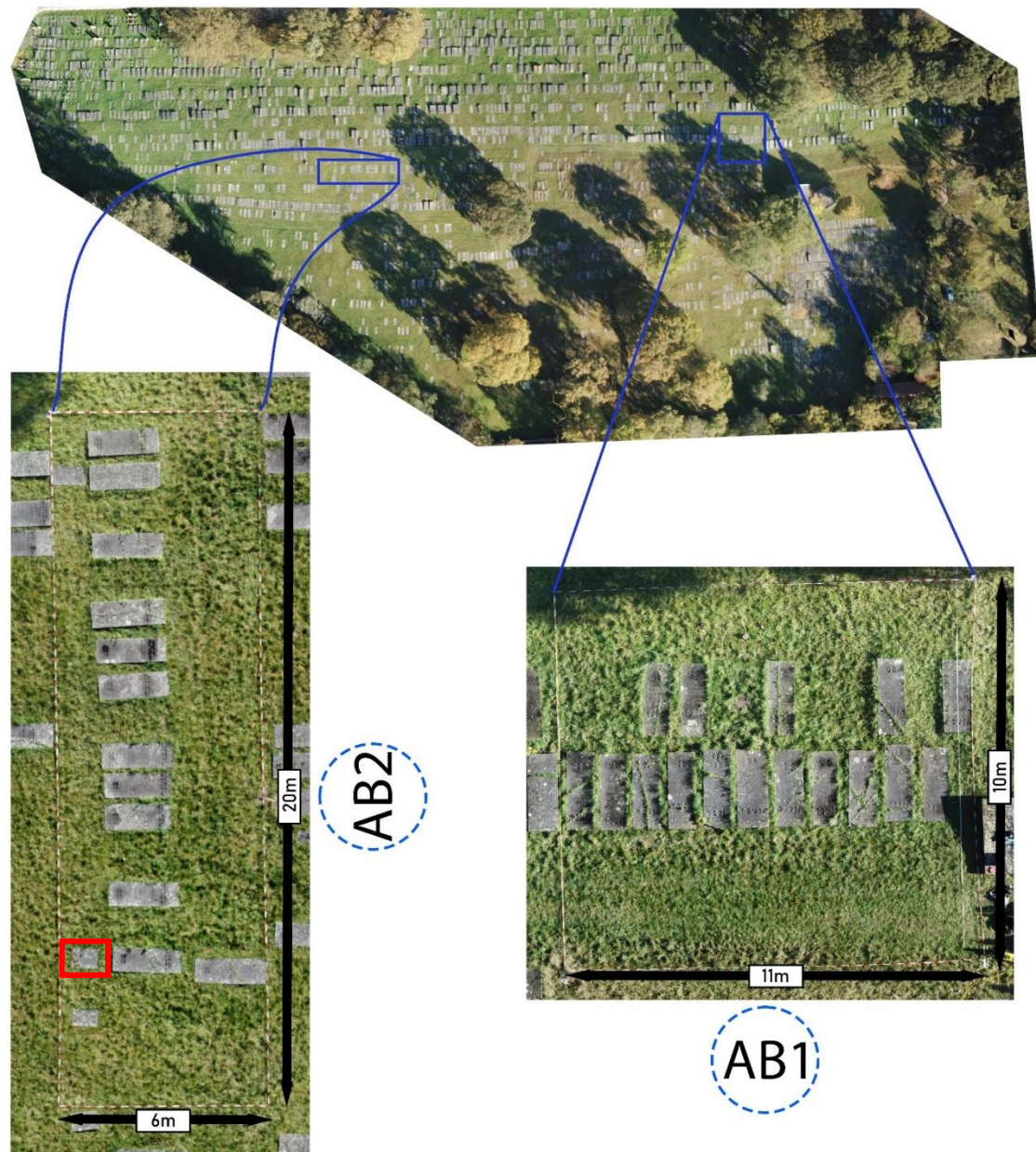
Annex B

Figure 31: Position and dimension of the two sample areas *AB1* and *AB2*. Note that the detail image of *AB2* is rotated 90° counter-clockwise. The red boundary box representing the patch *P1* for the visual evaluation. (Author).

Annex C

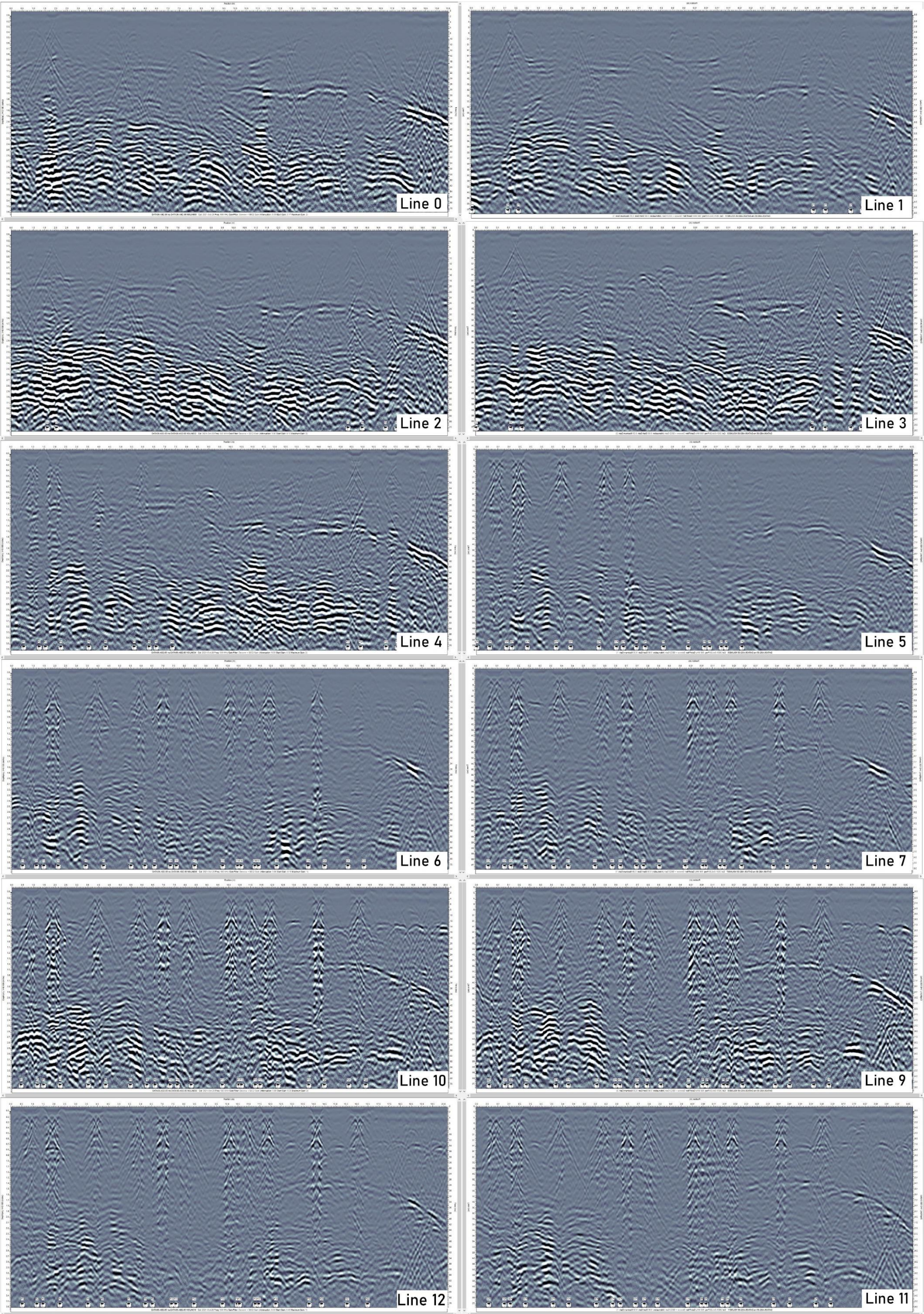


Figure 32: Scan lines (B-scans) Line 0–Line 11 of sample area AB2 (Author).

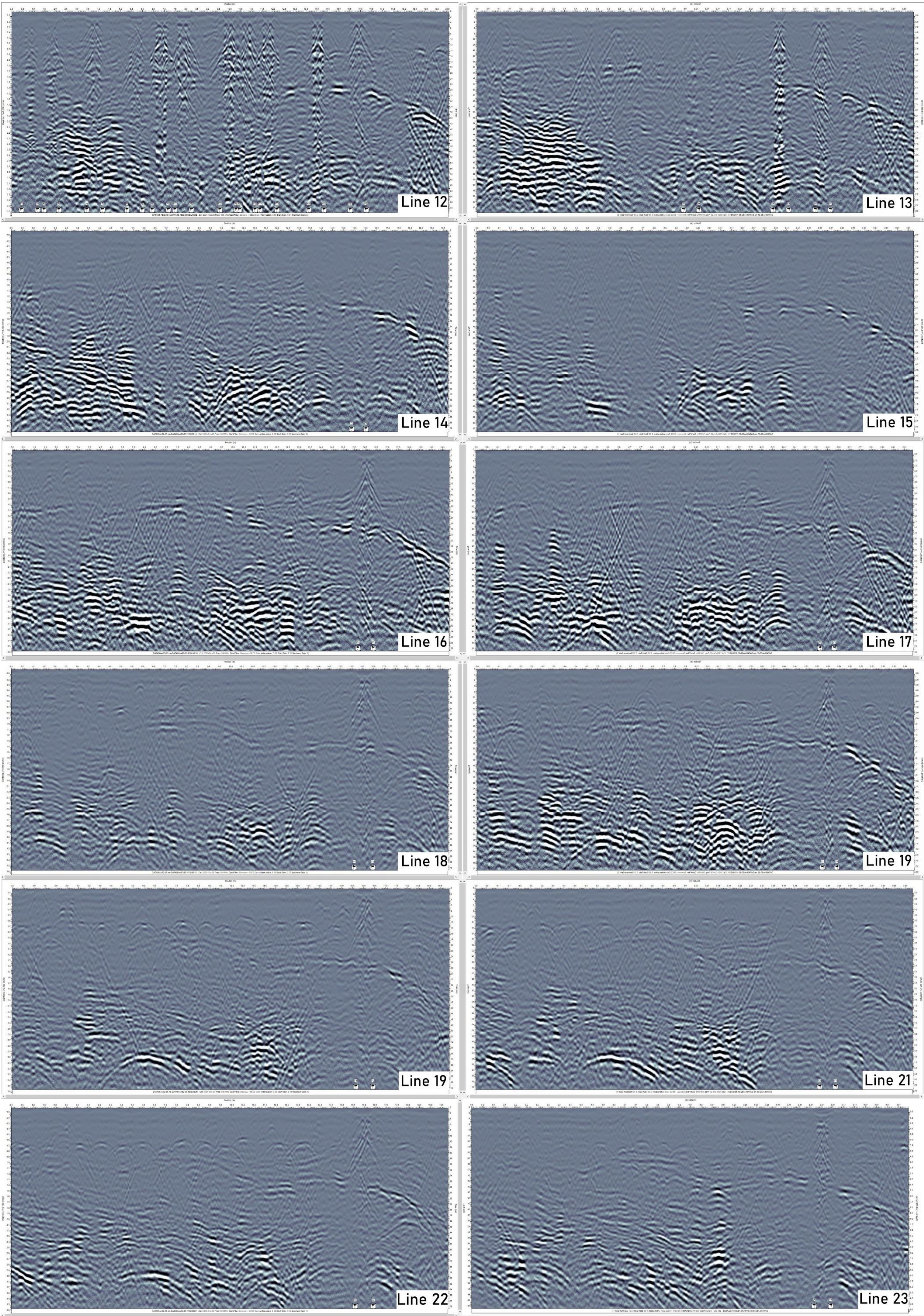


Figure 33: Scan lines (B-scans) Line 12–Line 23 of sample area AB2 (Author).

Annex D

Area	Point	x[m]	y[m]	h[m]	sx[m]	sy[m]	sh[m]	sxy[m]	sxh[m]	syh[m]
Area_AB1	100	80282.3842	456196.1483	4.0245	0.0029	0.0023	0.0045	0.0020	-0.0028	-0.0021
	101	80273.7286	456191.1479	3.9852	0.0014	0.0019	0.0035	0.0003	-0.0011	0.0008
	102	80268.1572	456200.6133	3.7707	0.0013	0.0015	0.0033	-0.0004	-0.0010	-0.0005
	103	80276.7459	456205.6745	3.3751	0.0010	0.0014	0.0028	-0.0005	-0.0005	0.0001
	900	80277.6219	456196.4266	3.8787	0.0011	0.0013	0.0027	-0.0004	-0.0006	0.0007
	901	80273.2244	456203.3122	3.7669	0.0010	0.0013	0.0027	-0.0004	-0.0005	0.0007
Area_AB2	200	80226.7645	456269.3332	4.0703	0.0012	0.0016	0.0034	-0.0002	0.0008	0.0008
	201	80221.5137	456266.4332	4.0371	0.0012	0.0015	0.0031	-0.0004	0.0010	-0.0005
	202	80211.6147	456283.8473	3.8383	0.0013	0.0017	0.0040	-0.0004	0.0003	-0.0006
	203	80216.8955	456286.6781	3.9195	0.0012	0.0014	0.0033	-0.0003	0.0004	-0.0011
	902	80217.8814	456282.6776	4.0570	0.0012	0.0014	0.0034	-0.0003	-0.0005	-0.0012
	903	80220.2043	456274.2509	4.1795	0.0013	0.0017	0.0036	0.0004	0.0008	-0.0007
Area_C1	300	80254.1353	456127.1661	5.6973	0.5483	0.6233	0.9820	0.2215	-0.2913	-0.2933
	301	80264.9224	456132.3085	4.3525	0.6791	0.9037	1.5354	0.2768	-0.3939	0.4200
	302	80267.6698	456127.0079	4.8365	0.5939	0.7133	1.1035	0.3694	-0.3508	-0.5270
	303	80256.6536	456123.6056	4.5727	0.5838	0.6483	1.0004	0.2699	-0.3733	-0.4328
	303	80260.2056	456131.7564	5.0291	0.0163	0.0258	0.0513	0.0096	-0.0205	0.0188

Table 10: Coordinates of reference points with error residuals (reference system RD new [Hz] and NAP [V]) (Author)

Annex E

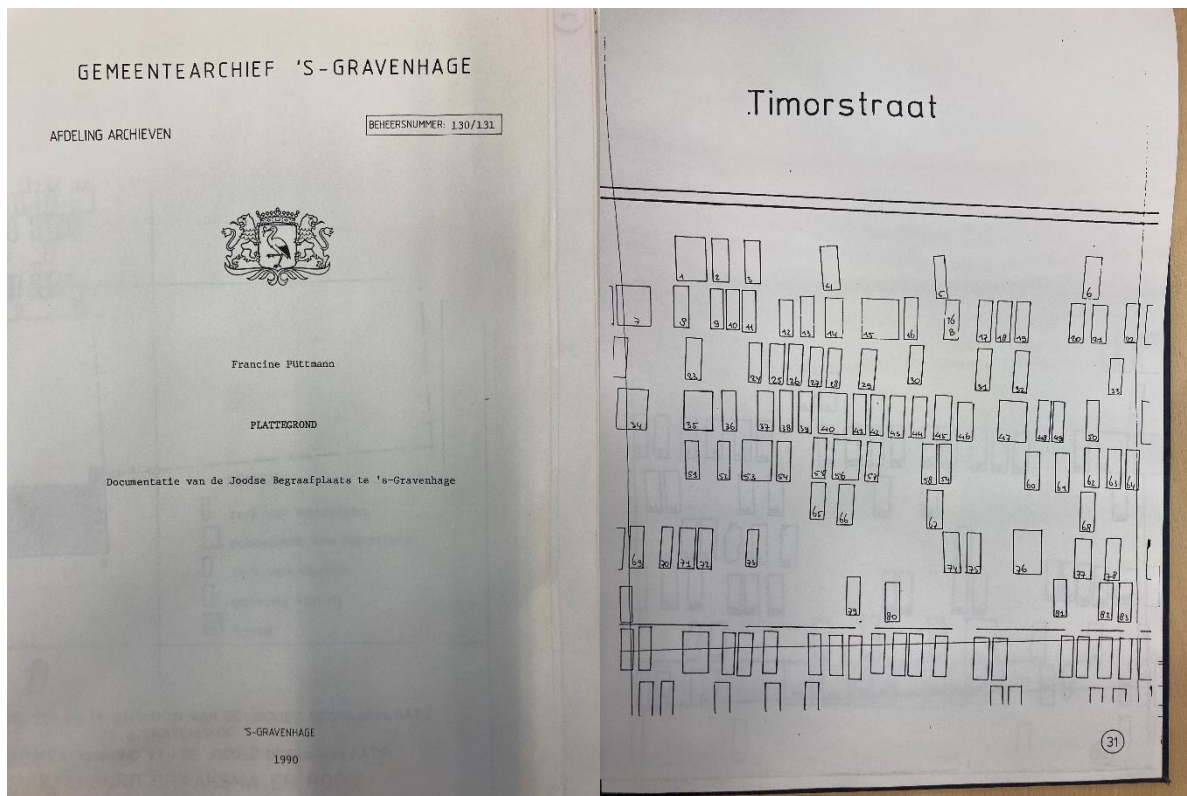


Figure 34: Coversheet (left) and example of detail layout plan (plattegrond) of the cemetery with individual gravestone numbers for each plan number (Author)

Naam	Overlijdensdatum	Zerknummer			formulier
(Cohen Joel van de Rool)					3138
() Boas	1729.07.22	2/48	dr v Chaim Boas	vr v Aberle Segal	398
() Hamel	1818.09.11		dr v Feij []	vr v () Hamel	2849
() Levi			dr v Jokef Levi		2860
() de Pinto	1818.07.05	7/45	dr v Ml.A. de Pinto		
() Dula()	1885.02.18	23/7	dr v Mordechai Dula()	vr v Samuel Prins	
()	1780.04.09		kind (meisje) v Abraham zn v Hirts		1311
() Hes	1779.02.12		kind v Ruben Hes		1246
() Perets	1815.12.31		zn v Natan Perets		2650
() Fresko	1804.06.28		1e dr v Jona Fresko	schoonzn v Natte Pos	2311
()eile Levi	1789.03.23		kind v Zanwil (zn? v) Mozes Israel	zn v Jakob Levi	1641
[] Suasso	1770.02.21	7/165	Suasso		5072
[]			Bakker		48
[] Abendana	1729		dr v [] Abendana		4757
[] Parde			dr v [] Parde		3471
[] Ephratty	1772.05.02		dr v Aaron Israel Ephratty		4979
[] Kuit			dr v Aaron Kuit		3070
[] Cohen	1770.02.09		dr v Aaron	zn v Mozes Cohen	954
[] Prins	1797.12.01		dr v Aaron	zn v Wolf Prins	1995

Table 11: Extraction (20 entries) of the database of the cemetery register with names (Naam), date of death (Overlijdensdatum), Gravestonenummer (Zerknummer), parents, church register number (formulier) (Francine Puttmann).

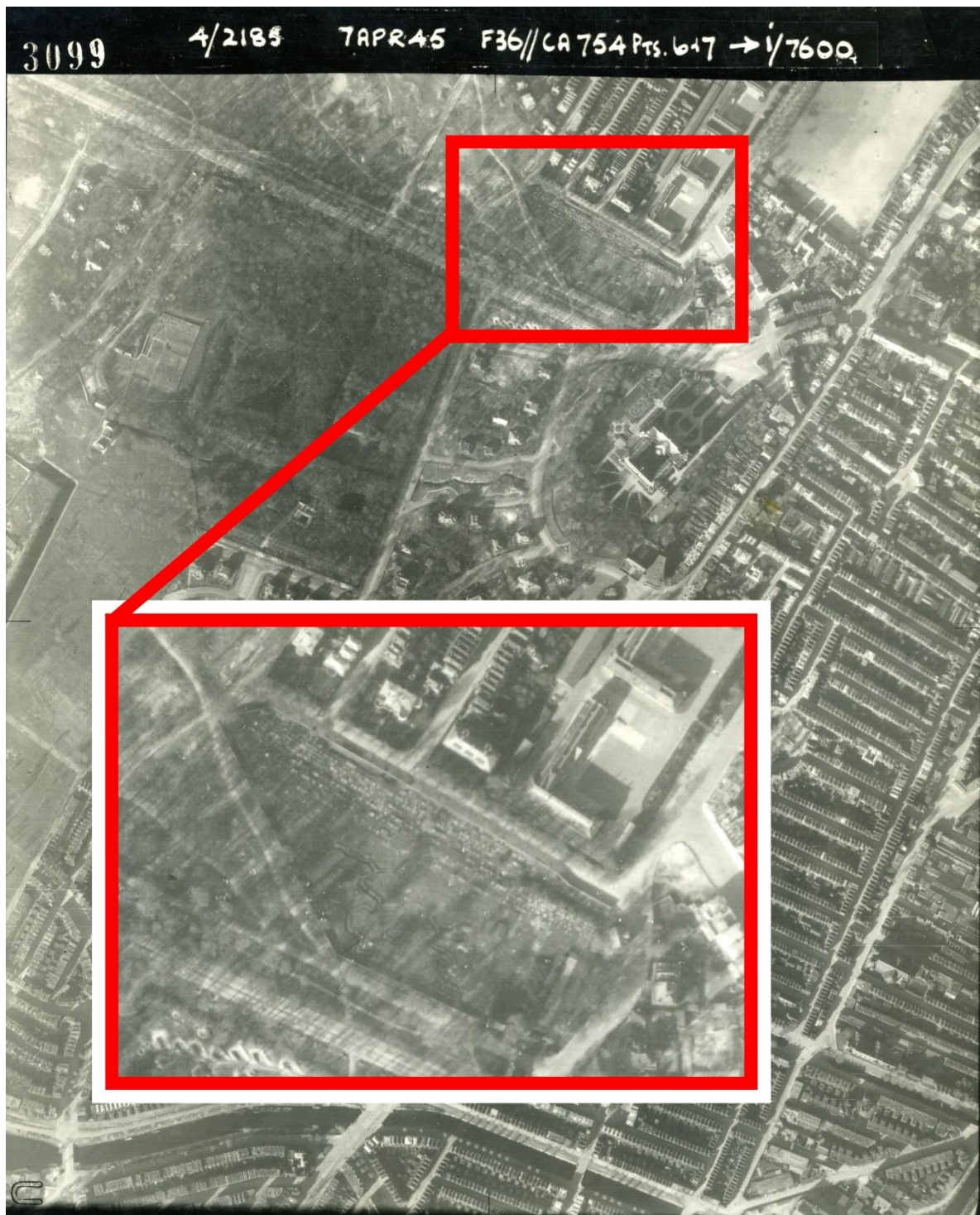
Annex F

Figure 35: Aerial imagery of the R.A.F. dated from 07/04/1945 covering the area of interest with a blow-up of the cemetery (haagsgemeentearchief).

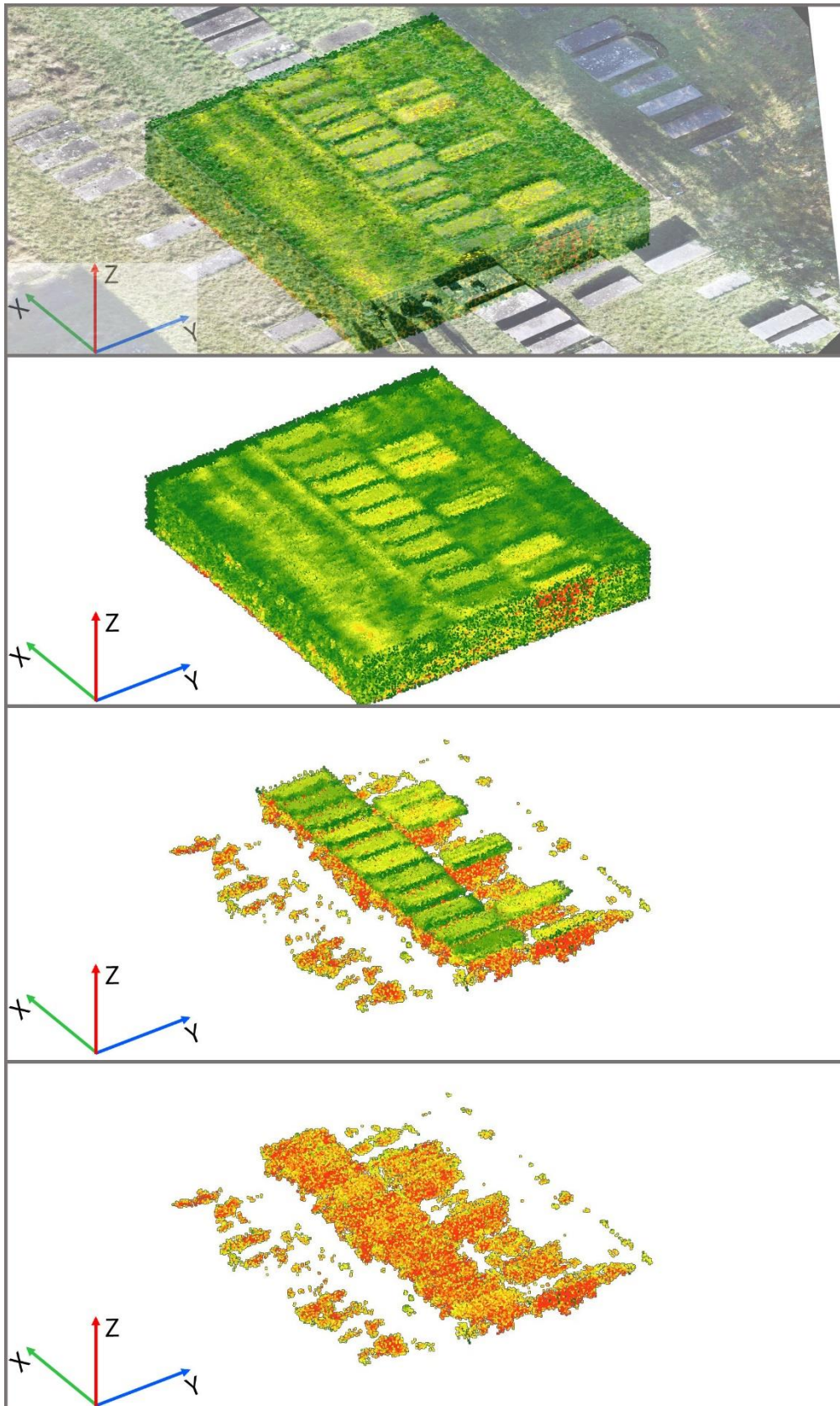
Annex G

Figure 36: Block-shaped point cloud area AB1 with attribute based colouring depending on the assigned intensity value. TOP to BOTTOM – with orthomosaic superimposed; point cloud displaying all three classes; point cloud displaying the classes #1 and #2; point cloud displaying class #1 (Author).

Annex H

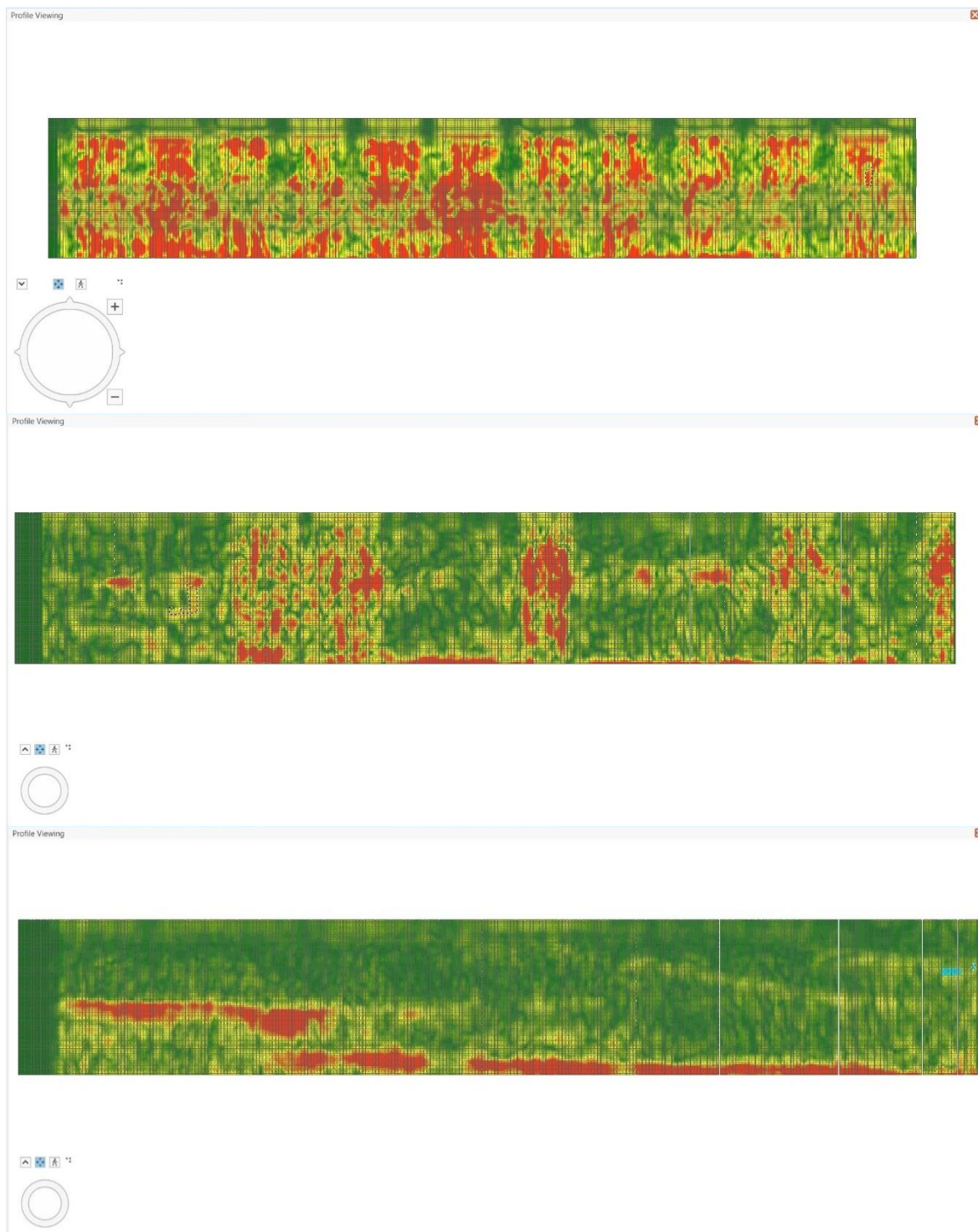
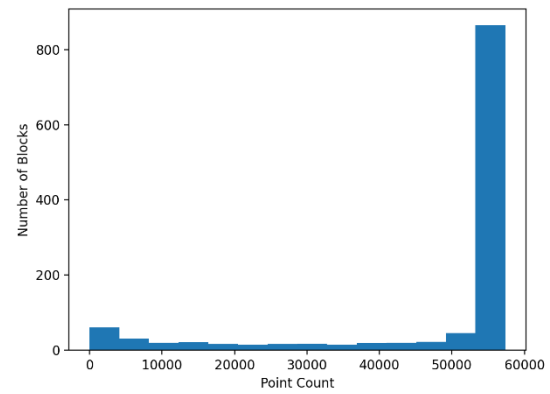
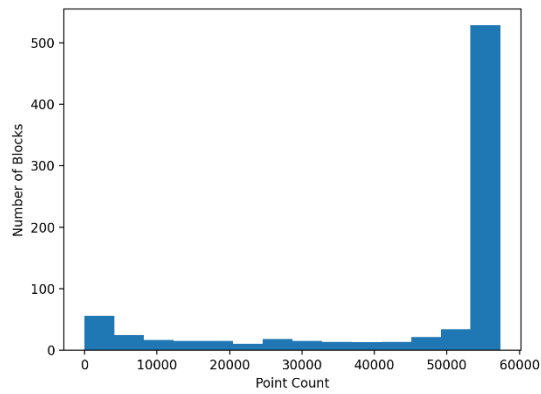
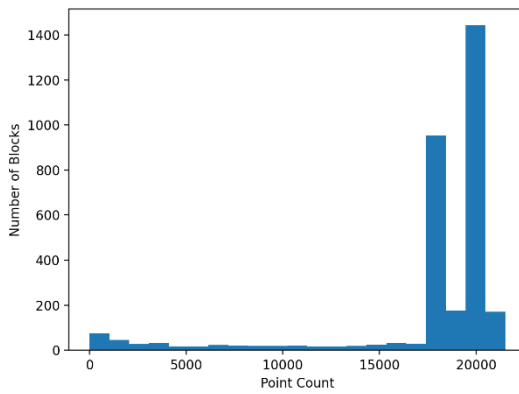
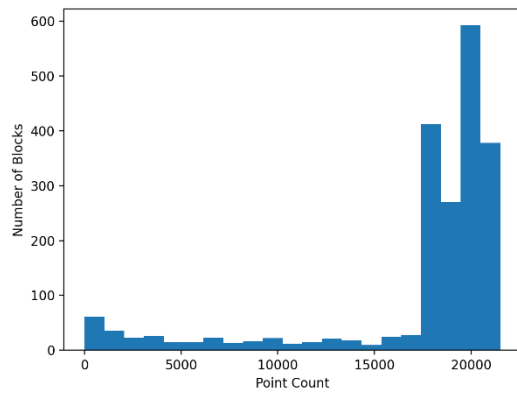


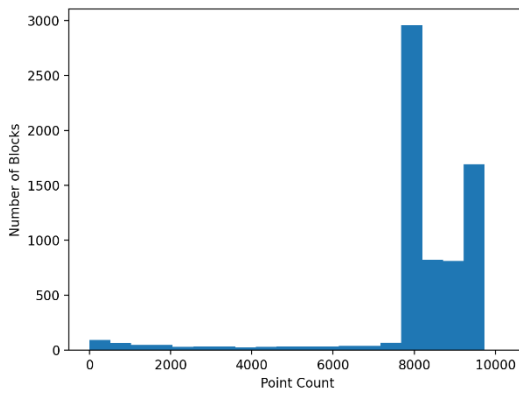
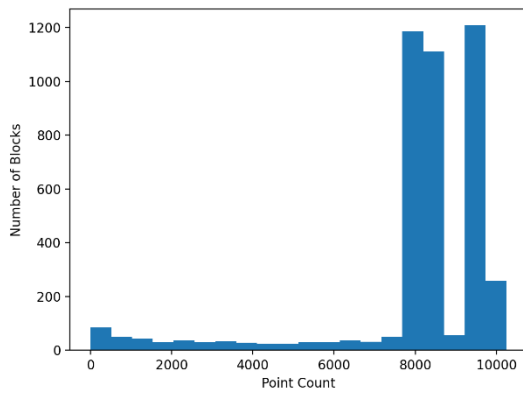
Figure 37: Three profile views showing the variety of the underground properties of the sample area AB1. In the top window the disturbances caused by the gravestones (light green square shapes at the surface) are prominent. The centre image displays the difference between burial locations marked and unmarked by a gravestone. The bottom image is a slice of undisturbed soil property where intensity values mark natural sediment changes (Author).

Annex J

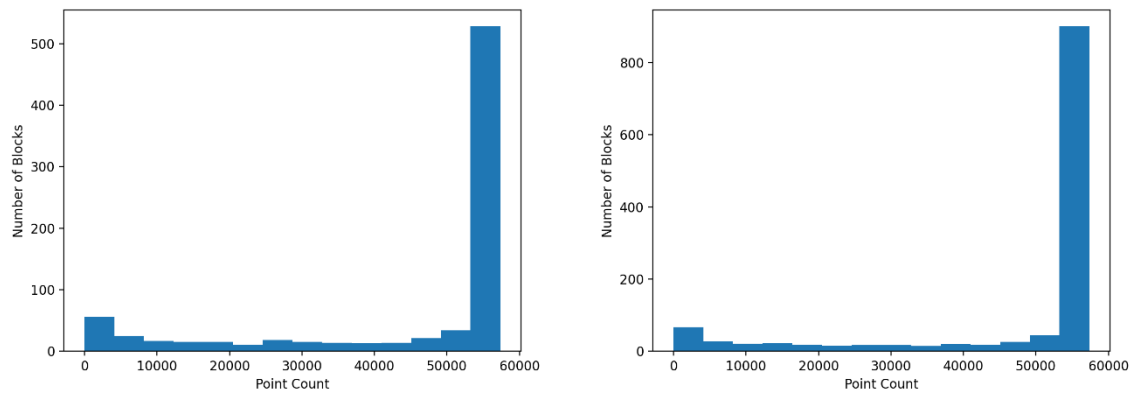
Training data #1



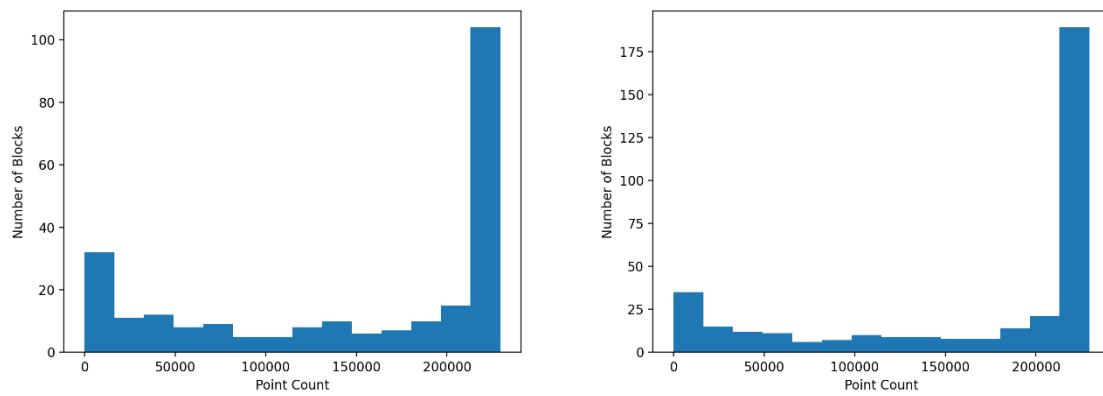
Training data #2



Training data #3



Training data #4



Training data #5

Figure 38: Histograms displaying the Block and Point-distribution of the different training- (left) and validation- (right) data sets (Author).

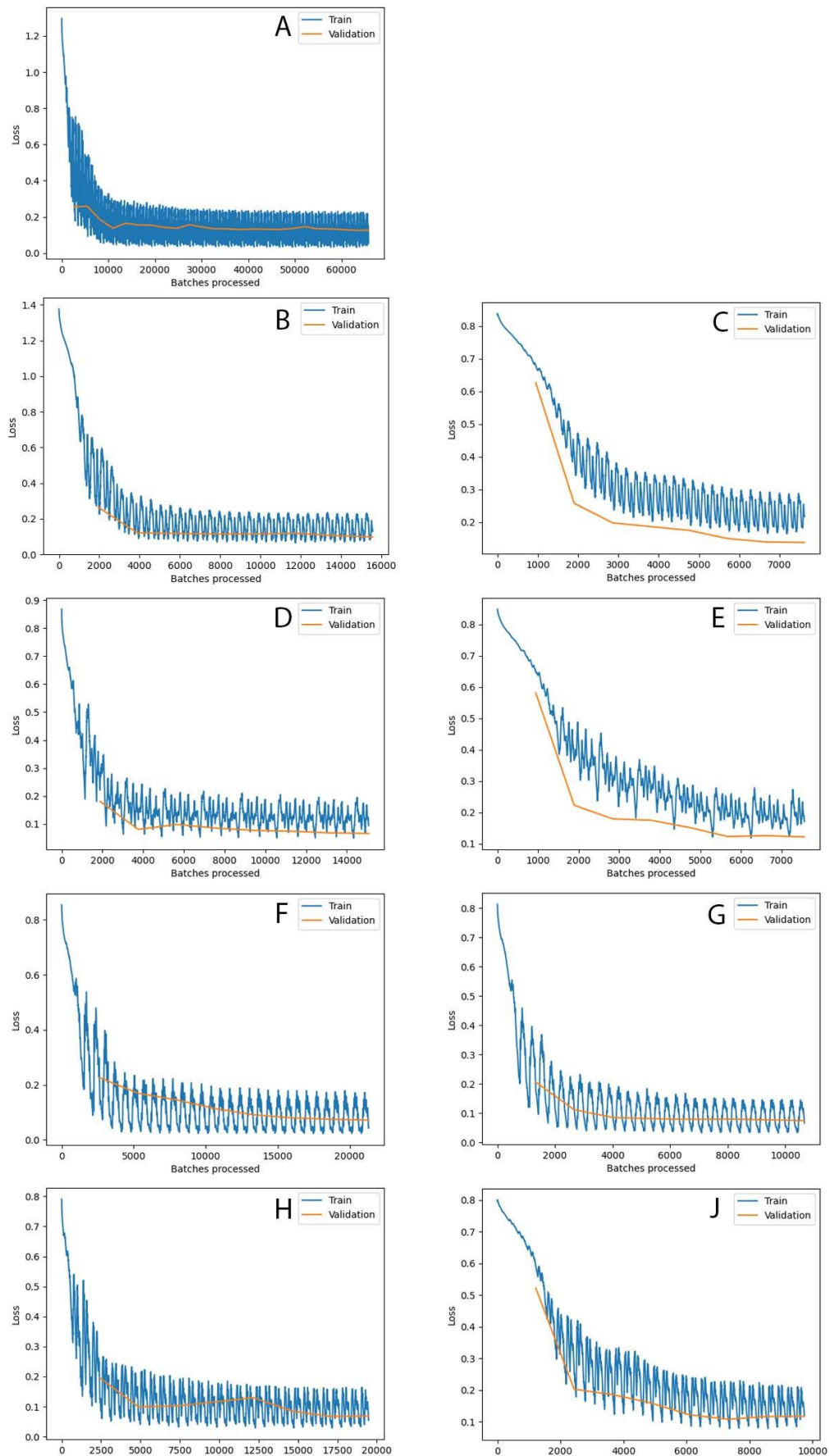
Annex K

Figure 39: Loss graphs of the nine model training sessions (A – J) (Author).

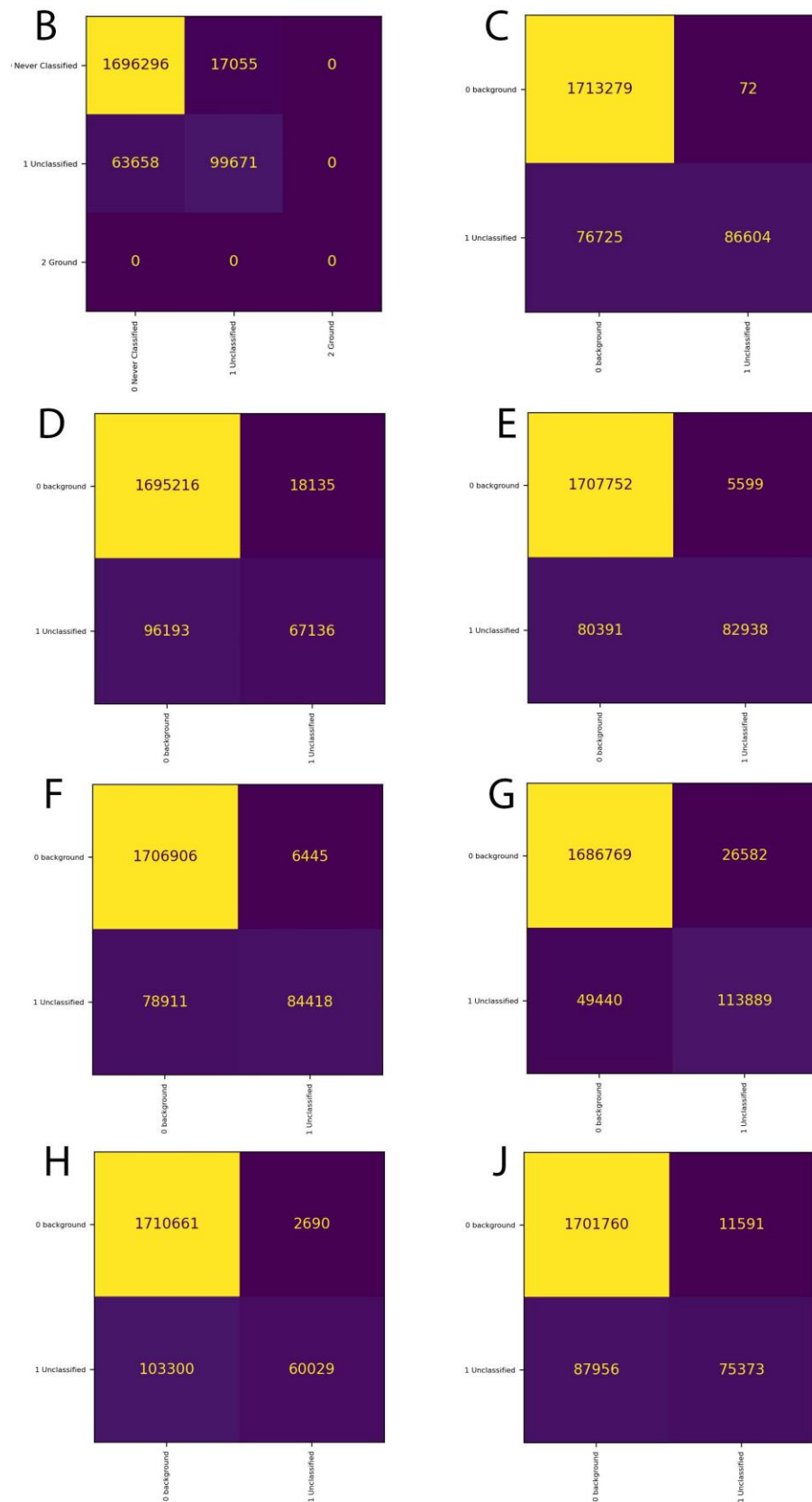
Annex L

Figure 40: Confusion matrices by model evaluation on area AB2 (Model B – J) (Author).

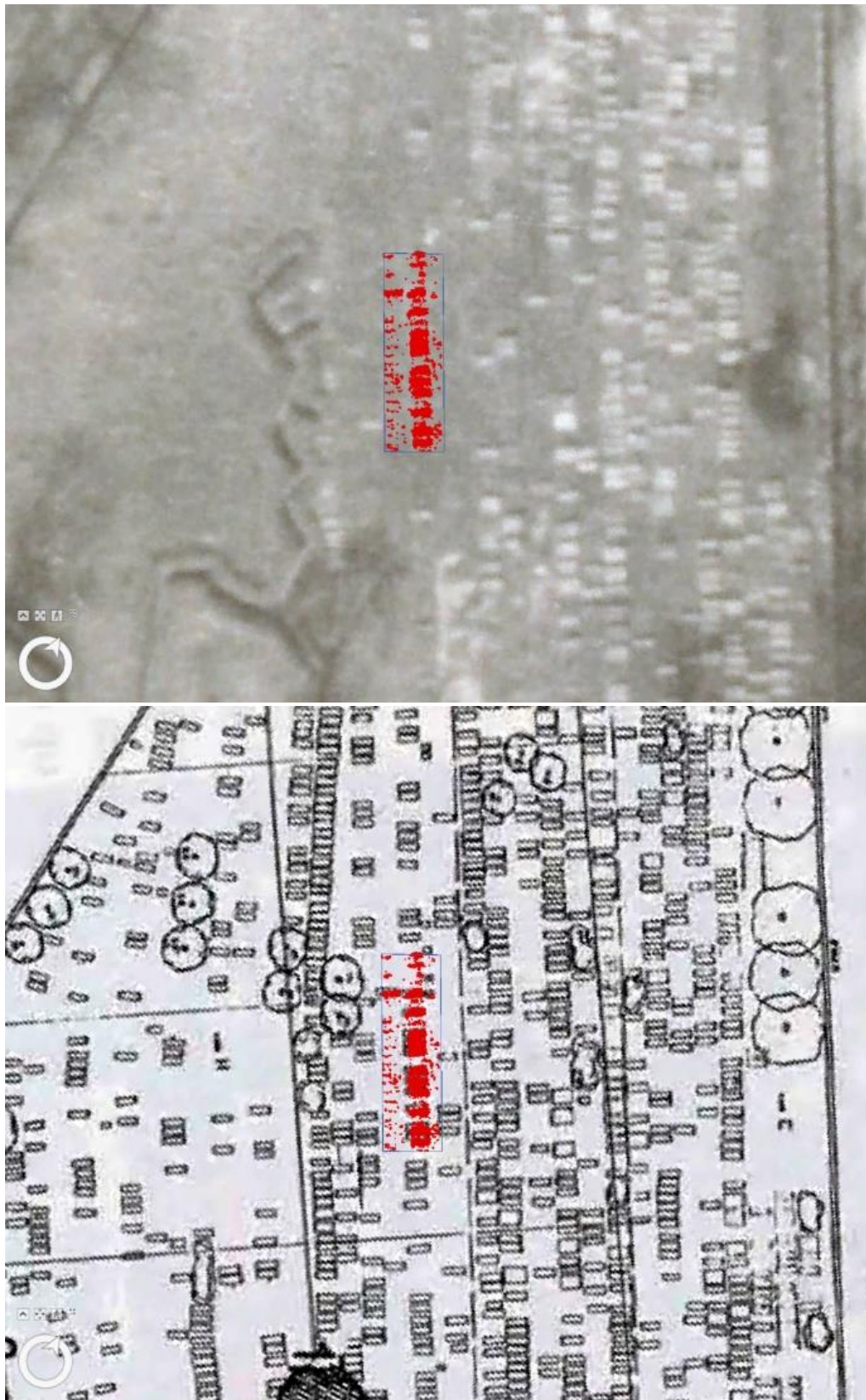
Annex M

Figure 41: Classified point cloud of area AB2 superimposed with historical aerial imagery (TOP) (DotKa, 2022) and layout plans (BOTTON) (Enthoven, drs. Francine Puttmann *et al.*, 1992).