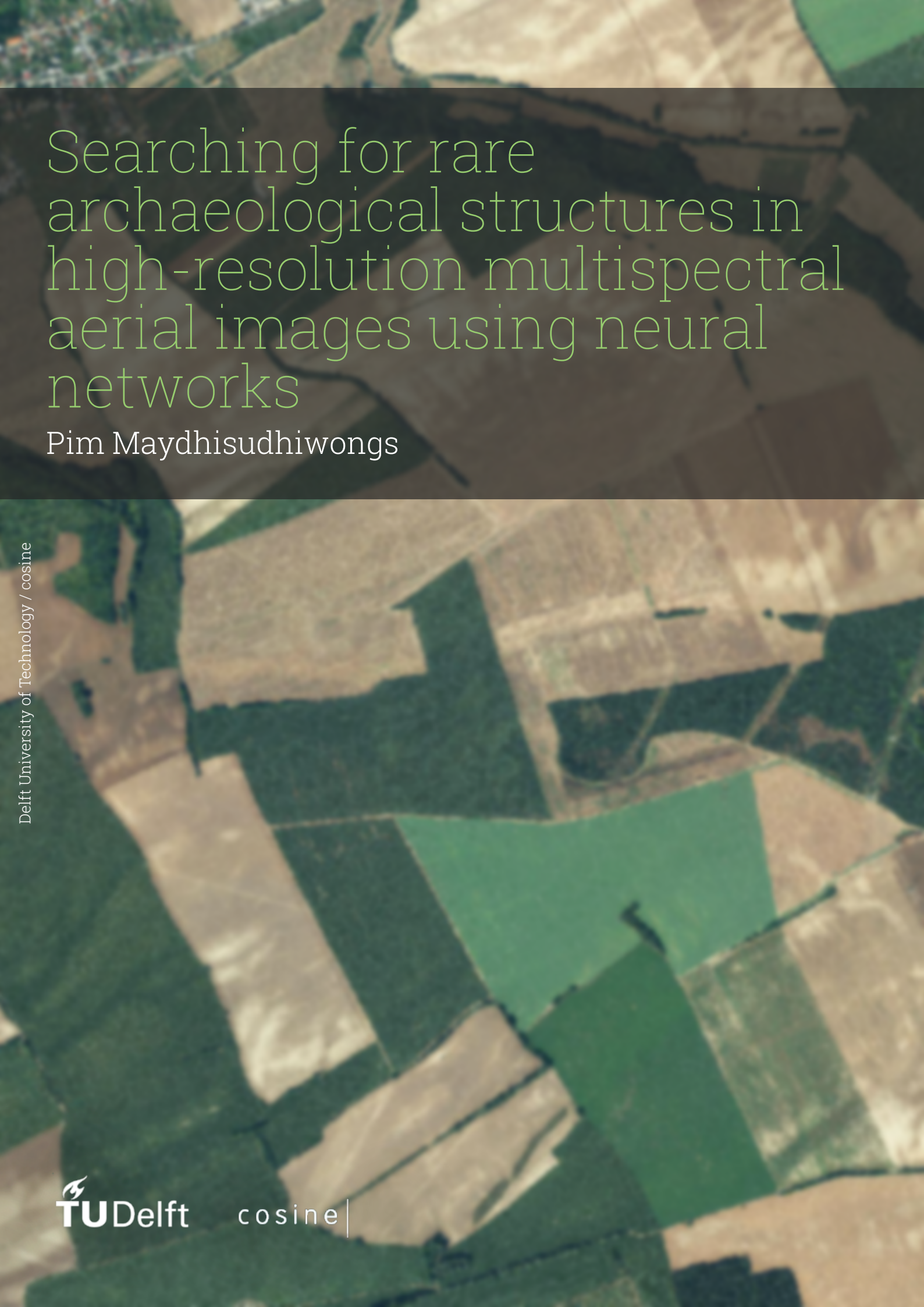

Searching for rare archaeological structures in high-resolution multispectral aerial images using neural networks

Pim Maydhisudhiwongs

Delft University of Technology / cosine

TU Delft cosine

Searching for rare archaeological structures in high-resolution multispectral aerial images using neural networks

by

Pim Maydhisudhiwongs

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on February 12, 2025 at 2:00 PM.

Student Name	Student Number
Pimmbhattra Maydhisudhiwongs	5691907

TU Delft Supervisor:	Roderik Lindenbergh
TU Delft Co-supervisor:	Marc Schleiss
Company Supervisor 1:	John Hefele
Company Supervisor 2:	Nathan Vercruyssen
Faculty:	Civil Engineering and Geosciences
External company:	cosine Remote Sensing BV

Style: TU Delft Report Style, with modifications by Daan Zwaneveld
Cover: Geoportál.sk

Abstract

Few rare, circular, concentric enclosure ditches called rondels were discovered in Slovakia, a country in Europe; within the ditches, material traces of Neolithic European culture can be excavated. For excavations to happen, archaeologists must first locate these rare structures. Most rondels were spotted on agricultural fields or searched for manually during aerial surveys in a time-consuming manner. With the release of a high-resolution multispectral aerial orthophotomosaic data set of Slovakia, detailed sites containing rondels may be discovered using machine learning techniques.

Machine learning techniques using convolutional neural network (CNN) models can be applied to the field of archaeology to search for excavation sites automatically. An obstacle remains: CNN models require a lot of training image data to be efficient at their task, which is to classify whether areas contain rondels or otherwise. There are only 20 visible rondels on the orthophotomosaic that can be used as training images for model input, creating an imbalanced data set of a class with a minority class of aerial images of rondels and a large majority class of aerial images without rondels. Sketches and recorded characteristics of rondels from current images and from archaeological publications were used to automatically and randomly replicate rondel appearances from above, resulting in a created balanced data set with sufficient rondel examples for CNN training.

Multiple ResNet-34 models and a ConvNeXt model with differing hyperparameters were trained. The most promising model, a modified ResNet-34, was selected based on validation loss from the cross-entropy loss function and on the number of correctly and incorrectly classified labeled images from a test set. The selected model is used to classify data from the orthophotomosaic for rondels using a sliding window technique. Over 9510 square kilometers of agricultural land cover in western and eastern Slovakia was selected from the CORINE land cover map for classification. 7 suspected rondel sites were found, and 2 were determined to likely be rondels, based on their circular ditch-like appearance in 4 sets of multispectral images and in LiDAR elevation data. Results indicate that exact rondel layouts can be delineated with high-resolution orthophotomasics, however identifying circular elevation patterns of ditches proves to be challenging without using additional LiDAR data.

Contents

Abstract i

1 Introduction 1
1.1 Objective . . . 1
1.1.1 What are rondels and why are they important? . . . 1
1.1.2 What are convolutional neural networks? . . . 2
1.1.3 What are RGBN aerial images? . . . 2
1.2 Problem Statement . . . 2
1.3 Research questions . . . 2

2 Background 3
2.1 Rondels . . . 3
2.1.1 Characteristics of rondels . . . 3
2.1.2 Known and located rondels . . . 5
2.2 Convolutional neural network models . . . 8
2.2.1 Object detection and segmentation . . . 8
2.2.2 Image classification . . . 9
2.3 Class imbalance . . . 9
2.4 Previous research . . . 10

3 Data and study area 11
3.1 Data sources . . . 11
3.1.1 Orthophotomosaic of Slovakia - 2nd cycle . . . 11
3.1.2 Reference data: 1st cycle orthophotomosaic . . . 12
3.1.3 Rondel sketches . . . 12
3.1.4 CORINE land cover map . . . 13

4 Methodology 15
4.1 Training, validation, and test set folder construction . . . 15
4.2 RGBN image extraction . . . 16
4.3 Synthetic training sample generation . . . 17
4.4 Hardware . . . 19
4.5 Convolutional neural network training . . . 19
4.5.1 Model setup . . . 19
4.5.2 Hyperparameter tuning . . . 20
4.6 Model selection . . . 22
4.6.1 Cross-entropy loss function . . . 22
4.6.2 Test set performance . . . 22
4.7 Sliding window inference . . . 23

5 Results and Discussion 25
5.1 Final data set size . . . 25
5.2 Evaluating and selecting a trained model . . . 25
5.2.1 ResNet evaluation . . . 25
5.2.2 ConvNeXt evaluation . . . 28
5.3 Final ResNet model evaluation . . . 29
5.4 Comparing results from the generated training set . . . 33
5.5 Rondel search and inference . . . 34
5.5.1 Site 1: 48.21540, 17.55654 SERED_5-8 . . . 35
5.5.2 Site 2: 47.94624, 17.69755 DUNAJSKA_STREDA_2-3 . . . 36
5.5.3 Site 3: 48.22847, 17.84106 MOCENOK_708 . . . 37

5.5.4 Site 4: 48.33399, 17.78244 MOCENOK_8-2 . . . 39
5.5.5 Site 5: 48.11326, 18.83470 PLASTOVCE_8-7 . . . 40
5.5.6 Site 6: 48.55702, 17.70455 SENICA_0-9 . . . 41
5.5.7 Site 7: 48.08548, 18.35974 SURANY_2-7 . . . 43

6 Conclusion and future work **45**
6.1 What worked? . . . 45
6.2 What did not work? . . . 46
6.3 Rondel visibility . . . 48
6.4 Answers to research questions . . . 49
6.5 Future work . . . 50
6.5.1 Data level improvements . . . 50
6.5.2 Model level improvements . . . 50

References **51**

1 Introduction

1.1. Objective

The objective of this thesis project is to automatically search for rare, circular, Neolithic era structures in Slovakia called rondels with few remaining positive examples in RGB and near-infrared aerial images using a trained convolutional neural network model.

1.1.1. What are rondels and why are they important?

Rondels are enclosure ditches arranged in circular patterns. For reasons currently unknown, they were built around 6800 years ago during the Neolithic [5]. The ditches are concentric with entrances. The circular pattern can be seen more clearly from above; figure 1.1 shows an example of a rondel from an aerial photograph, the rondel has two clear rings with an opening to the right of the rings [4].

Slightly over a hundred suspected rondels have been reported across Slovakia, therefore rondels are confirmed to be sparingly sighted [5]. Outside of Slovakia, similarly, rondels are found scattered across other central and eastern Europe countries in less urbanized areas [5], usually in agricultural fields where their remains have not been removed by construction projects.

We are interested in searching for rondel sites because, for archaeologists, rondels contain unearthed material remains of past Neolithic cultures and are one of the earliest examples of ancient European architecture. It has been theorized from objects at excavated ditches that rondels may have served a ceremonial purpose. At a few sites, declinations in celestial bodies can be observed during equinoxes or solstices at their openings [5] [51]. For archaeologists to excavate rondels, the sites should first be located using a non-destructive method, preferably by using aerial imagery [4].

Figure 1.1: Remains of a rondel in Golianovo, Slovakia seen from an aerial photograph [4].

1.1.2. What are convolutional neural networks?

Convolutional neural networks (CNNs) are a type of layered neural network model commonly used to classify data with multiple channels and dimensional sizes such as colored images. Convolutional and aggregating operations are applied to multiple layers to summarize and reduce spatially significant patterns from images for the classification process, making CNNs a viable choice for classifying archaeological fields in large, terabyte scale spatial data sets the size of a country [28]. To do this, CNNs must first be trained on 2 classes of aerial images- in this case images of rondels and images without rondels. The more a CNN 'sees' a class of trained images, the more the model is updated to effectively classify images of that class [25].

1.1.3. What are RGBN aerial images?

The Slovak cadastre released a complete second orthophotomosaic data set of Slovakia in 2023 created from processed airborne photos [21] at a 20 centimeter resolution. The orthophotomosaic contains processed multispectral imagery of the country collected via 4 channels: visible red, green, and blue colors, and a near-infrared channel. We will be using this data set to train our CNN and search for rondels from above.

1.2. Problem Statement

Not a lot of machine learning techniques have been implemented in archaeological studies as they have been used in engineering or scientific research [2]. Slovak archaeologists have been using aerial surveys since 1963 to search for Roman forts [4]. We are planning to close this research gap by training neural network models to search for rarely spotted archaeological sites on aerial orthophotos. Physically searching for rondels during field campaigns is time consuming for archaeologists. With a published dataset of Slovak aerial photos, it is possible to 'scan' through the photos for rondels with a neural network model.

However, few problems remain: First, neural networks require large amounts of inputted training images to be effective at classifying an object of interest. A hundred positive examples of visible rondels are not enough to train the network adequately [30]. And second, seasonal changes in crop growth over agricultural fields can potentially obscure the remains of a rondel from above [3]. We are going to address the problem statement by conducting research to answer the following research questions:

1.3. Research questions

Main question: How can we detect new rondels in RGBN aerial images with limited positive examples?

Sub-question 1: Which configurations and parameters can we use for the training data set and the convolutional neural network models that will result in an effective model for finding new rondels?

Sub-question 2: Can we differentiate any seasonal vegetation changes that would affect rondel appearances in the orthophotomosaic data set?

Sub-question 3: What additional value do high-resolution multispectral images bring when it comes to rondel detection? How useful are they compared to LiDAR data?

2
Background

This chapter introduces the characteristics and locations of Slovak rondels, a brief overview of convolutional neural network (CNN) model mechanics, and what the CNN models are used for. We will also discuss previous archaeological research utilizing digital images and neural network models, and past research specifically on searching for Slovak rondels.

2.1. Rondels

The exact purpose rondels serve and how they originally looked like around the time they were built remains unclear to researchers. Archaeologists have theorized that rondels could have served as a place for social gatherings such as ceremonies and celebrations, or as a marketplace, based on tools, ceramics, and stone materials found at excavated sites. Rondels could have served as burial sites based on human and animal remains found inside the ditches or as astronomical sites, similar to the British Stonehenge where openings align with declining celestial bodies at certain times of the year [5].

2.1.1. Characteristics of rondels

Most recorded rondels range from 40 to 200 meters across- there may be a large variance in their sizes but all rondels are arranged in circular and oval patterns. The ditches contain 1 to 3 main concentric rings with 2 to 5 entrances leading from the field outside the largest ditch to the innermost ditch [5]. A general ”blueprint” of their layout is shown in figure 2.1.

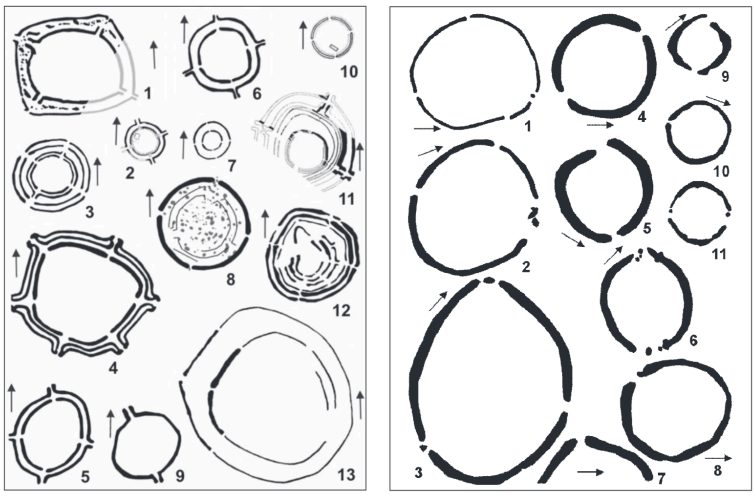

Figure 2.1: Recorded Slovak rondel ”blueprints” from Slovakia from a publication by Ridky (2007) [4].

When viewed from the ground level, rondels appear as circular ditches like in figure 2.2; this is particularly visible at excavated sites. The majority of spotted rondels are on agricultural fields where they are covered by crops, but their circular ditch structures remain beneath. The entire ditch-like structure becomes more apparent from higher ground and from aerial photos [42] [4].

Figure 2.2: An excavated ditch near Prague in the Czech Republic. Photograph by Google user Vladislav Urban. [20]

Circular rondel features can be spotted in visible RGB and near-infrared imagery as crop or soil marks because they lead to anomalous vegetation growth over ditches. Soil inside rondel ditches on vegetated terrains (like agricultural fields) are able to culminate more organic materials and water than their surroundings, leading to higher moisture retention. This causes vegetation growing over the ditches to grow healthier and taller than ones growing over non-ditched areas [46]. During drier seasons of the year, the contrast is more visible from above than during colder and wetter seasons [3].

The data set used in this project is an orthophotomosaic generated from multispectral aerial photography in the visible (RGB) and the near-infrared channels [21]. Multispectral refers to the data set's multiple channels. Although various crop types and camera sensors have different reflectance characteristics, generally, near-infrared light is reflected more strongly by chlorophyll in healthier vegetation and RGBN aerial imagery captures this contrast well [8]. Green light is reflected the most out of the three visible channels, creating a stronger and darker color over rondel ditches [46].

2.1.2. Known and located rondels

Figure 2.3 shows a total of 129 suspected rondels have been reported in Slovakia; but only 20 of them are clearly visible on the orthophotomosaic data set. Even fewer in Slovakia were well-studied or excavated [5]. Most of the recorded rondels were spotted from surveys and previous research studies in western Slovakia's fields overlooking rivers [36] [4]. It should be noted that a majority of archaeological aerial surveys done by Slovak researchers coincidentally take place in the western part of the country which has been Slovakia's most densely populated area since the Neolithic to the present day [4].

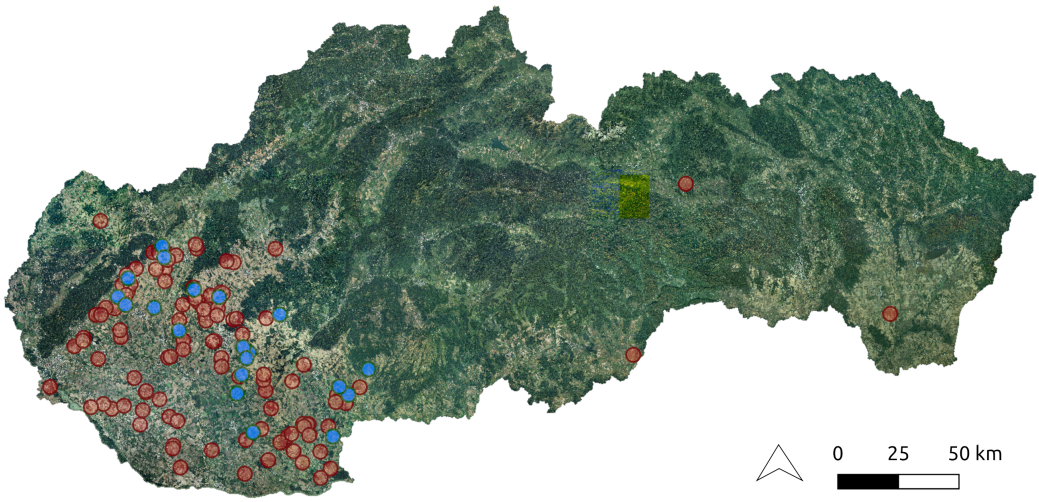

Figure 2.3: Map of 129 rondel locations on the orthophotomosaic data set provided by projects and publications [36] [4] [5]. 109 red circles are non-visible, but reported rondels, while 20 blue circles indicate visible rondels.

The 20 visible rondels marked in blue are shown in figure 2.4 in the red, green, and blue (RGB) channels. They are also shown in figure 2.5 in the near-infrared (N) channel. The majority of visible rondels appear as a darker circle on the field in the RGB and N channels.

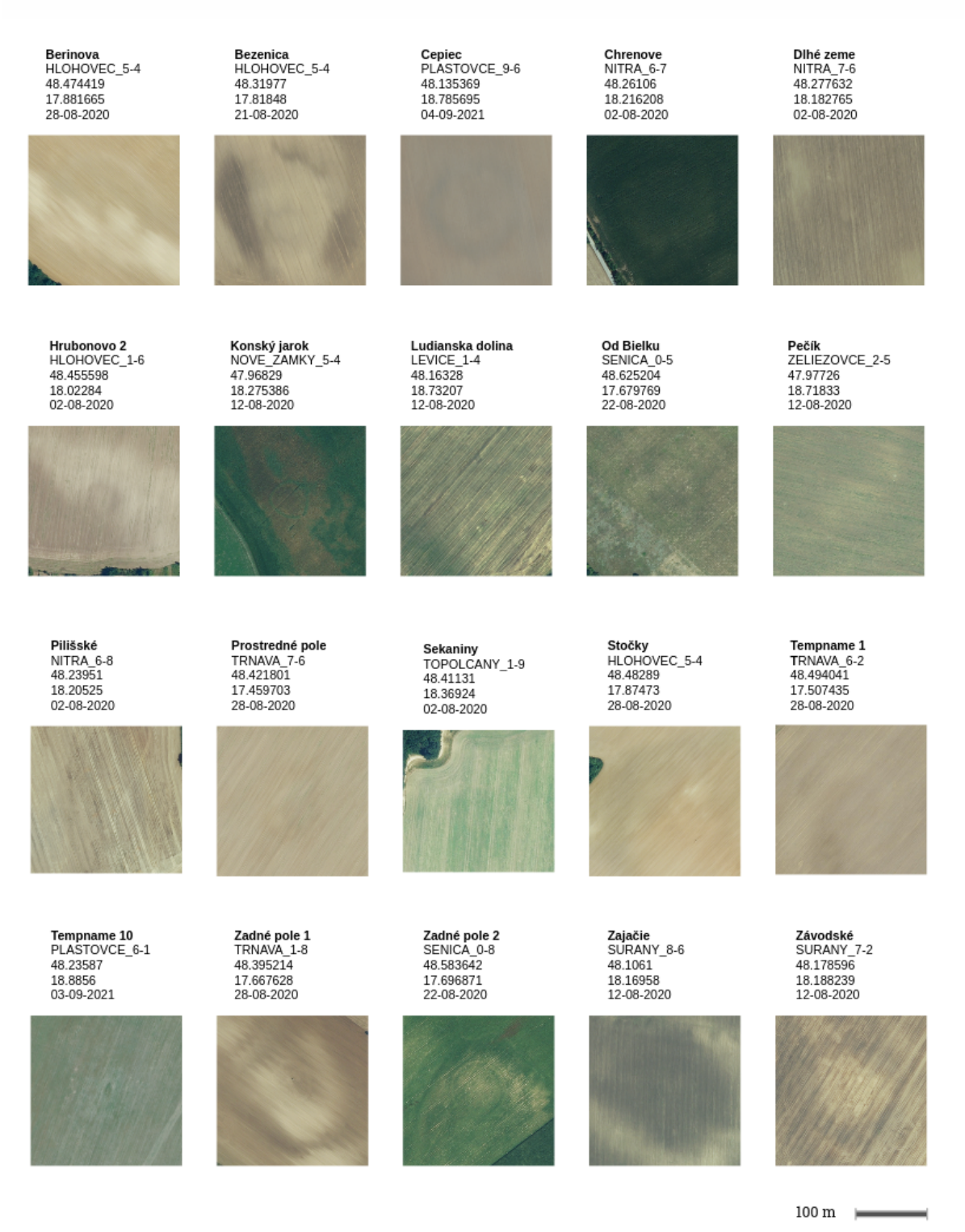

Figure 2.4: Rondels (with location names) visible in the 2nd cycle orthophoto, file name, WGS-84 latitude and longitude, and image date is included. RGB channels.

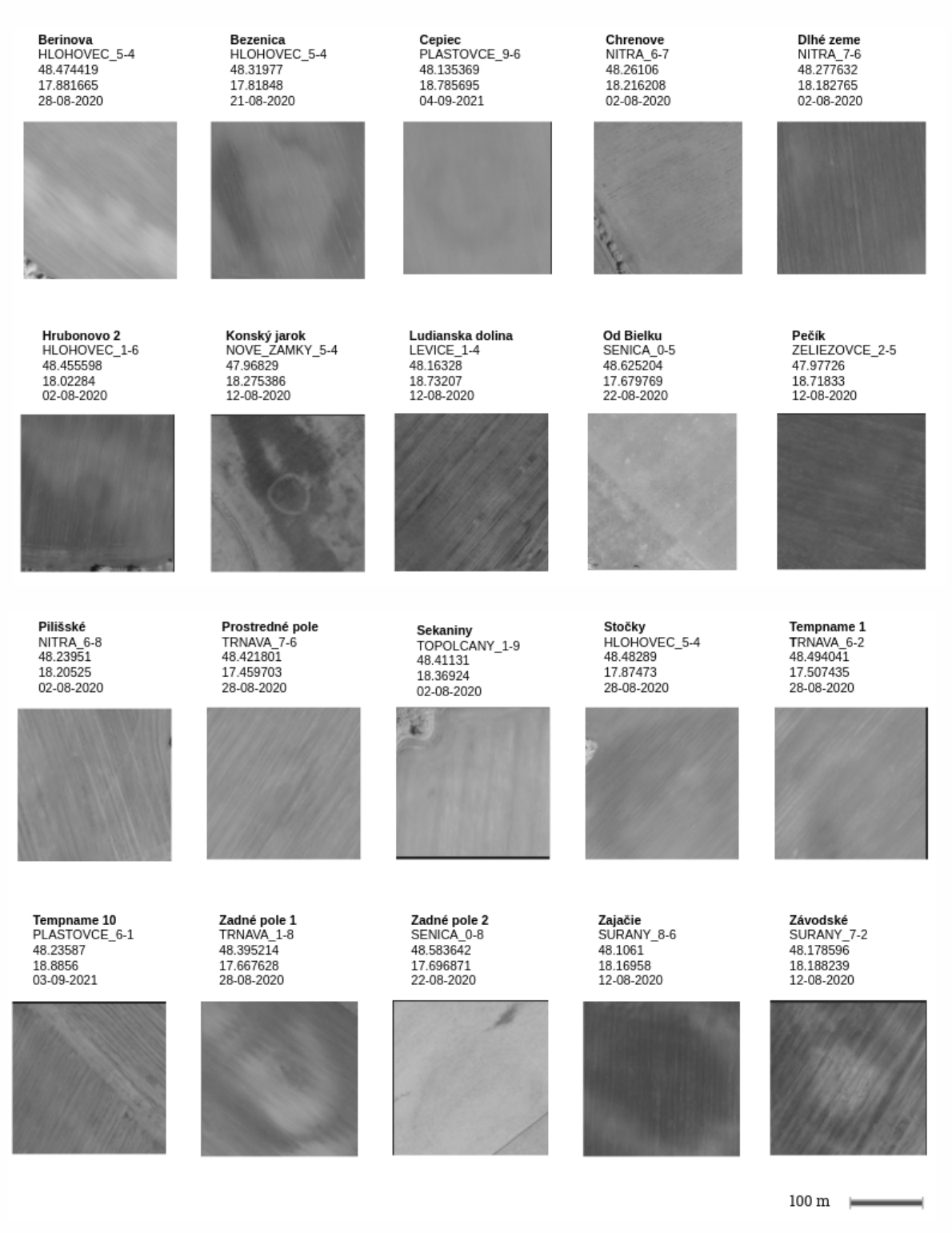

Figure 2.5: Rondels (with location names) visible in the 2nd cycle orthophoto, file name, WGS-84 latitude and longitude, and image date is included. Near-infrared (N) channel.

2.2. Convolutional neural network models

Convolutional neural networks (CNNs) are a type of layered deep learning model used for tasks such as image classification, segmentation, and object detection on images [6]. The model takes in image data and, after data transformations, outputs a prediction [7]. Dimension reducing transformations on inputs make CNNs in general, effective for usage with extracting patterns from large scale data sets [28].

During the training process, the more data we input into a model, the more its predictive abilities are expected to improve. Likewise, supervised models work better with more labeled data [25]. Our CNN models are supervised, meaning that training images contain a class label identifying whether or not they contain a rondel. After feeding in training data at each epoch, another set of unseen ground truth validation data is used for evaluation [11]. A model's effectiveness is measured through values from a loss function by comparing incorrectly or correctly classification results to labeled ground truth data. We are aiming for a model which minimizes the loss value since a divergence from the ground truth label increases the value [7] [19]. With each epoch, the model updates it's weights. We also want to hold out another data set called the test set that has not been used in the model evaluation process, to provide a final, unbiased view on the model's effectiveness.

Most CNNs consist of convolutional layers, pooling layers, and fully connected layers.

In convolution layers, filters performing matrix operations are applied across inputted images as sliding windows, or kernels, to extract relevant features. Common features extracted are patterns such as curves, lines, edges, or even whole shapes, depending on the side of the filter and kernel. The extracted features would get passed on to the next layers as a feature map. Pooling layers "summarize" the features found by reducing the dimension count of the feature map through aggregating functions. In the final connected layer, the features and their probabilities are flattened and passed to a prediction layer for output label generation [28].

3 tasks CNNs are commonly used for in image operations include: image classification, object detection, and object segmentation; shown in figure 2.6. Image classification is used in this project.

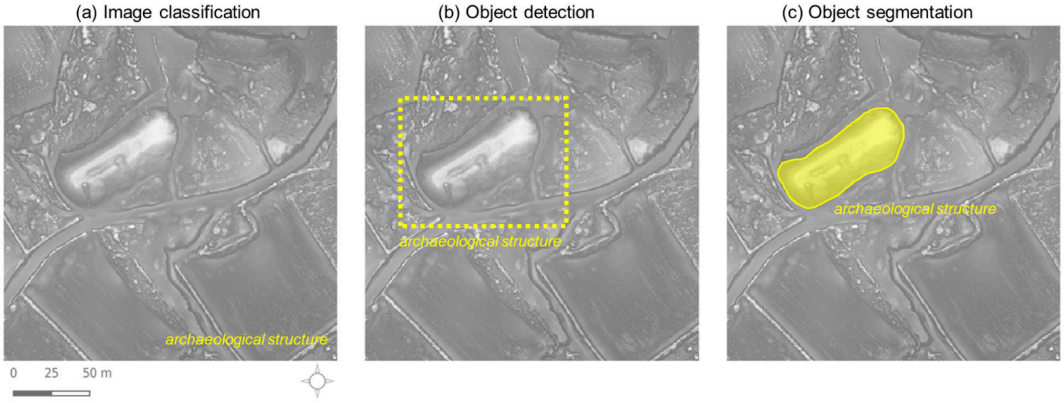

Figure 2.6: 3 tasks with CNN usage in archaeological research. (a) shows a digital elevation image classified as containing a Neolithic French burial mound. (b) shows the detected location of the structure with a bounding box on the image. (c) shows a delineated, or segmented burial mound. [2]

2.2.1. Object detection and segmentation

In an object detection study, the neural network outputs a located bounding box surrounding an object of interest belonging to a class within an image [2]. The advantage of object detection is that the output directly shows the object's location, but a drawback is that it takes time to draw bounding boxes. To train a supervised detection model, training samples must contain four coordinates pairs for drawing bounding boxes. With over a thousand training samples required to effectively train a CNN, we would be spending too much time; therefore we are not going to use object detection in this project [36].

Similarly, object segmentation outputs a located and delineated mask around an object of interest. As non-archaeologists, we have limited expertise on the details and layouts of rondels; additionally,

the rondel ditch could be covered by more or less vegetation depending on the season [3]. Creating delineated masks over rondel samples therefore becomes time consuming and may require input and expertise from actual archaeologists.

2.2.2. Image classification

In an image classification study, the neural network outputs a class associated with an input image [2]. An image containing a rondel has a 'rondel' class, similarly, an image without a rondel has a 'not a rondel' class. Image classification does not require bounding boxes or masks to be drawn around the object of interest (in our case, rondels) in training images, therefore this reduces time used to create training images for supervised models. We are using this method since the images can be rapidly generated without intervention to add in masks or boxes [36]. Large areas can be classified in smaller parts using the sliding window technique; for example, searching for smaller archaeological sites of interest on a larger map or a high-resolution aerial image [41] [12].

ResNet

A ResNet, or a residual neural network, is the name of a CNN model that we will be training for the project's purpose. Having a deep layered neural network results in improving target feature recognition-yet they comes at the expense of over-complexity, overfitting, and causes exploding gradients that could prevent converging during training. ResNets contain a 'skip-connection' at each residual block (or, a group of layers) to skip through multiple layers and prevent such issues that arise when layers are increasingly added [23]. The number behind ResNet indicate the amount of layers in the model. We will train a modified version of a ResNet-34 with a stride of 1 for it's moving filter to detect finer image details; we call this a simple ResNet.

ConvNeXt

Another classification model we will be training is the ConvNeXt. This model is 'inspired' by vision transformers and is said to have achieved better results than the ResNet. A 'tiny' version of the ConvNeXt comes with more parameters than a ResNet model with 49 convolutional layers (and a fully connected layer) [33]. The ConvNeXt splits input images into patches and classifies them individually with depth-wise convolution.

2.3. Class imbalance

We are performing a binary image classification with two classes: 'rondels' and 'not' rondels.

The entire country of Slovakia contains much more areas without rondels than there are areas with, therefore we are dealing with a large class imbalance in the data set [30]. When a neural network model is trained on imbalanced data, the model's weights are updated based on more gradients coming from the majority than the minority class, increasing the error coming from the minority class and results in worse classification performance. Other examples of imbalanced data include oil spill detection from satellite images, catching fraudulent transactions, and diagnosing rare medical cases [25] [30]. This is an intrinsic imbalance in the data where the imbalance of data is naturally occurring due to their low frequency. Oil spills are infrequent, most patients don't have rare disease diagnoses, and most monetary transactions are legitimate [31].

Our class imbalance can be partially solved regardless of class disproportion by equally representing both classes through sampling. In this case we can intentionally over-sample the 'rondels' class [30]. With a binary dataset with few positive class examples, only outputting an negative class classification means the model is accurate. We are reducing this problem on the data-level by reducing the imbalance through sampling more "rondels" by randomly generating synthetic training images of them [31].

2.4. Previous research

A ResNet-18 CNN model was previously trained specifically for searching rondels in digital terrain rasters from processed LiDAR data over western Slovakia [36]. The data set used was processed at a resolution of 1 meter per pixel and at that time, was only available over the western part of the country [21]. The model was trained on augmented versions of existing visible rondels pasted on terrain maps. It was then used to search for new rondels in the surroundings of an area called LOT5, northeast of Bratislava. 32 possible new rondels in figure 2.7 were found within LOT5.

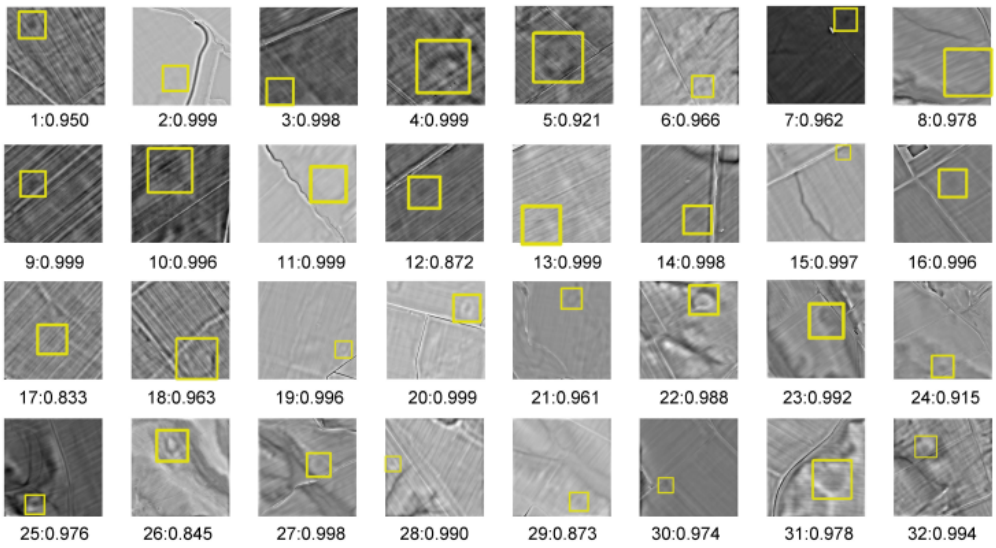

Figure 2.7: 32 new possible rondel sites and their probabilities in western Slovakia from a LiDAR data study [36].

Most of the new rondels shown tend to be located on fields and close to abandoned rivers [36]. It is theorized that rondels found on fields were not yet removed by urbanization, since rondels have been found in urbanized areas close to buildings, for example the rondel in Prague from figure 2.2 [20]. Another study using terrain data reported that most rondels were built on sloped areas [51].

A following archaeological project has shown how rondels appear in each channel and combination in multispectral aerial images- but without taking into account seasonal differences [42]. Multispectral RGBN images and their combinations containing archaeological sites are claimed to appear different seasonally in research papers; usually during drier months without heavy rainfall the sites can be more easily discerned [3]. Fall and spring are the months where less vegetation differences can be seen over other non-rondel buried archaeological sites in Greece and Turkey because of low plant growth and overall wetter conditions- an ideal time to carry out multispectral studies in Europe would be from July through September [27] [3].

3

Data and study area

This chapter details the specific data sets we'll use in this project. All data used are stored on a remote computer and accessed through a remote server.

3.1. Data sources

3.1.1. Orthophotomosaic of Slovakia - 2nd cycle

We are searching for rondels on the official second cycle orthophotomosaic of Slovakia. An orthophotomosaic is a digital aerial photograph created from multiple mosaiced photos, or orthophotos, where corrections were made for terrain reliefs, sensor and camera calibrations, and feature displacements [17]. A full version of the orthophotomosaic is shown in figure 3.1 below. This cycle's orthophoto was generated from aerial photos taken between 2020-2022 May to October and was released by Slovakia's cadastre in 2023 via Geoportal [21].

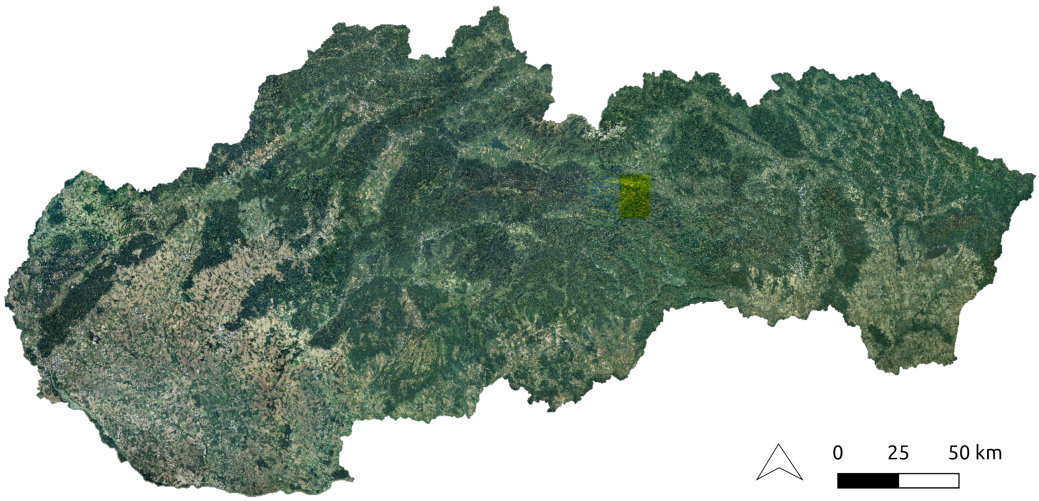

Figure 3.1: Full view of the 2nd cycle Slovak orthophotomosaic provided by Slovakia's cadastre [21].

The full orthophoto of Slovakia is split into 13,000 smaller rectangular 12,000 by 10,000 pixel tiled images (5 square kilometers per tile), totaling at around 2 terabytes in file size. 4 image channels (red, green, blue, near infrared) are available as a .TIFF raster alongside a .TFW sidecar file containing its projection and co-location [21]. The images have a ground sampling distance, or resolution, of 20 centimeters. A square pixel on the orthophoto represents 20 by 20 centimeters on the ground [34].

Every .tif image tile and it's .tfw sidecar file containing geolocated information is named using the following convention: AREA_##-##.tif or .tfw. Following digits indicate the tile's position in a named area of the country [21].

As of 2024, only the second orthophoto cycle is fully available over the entire country in all 4 channels.

Table 3.1 lists the full details on this cycle. The first cycle does not contain near-infrared image channels [21].

The data set comes with a .shp shapefile containing metadeta on the dates at which aerial photos used to generate the tiles were taken [21].

Orthophotomosaic properties		
	1st cycle	2nd cycle
Resolution	25 cm per pixel	20 cm per pixel
Channels	RGB	RGB + Near-infrared
Cycle duration	2017-2019 May 1 - September 30	2020-2022 May 1 - October 17
Format	.TIF images + .TFW sidecar	.TIF images + .TFW sidecar
Size per image tile	10000 x 8000 pixels	12500 x 10000 pixels
Coordinate system	S-JTSK(JTSK) - EPSG:5514	S-JTSK(JTSK) - EPSG:5514

Table 3.1: First and second orthophotomosaic properties covering Slovakia

3.1.2. Reference data: 1st cycle orthophotomosaic

RGB orthophotos (without the near-infrared channel) are available for the first 2017-2019 cycle over Slovakia. With the photos, we may be able to see slight seasonal vegetation changes in the visible spectrum because the images were taken between 1.5. - 30.9. during different years [21]. Slight vegetation differences between two cycles are show in figure 3.2.

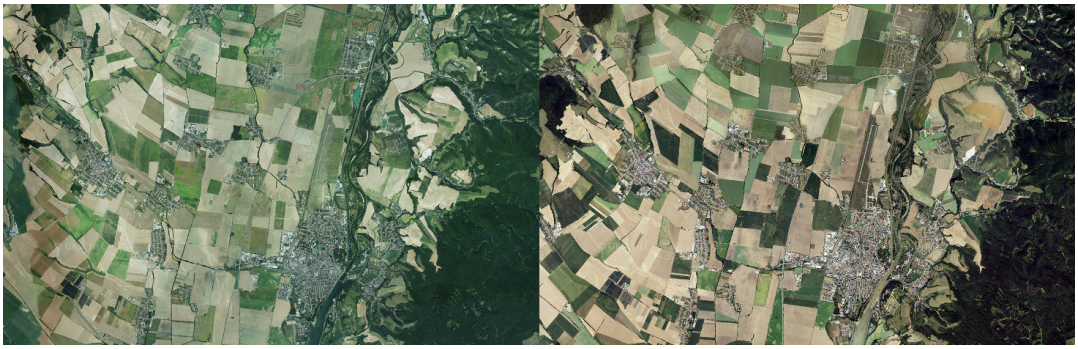

Figure 3.2: First cycle (left) and second cycle (right) images over a western Slovak village up close - this region was previoiusly discovered to have more detected rondels than in other areas in the country. [21] [36]

3.1.3. Rondel sketches

We are using rondel sketches from archaeological publications and textbooks to create new samples and address the inherent class imbalance in our data set [30]. The rondel sketches below in figure 3.3 are blueprints of their layouts based on publications recording rondel excavations and sightings in Austria, Slovakia, and Germany [4] [5] [48].

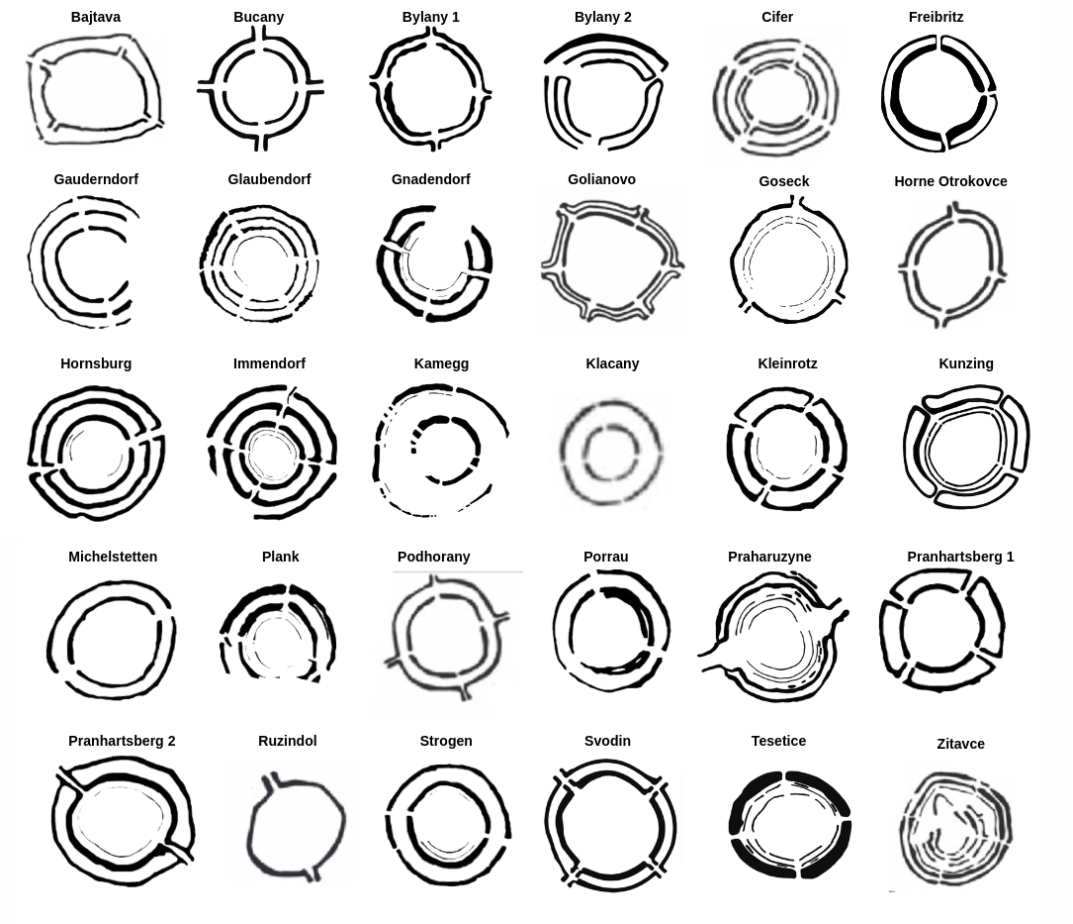

Figure 3.3: All rondel sketches used in this project from Austria, Slovakia, Czechia, and Germany [4] [5] [48]

3.1.4. CORINE land cover map

New training images are generated based on sketches and on empty agricultural land containing no rondels. The majority of remaining rondels, visible or not, are found on agricultural fields [36][5]. The CORINE 2018 land cover map made available by EU Copernicus' land monitoring service confirms this in figure 3.4 below where agricultural land cover is colored in light green. 2018 was the year with the latest land cover data update provided [44].

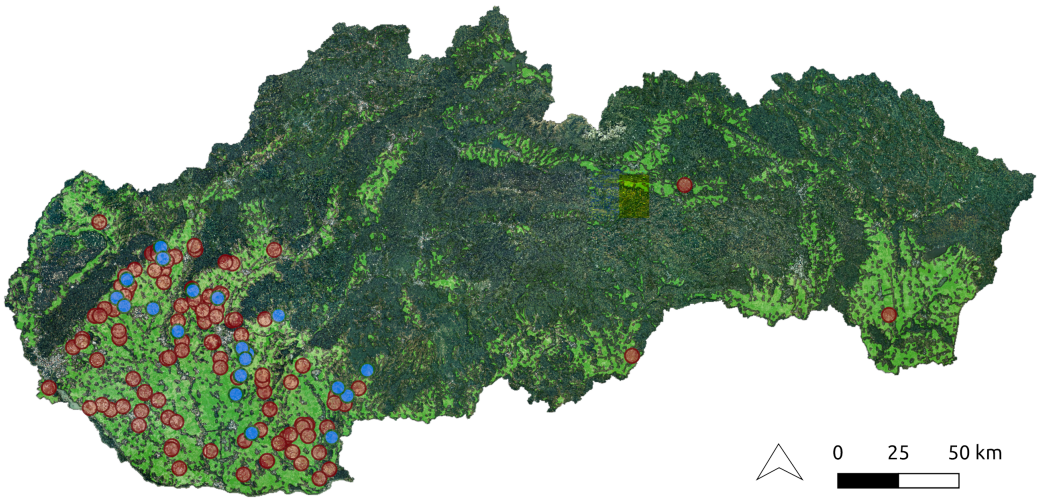

Figure 3.4: A CORINE shapefile (in green) over the 2nd cycle Slovak orthophotomosaic with reported rondel locations [21].

The land cover map is supplied as a raster and as a vector shapefile containing land cover type for each EU country. Land cover types can range from agricultural areas to water bodies, forests, industrial and commercial areas [44]. The shapefile is used in Python's GDAL package and in the QGIS software to constrain areas where generated rondels are set to appear on only fields. It is also used later on in the project to narrow down areas to search for rondels in order to limit data storage and runtime [41].

4 Methodology

This chapter lists out all steps required to carry out this thesis project. Figure 4.1 contains a visual overview.

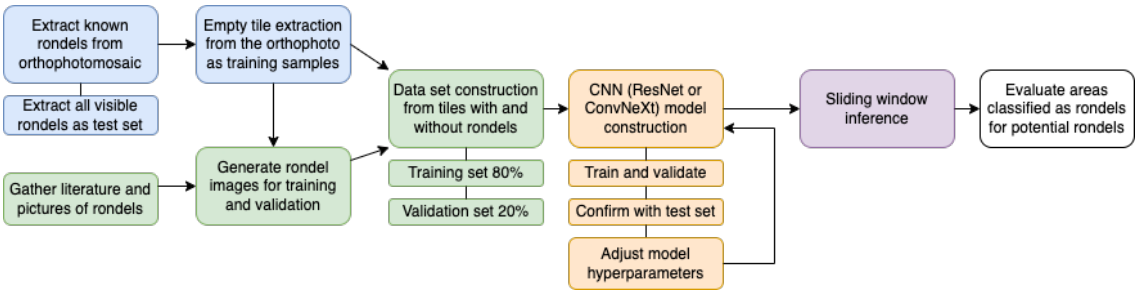

Figure 4.1: Flowchart showing the methodology, step by step, of this project.

4.1. Training, validation, and test set folder construction

To run a binary classification of areas which either contains a rondel or none at all, a CNN model must first be trained on images split into separate folders. Images containing rondels are positive examples, images without rondels are negative examples. There are only 20 clear rondel examples from the Slovak orthophotos in comparison to thousands of areas without rondels. This imbalanced data set is not enough to train a neural network model, but this issue can be resolved by using synthetic rondel examples [30].

Our data set consists of a training, validation, and test set with a split shown in figure 4.2. The training set is what we feed to train the model before using it as a classifier, while the validation set is the one we use to evaluate the results of the model at each epoch using the training set. We will use approximately a 8:2 training to testing data set ratio since it is the most commonly used ratio with resulting in the best model results empirically [26] [22].

Data in each set should be independent of each other so that the model does not evaluate the validation set based on data it has already seen; therefore, for synthetic rondels, rondels generated from the same sketches should be grouped together in either set, but never in both. For both sets we chose to use a balanced 1:1 number of positive and negative examples [26].

A test set consisting of all 20 real, visible rondels and 625 empty areas is held out separately for the final evaluation step. No real rondel examples are used in the training or validation stages to prevent overfitting.

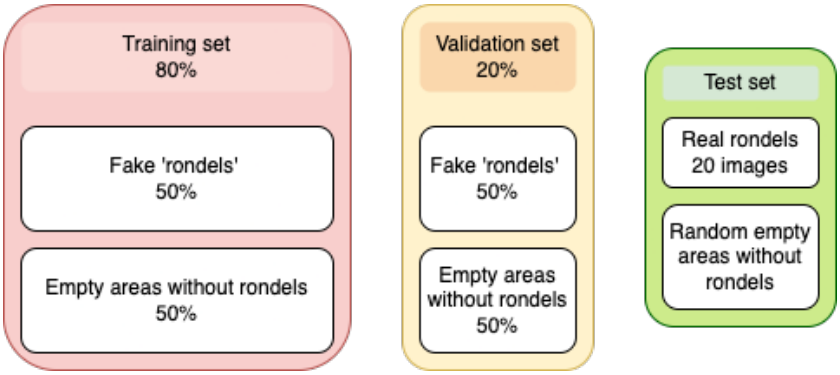

Figure 4.2: Training, validation, and test set split.

4.2. RGBN image extraction

Rondels appear as dark and faint circles on fields and on all 4 image bands because ditches cause vegetation growing above them to retain more moisture [46]. Most rondels have been discovered on agricultural fields in Central and Eastern Europe [5]. Crops growing over rondel ditches and images show a relief difference formed where vegetation overlays near-surface archaeological remains. In this case, archaeological features can retain soil moisture with a different percentage of moisture compared to non-archaeological areas. [1].

A script was written to extract square images and numpy arrays from RGBN bands centered around any given S-JTSK or WGS-84 formatted coordinate located within the orthophotomosaic dataset. We use the extracted images to construct our data set. An example of extracted visible rondels (and their locations plus tile names) the RGB and N bands was shown previously in figures 2.4 and 2.5. Additionally, these rondel images make up the 20 real examples in our test set.

4.3. Synthetic training sample generation

20 out of all recorded rondels are visible in the Slovak orthophoto. A lack of positive rondel examples makes it difficult to construct an adequate training and validation set for CNN models. To compensate our imbalanced data set, we sampled our own synthetic training images resembling the appearance of existing rondels on RGBN bands; otherwise, using an imbalanced data set can adversely affect the classifier's results [25]. The generated rondels were created using sketches of Austrian, German, and Slovak rondels from archaeological textbooks. Few rondels in Austria appear to have triple concentric rings [5] [48].

Our randomly generated rondel images are made to closely match the appearances of vegetation with more moisture retention growing in a circular crop mark pattern on top of arranged ditches on fields [8] [46].

First, rondel sketches were individually cropped out of the textbook and edited to appear as greyscale sketches on a transparent alpha channel. An example of an cropped image is shown in figure 4.3.

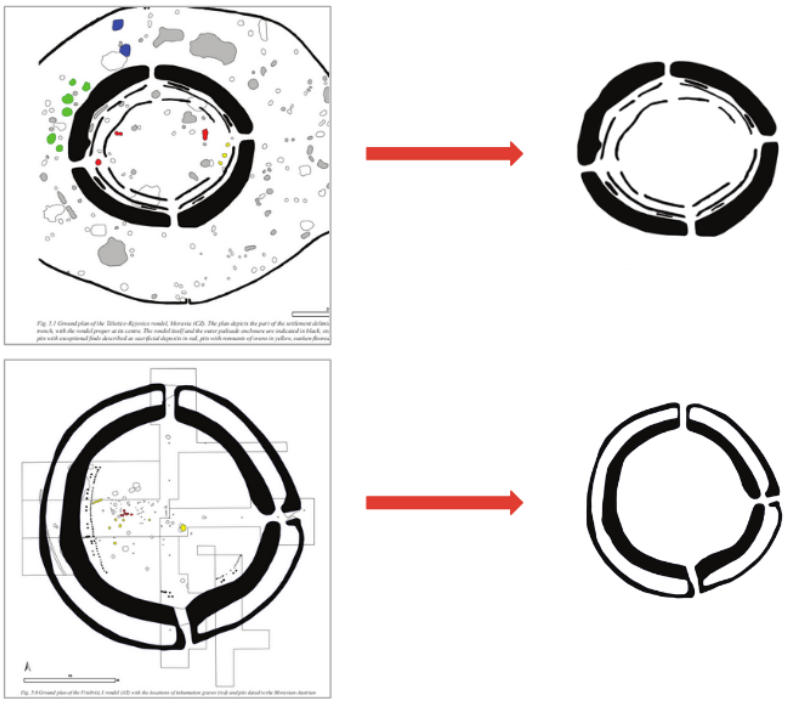

Figure 4.3: A rondel in Slovakia edited and cropped out of a Slovak archaeology textbook authored by Ridky et al. [5]

The cropped sketches are then randomly and automatically modified according to figure 4.4 using image editing functions in openCV and sklearn to increase training sample variety while still retaining the characteristics of a rondel [36]. An increase in training sample variety and augmentation increases the trained CNN's performance [10].

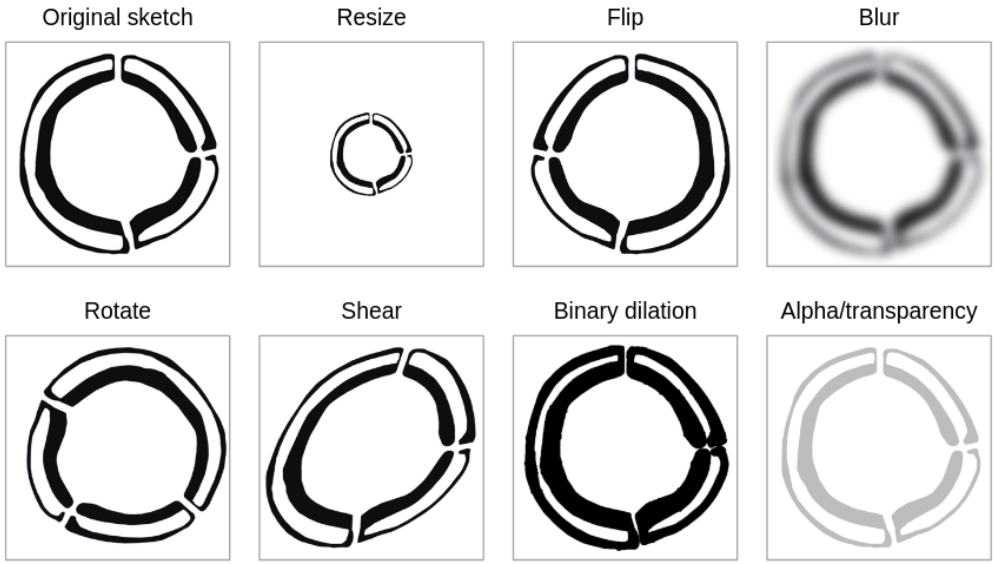

Figure 4.4: Randomly and automatically augmenting a rondel sketch with random openCV and sklearn image editing functions.

Next, a specifiable number of randomly geolocated images are cropped out of the orthophoto if the ground below has a land cover type indicating agricultural activity. Areas containing known rondels (to our knowledge) within half a kilometer are excluded from the process. Since all of our visible rondels fall within the boundaries of an agricultural area as delineated by the CORINE 2018 land cover map; picking random agriculture fields to place generated rondels on should give a representative sample. The modified rondel sketch is pasted onto the cropped area on a random location of the image on all 4 (RGBN) channels like in figure 4.5.

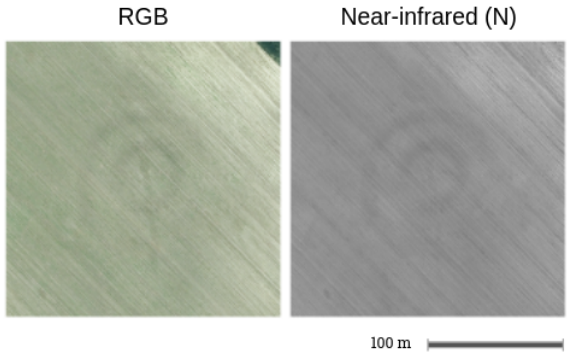

Figure 4.5: A modified rondel stamped on a random field on RGB and N channels.

We inserted the modified sketches on random locations on extracted fields since CNNs are reported to not be shift invariant, therefore a difference in the location of an object of interest affects CNN performance to a significant extent [9]. Training and validation samples resemble pictures in figure 4.6. RGB channels are shown for easier visualization.

Figure 4.6: Examples of edited rondels on empty fields at different locations on each extracted field. RGB channels are shown.

Negative examples of the 'not a rondel' class are extracted randomly from all land cover types over the country. This includes infrastructure, villages, water bodies, and forests where rondels are not typically found. Figure 4.7 shows several examples of images containing no rondels, generated or otherwise.

Figure 4.7: Examples of images containing no rondels for the 'not a rondel' class. RGB channels are shown.

The images are pre-processed into numpy arrays and to a normalized data type called tensors for use in the Pytorch machine learning package. Every sample is manually inspected before being fed into the model. Generated training images labeled in the 'rondel' class showing no clear rondels are discarded. Horizontal and vertical flips are also applied to the images for further augmentation to increase training sample variability and rondel placements.

4.4. Hardware

To train and evaluate our models, and to search for rondels, we used a remote machine accessed through a remote server with 4 GPUs and 12 CPU cores. 2 NVIDIA GeForce RTX 2080Ti and 2 NVIDIA RTX A4000 GPUs were used. All data is accessed through the remote server. To speed up training and search times with data parallelism, Pytorch utilizes all 4 GPUs by splitting loaded data into individual GPUs through the DataParallel function. We increased CPU memory usage and training speed further with a higher num_workers value up to the number of CPU cores during the data loading process [29].

4.5. Convolutional neural network training

4.5.1. Model setup

We are using the Pytorch library to set up a modified simple ResNet-34 and a modified ConvNeXt tiny classifier for 4 channel multispectral image inputs. The expected output is a binary classification of an input image. The models take in tensors; matrices containing elements of a single data type with a square input size of 512 by 512 pixels and 4 channels (RGBN). In our models, the input data type is float32. All square multispectral images are converted to multichannel (4) numpy arrays, then normalized and rescaled to the input size to train the classifiers. Bilinear interpolation is first used to rescale training images. Then the multi-channel numpy arrays are converted to float32 tensors.

Each image channel is normalized from 'color' values from [0-255] in their original RGBN format to tensor values between [0-1] and with a mean and standard deviation of [0.5, 0.5, 0.5, 0.5] before being fed to the neural network model for training or classification. Data normalization also speeds up the

training process [24]. An example a normalized image in comparison to it's original image is shown in figure 4.8.

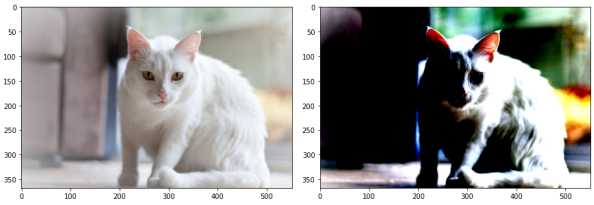

Figure 4.8: A cat image to the left before normalization on all channels, and a normalized version of the same image to the right [47].

A softmax layer at the end of the classifier produces a decimal output containing the probabilities of each class the image is is given an input [13], with all class probabilities adding up to 1. The softmax's layer output is used to quantitatively evaluate a model at each epoch with the test and training sets through the cross-entropy loss function [19]. Figure 4.9 is an example of a softmax probability output with 3 classes.

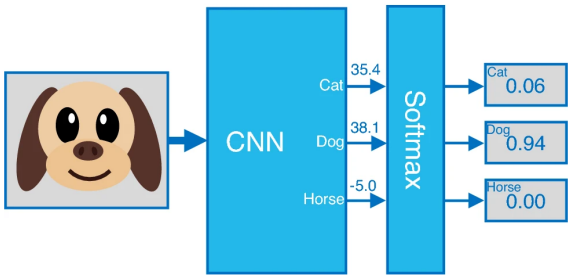

Figure 4.9: The softmax layer outputs the probability of an inference output belonging to one of three classes [13]. The input picture has a 0.94 probability it belongs to the 'Dog' class.

After training, a generated .pth file containing model weights is used to run the trained model. With a trained model and its corresponding .pth file on hand, the model can be set to inference mode to classify input images and output classification probabilities. Images are classified as belonging to the 'rondel' or the 'not a rondel' classes when their softmax probabilities for the respective class is above 0.5 [28].

4.5.2. Hyperparameter tuning

Before training each ResNet or ConvNeXt model, we can configure hyperparameters to optimize their training times and performances, and to minimize test and validation loss from the cross-entropy loss function [11].

Epoch: The number of times the model updates its weights based on trained data passed through it [11]. We initially train models with different hyperparameters for 20 epochs to see their perforamnces and validation loss values, then for the best model, train up to 200 epochs. At each epoch the model iterates over the training set to update weights, and over the validation set to check it's performance [39].

Batch size: The number of training samples used in each training epoch before the model updates its weights [11]. Batches in powers of 2 work well on hardware [12]. We're using batch sizes of 8, 16, 32, 64.

Learning rate: We use values of 0.01, 0.001, 0.0001. The learning rate is the rate at which the model adjusts it's parameters at each epoch [11]. Models with low learning rates risk not updating weights enough at each epoch while having high learning rates can result in the model going too far with weight updates and overfitting.

Optimizer: We're using AdamW as our optimizer. AdamW is a modified version of the Adam optimizer which was claimed to work better than both Adam and SGD optimizers when it comes to classification tasks [35].

Weight decay (L2 regularization): This is a regularization technique used to penalize weight updates in a neural network. Smaller weight values penalize weight updates less and the model is prone to overfitting, while the inverse is true with larger weights and underfitting [35]. The default weight decay for AdamW in Pytorch is 0.01. We're using values of 0.1, 0.01, 0.00001, 0.000001.

4.6. Model selection

We want to chose the most promising model we could train; preferably one with the lowest validation loss and a low false positive rate for our rondel search. This particular model will be discussed in detail in the following results chapter.

4.6.1. Cross-entropy loss function

To evaluate how well a model classifies images based on its true labeled class, we're using the cross-entropy loss function at each epoch on the validation set. This is an error log function measuring the performance of a model using its classification output is a probability of each class between 0 and 1 with possible values shown in figure 4.10. The value it outputs is higher as the classified output probability diverges from the true label, therefore, more promising models incur low validation loss values [19]. At each epoch, we calculate the validation set's loss for each training run's first 20 epochs and use it for evaluation before training further for 200 epochs [11].

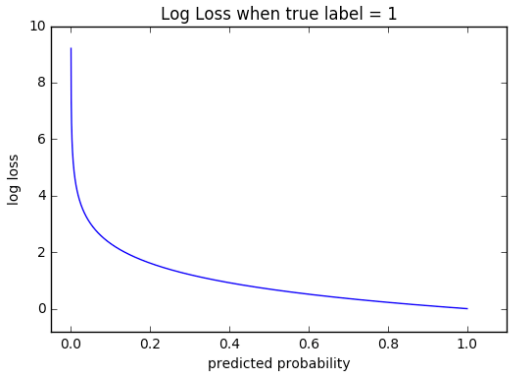

Figure 4.10: Log-loss values from the cross-entropy loss function. Loss values increase as the output probability is further from 1, the ground truth class. Values decrease as the probability is closer to 1 [19].

4.6.2. Test set performance

A confusion matrix is a table resembling figure 4.11 which we are using to evaluate the CNN model's binary classification performance. To save time sifting through the output for field patterns resembling rondels, we want a low false positive rate while maintaining a sufficient true positive rate to detect a few obvious examples [36]. The model's classification result for each image is compared quantitatively to the image's true labeled class.

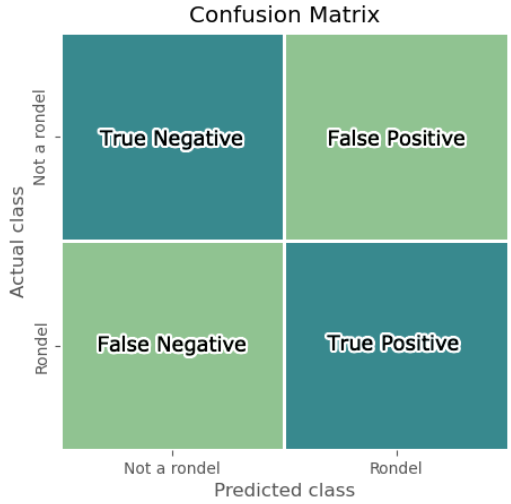

Figure 4.11: An example of a confusion matrix showing classification results versus ground truth classes. [19]

There are four types of results for binary classification as follows:

True positive: The model correctly classifies an image containing a rondel pattern as such.

True negative: An image without a rondel pattern is correctly classified as not containing a rondel.

False positive: Image is classified as containing a rondel pattern when it is labeled as not containing one.

False negative: An image with a rondel pattern is misclassified as not containing one.

As stated previously, we want a low number of false positive classifications to save time searching for rondels. It is more valuable to discover at least one item of the rare class (rondels), than it is to run into too many false positives of the more common class (not a rondel) [49], and finding more false positives is costlier time-wise both for us and for archaeologists.

4.7. Sliding window inference

Although our goal is to search for more rondels in the country with a sliding window, we will not search through the entire orthophoto dataset due to storage constraints and hardware limitations, since sliding windows are computationally expensive [12]. Our hardware limitations increases the time it takes to read orthophoto tiles from the .TIF format. Therefore, each search tile must first be preprocessed and saved as a numpy array before we can run an inference to reduce runtime. Without preprocessing, reading in a tile takes longer than the classication process.

And to save storage space, we chose to search only through tiles in areas near rondels and exclude tiles which are known to contain rondels (visible or otherwise). All search tiles are over agricultural land cover from the CORINE data set [44]. The search area is shown in the following figure, figure 4.12, with a comparison to a previous LiDAR research project at cosine [36].

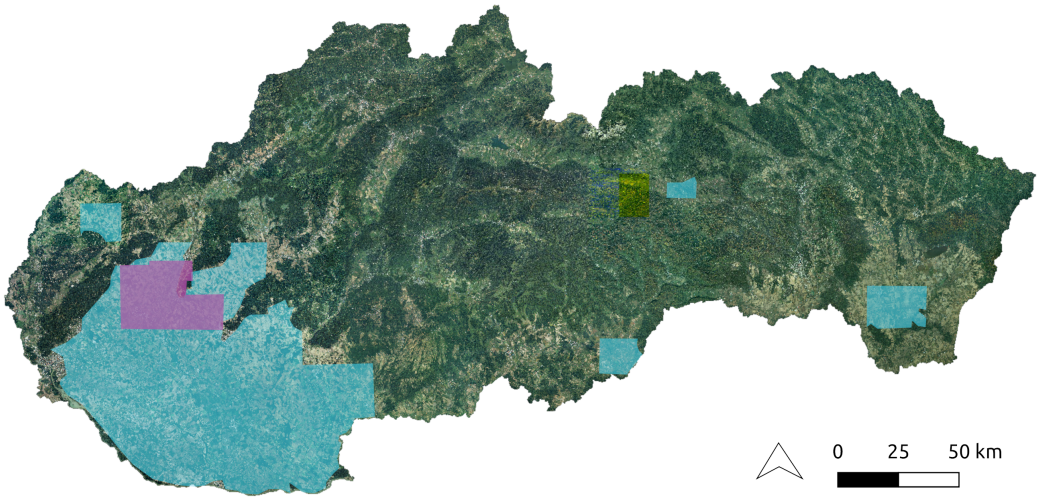

Figure 4.12: Our search area colored in blue over orthophotomosaic tiles. Previous search area, LOT5, from a LiDAR search in purple [36].

The selected model is set to inference mode for classification then it is used to search through orthophotomosaic tiles for rondels with the sliding window method, which has successfully been used in archaeological research to search for ancient Nazca lines in Peru from aerial RGB imagery [41]. Each tile in the data set is split up into 456 smaller patches of 512 by 512 pixels with an overlap of 256 pixels between each patch (or, half a patch), in case a rondel lies between two patches. The patches are created with PyTorch's nn.Unfold function and are subsequently inputted into the model for classification.

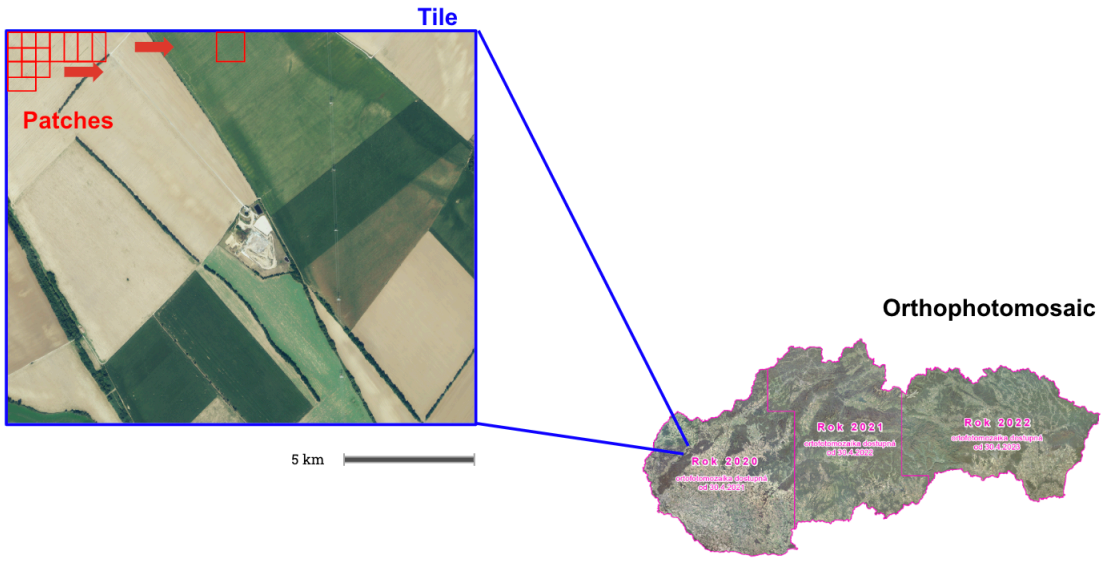

Figure 4.13: A sliding window classifier moving through an orthophotomosaic tile, classifying each patch by patch.

We then obtain a list of tiles and the patch location of a suspected rondel where the model encounters a patch with a high softmax probability output of the 'rondel' class. After manually inspecting the highest probability patches, we create a list of potential rondels [49].

5
Results and Discussion

This section discusses the final generated data set for model training, trained model evaluations and the selected model which shows the promising classification performance out of all evaluated models. Furthermore, this section discusses the results of the test set classification and gives an overview of the rondel search process with the selected model.

5.1. Final data set size

The final data set consists of RGBN channels of 2466 synthetic rondel images and 2466 images without synthetic rondels as our training set. Figure 5.1 contains the number of exact images used. Our validation set consists of 417 synthetic rondel images and 417 images without rondels; roughly equating to a 8:2 split. 20 real examples of rondels were held out (seen in figure 2.4 and 2.5) with 625 random images without rondels as out test set for evaluating the most promising models.

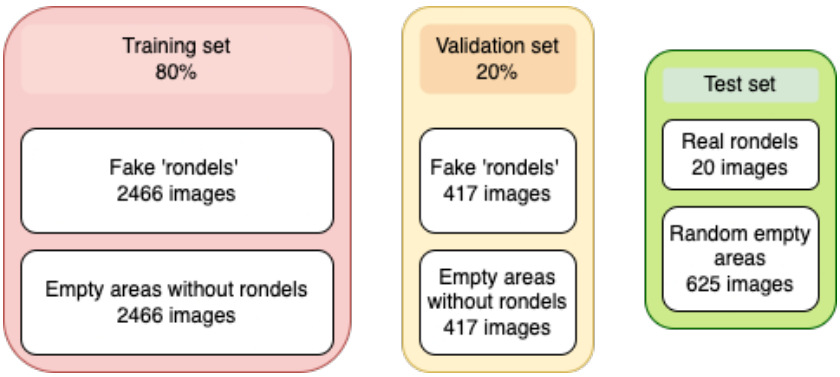

Figure 5.1: Training, validation, and test set final data set sizes.

5.2. Evaluating and selecting a trained model

We trained multiple models with differing hyperparameters and used the cross-entropy loss function after each epoch to evaluate the validation set's loss determined based on the differences between correctly or incorrectly classified images from the validation set. A model's lower loss value from the cross-entropy loss function indicates that there were fewer incorrectly classified samples [19]. 20 epochs of a simple ResNet model was trained initially with different hyperparameters. A ConvNeXt model (reported to result in better classification) was also trained to check whether performance improves with more parameters or otherwise [33].

5.2.1. ResNet evaluation

Validation loss

At 20 epochs, the modified ResNet-34 model with hyperparameters resulting in the lowest validation loss values at most epochs has a batch size of 32, an AdamW optimizer, a learning rate of 0.001, and the default weight decay value of 0.1. This model shows the most promising performance and

we chose to train this model longer for more epochs. Figures 5.2, 5.3, and 5.4 shows all the different hyperparameter combinations tested.

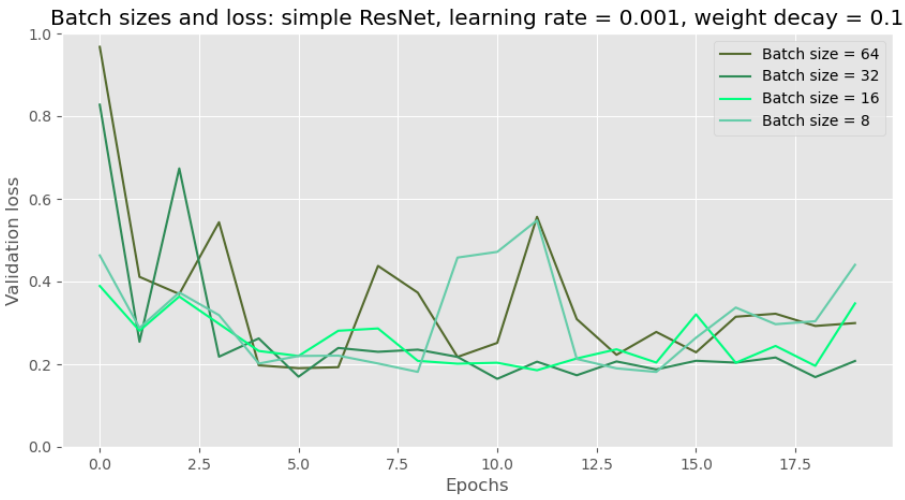

Figure 5.2: Batch size variations and validation loss.

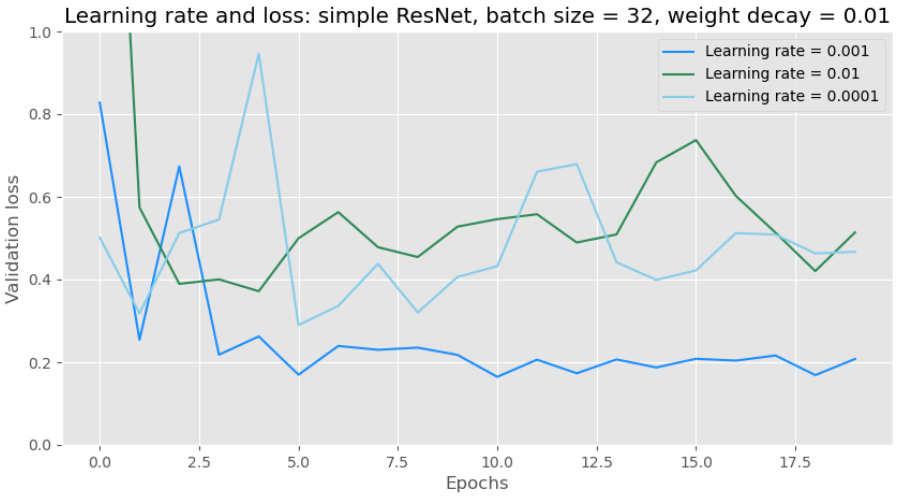

Figure 5.3: Learning rate and validation loss.

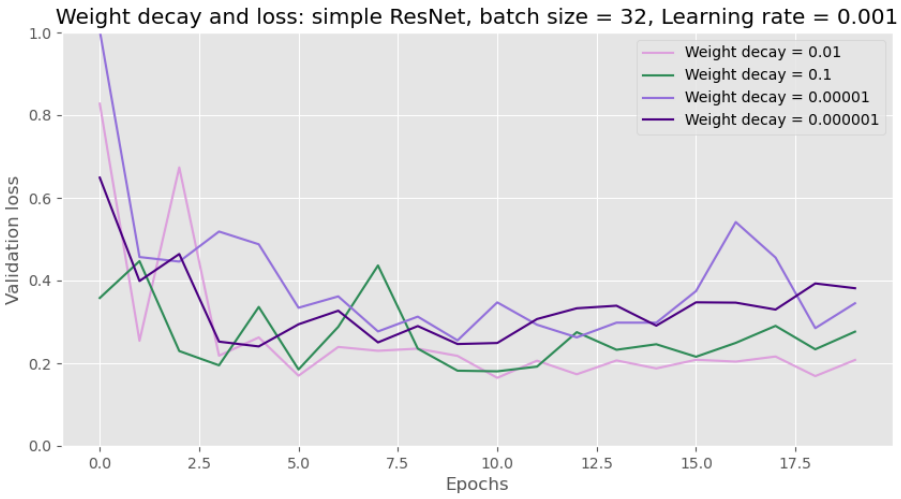

Figure 5.4: AdamW weight decay and validation loss.

Test set

We ran our test set's classification using the first 20 trained epochs for an initial check. Figure 5.5 shows the results as a confusion matrix. More than half of our rondel images were correctly classified, and a significant majority of images without rondels were also correctly classified. 52 negative examples were misclassified as the 'rondel' class. We have trained this model further to see whether this false positive number would decrease, since more false positive classifications are costly time-wise for archaeologists.

Figure 5.5: Confusion matrix for the simple ResNet model with 20 epochs.

573 out of 625 images without rondels were correctly classified as not belonging to the rondel class (figure 5.5). The distribution of the test set's softmax probabilities are imbalanced, falling on either the 0 or 1 end of the histogram.

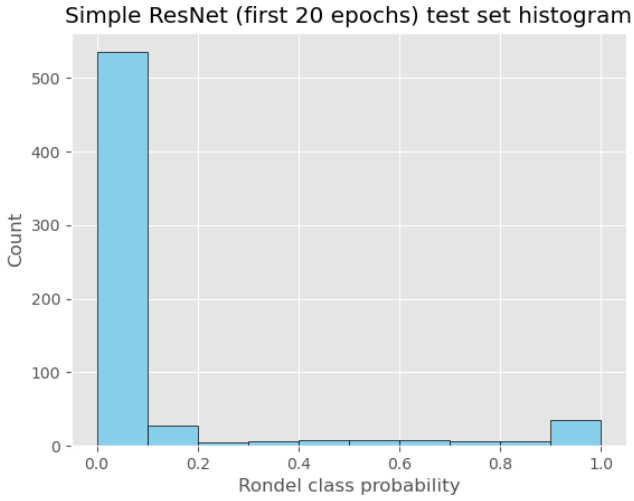

Figure 5.6: Probability histogram of images with the rondel class. Classifier: ResNet model with 20 epochs.

5.2.2. ConvNeXt evaluation

Similarly, we also tested a newer ConvNeXt 'tiny' classification model using the hyperparameters resulting in the lowest validation losses from our previous simple ResNet model. The ConvNeXt 'tiny' model has 28 million parameters, takes longer to train per epoch, and uses more GPU resources overall. Although the original model's publication claims that a ConvNeXt is better at classification tasks than a ResNet with similar number of parameters [33], our validation loss and classification results shows that a more complex model with more parameters does not lead to better performance.

Validation loss

With identical hyperparameters, the ConvNeXt tiny model did not classify images as well as the simple ResNet according to the validation losses in figure 5.7, contrary to published claims [33].

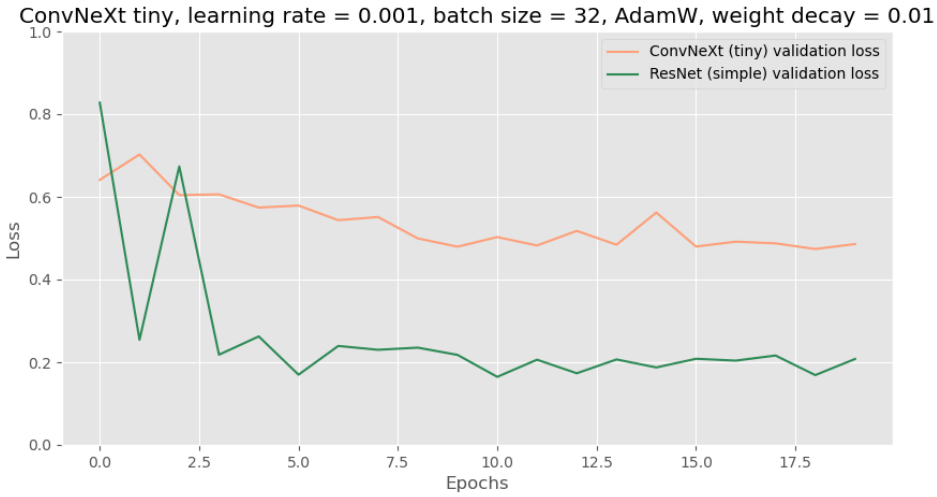

Figure 5.7: ConvNeXt tiny validation and training loss graph.

Test set

We can also perform an identical test set classification using the first 20 trained epochs but with a ConvNeXt model. The results are in figure 5.8. Less than half (239 out of 625)of our images without rondels

were correctly classified. Over half (386) were incorrectly classified as containing a rondel, indicating a high false positive count. We want to reduce the number of false positives during classification to decrease manual inspection time, therefore this model is not good for our objective. All 20 images containing a rondel, however, were correctly classified without any false negatives.

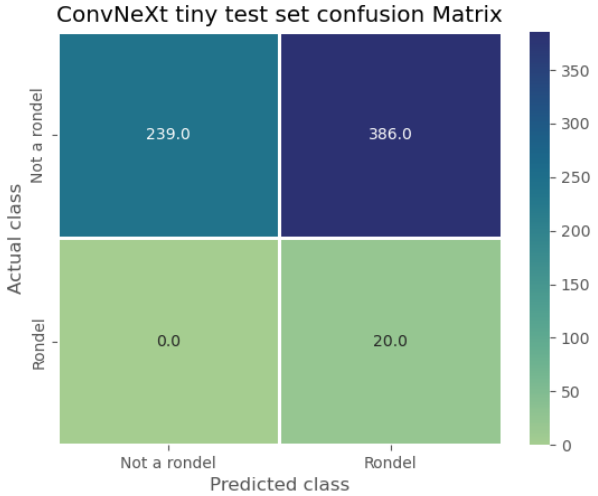

Figure 5.8: Confusion matrix for the ConvNeXt tiny model with 20 epochs.

Most classified images have a lower (less than than 0.9-1) probability of belonging to the 'rondels' class with the ConvNeXt classifier (figure 5.8). There is a higher spread in the classification's softmax probabilities, but there are also more empty images incorrectly classified as rondels than before with the simple ResNet model.

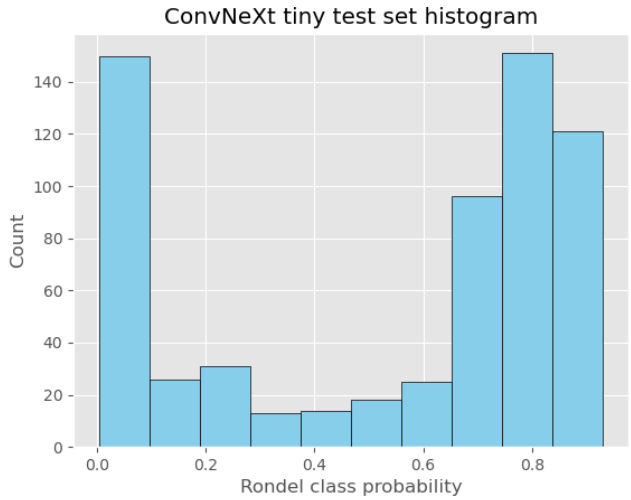

Figure 5.9: Probability histogram of images with the rondel class. Classifier: ConvNeXt 'tiny' model with 20 epochs.

5.3. Final ResNet model evaluation

The selected model with the lowest validation loss and low false positive classifications is a simple ResNet with a batch size of 32, learning rate of 0.001 and an AdamW optimizer with a weight decay of 0.01. We chose this model to train further for 200 epochs.

Figure 5.10 below is the training and validation loss and accuracy plot for the final selected model

trained up to 200 epochs. The graph indicates that this model begins to overfit starting from the 25th epoch as the validation loss increases up to twofold, while training loss continues to decrease [15]. Fluctuating loss values in the first 50 epochs are likely due to the high variations of augmented and randomized rondel appearances over different fields in the training dataset [16].

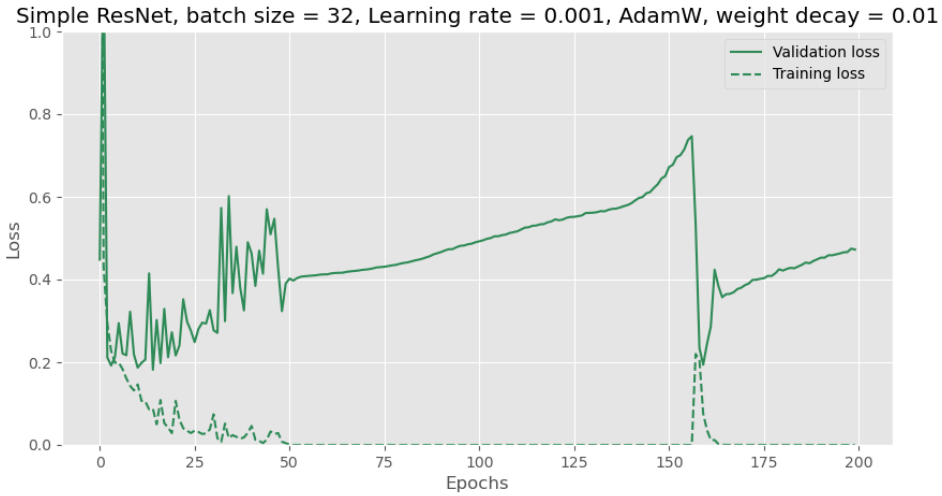

Figure 5.10: Simple ResNet model training loss and evaluation loss graph over 200 epochs.

Figure 5.11 shows a confusion matrix from classifying the test set with the selected model trained up to 200 epochs. This model trained for more epochs resulted in a lower false positive count (18 out of 625) and most images without a rondel were correctly classified as 'not a rondel'. Only 18 out of 625 empty images were incorrectly labeled as containing a rondel. Yet, 7 existing rondel images were correctly classified. Keeping the false positive count low is important when a high count can cost us more time, which we're trying to minimize through an automatic searching method, therefore a false negative classification is preferred over a false positive in this study and this model is a good choice.

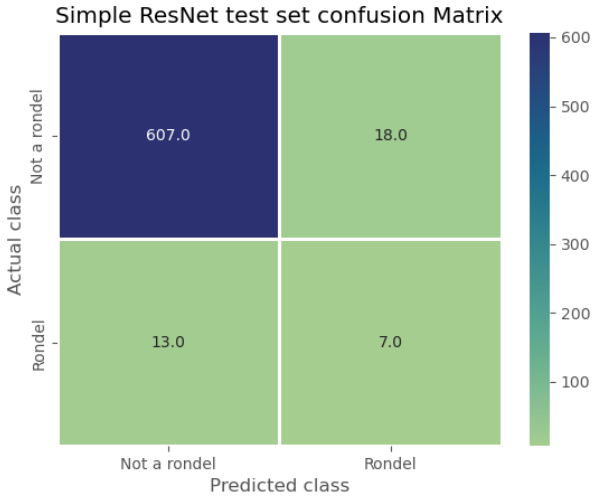

Figure 5.11: Selected ResNet model's confusion matrix from our test set classification.

After training the model up to 200 epochs, a majority of test set images were correctly classified as not containing a rondel (figure 5.12). There are now lower false positive classifications, but this results in less true positives classifications.

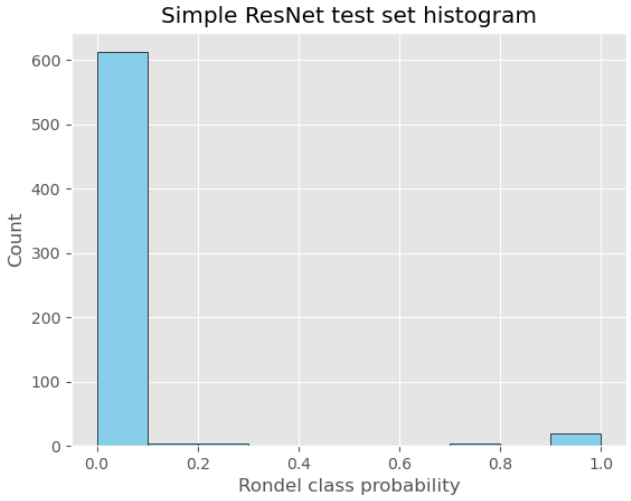

Figure 5.12: Probability histogram of images with the rondel class. Classifier: ResNet model with 200 epochs.

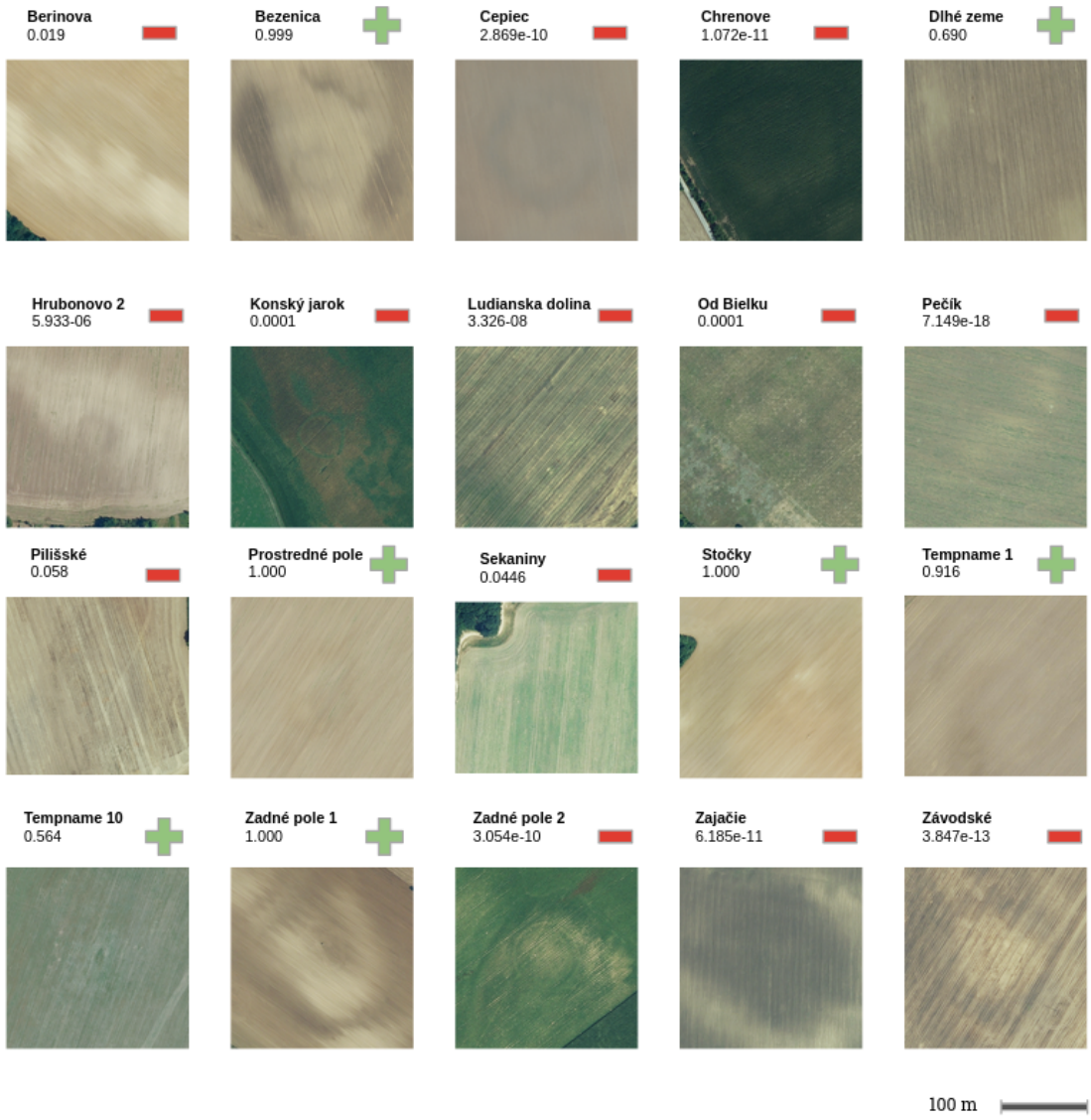

Figure 5.13: 20 visible rondels and their classified results alongside a softmax probability output for the 'rondel' class.

5.4. Comparing results from the generated training set

To compare the effectiveness of using a generated training set of rondels, we trained a model with identical parameters to the selected model using only training images consisting of the 20 real visible rondel images from our test set. The validation loss is perpetually much higher when the model is trained with only 20 images (figure 5.14) indicating that a generated training set is required for better classification results in this study.

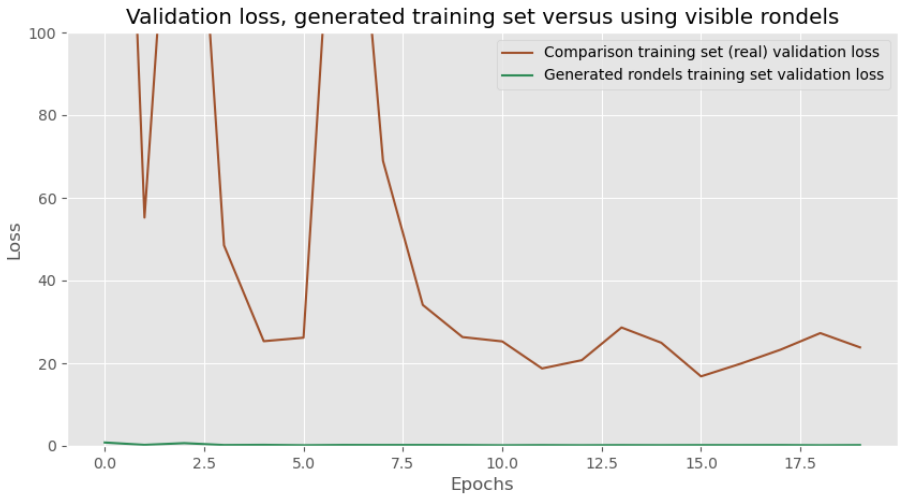

Figure 5.14: Validation loss from training the simple ResNet model on only our visible rondels (batch size = 20, the entire data set) versus on the generated training set.

Interestingly, upon classifying the test set with this model, all but one test set image were correctly classified as seen in the confusion matrix in figure 5.15. This does not equate to the model being near-perfect. Despite the test set performing well with classification, this model does not generalize well to classifying unseen rondels beyond the 20 existing rondel images it was trained on. This explains the large discrepancy in model performances from the validation loss between the model trained on generated rondel examples versus real visible rondels which are those in the test set.

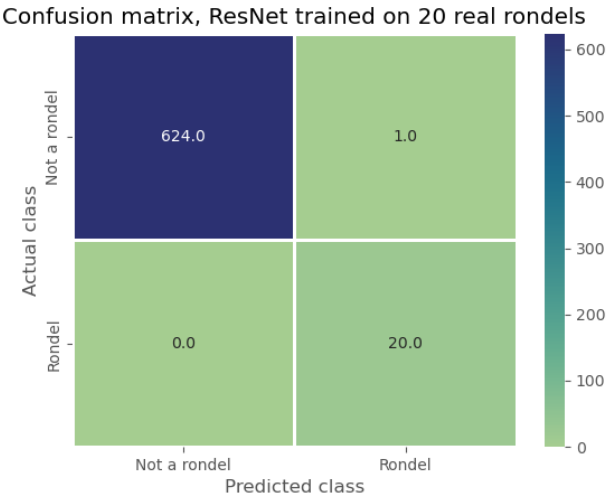

Figure 5.15: Confusion matrix for a simple ResNet model trained using only visible rondel images.

5.5. Rondel search and inference

The sliding window method resulted in 426 patches classified per tile with the CNN. A total of 1902 tiles covering 9510 square kilometers, and around 1 terabyte of data was preprocessed then classified.

This resulted in a much higher false positive count than expected; but overlapping patches may have contributed to similar scenes being classified consecutively. Probabilities are high for patches classified as the positive rondel class. Overall, probabilities fall either on the far 0 or 1 end of the histogram, with all 1902 tiles containing at least a 0.5 up to 1 probability for the rondel class, signifying that the classifications are rather confident, resembling the histograms of test set probabilities. We only looked at tiles with patches containing the probability '1' for a rondel, given histogram distributions for our test set.

Some tiles may contain more agricultural areas than others, and puddles or pools were misclassified. Small, but darker areas of grass in villages or near were also misclassified as rondels.

Tiles containing the highest count of positive patches have a count of 65, 68, 56, 48, and 45. We've looked at these tiles for possible rondels or false positives. There are 167 tiles containing only 1 patch within classified as containing a rondel, which were inspected. We also inspected tiles containing 2 to 6 patches classified as such to expand search options.

Once we extracted RGBN images from the above 3 groups of positive tiles, we found 7 possible sites to further study in detail. 2 (sites 1 and 3) out of the 7 new suspected rondels sites from our list were found with their locations mapped in figure 5.16. These 2 rondels display a clearer circular pattern. Close up images and an in-depth review of their characteristics are found in figures in section 5.4.1.

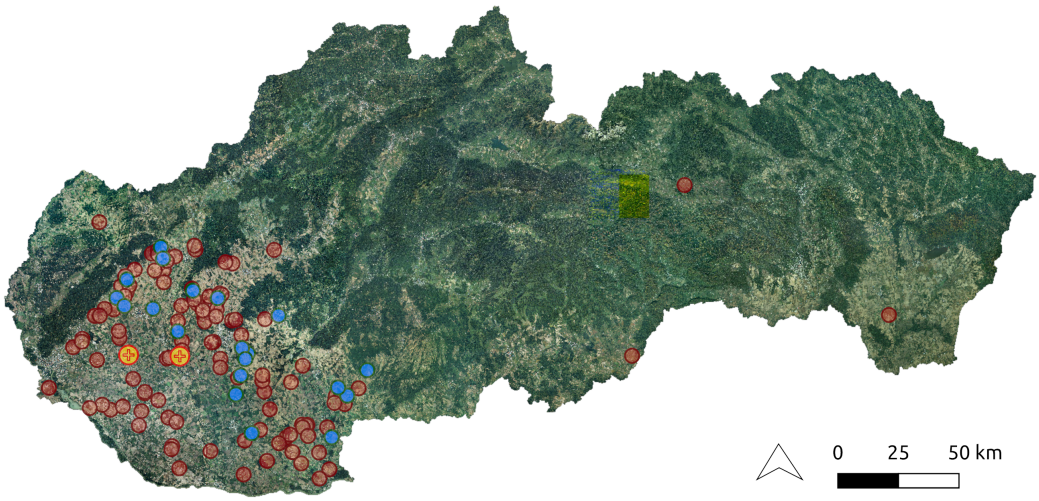

Figure 5.16: 2 new suspected rondels (sites 1 and 3). Both marked in yellow with a plus sign located on the orthophotomosaic.

5.5.1. Site 1: 48.21540, 17.55654 SERED_5-8

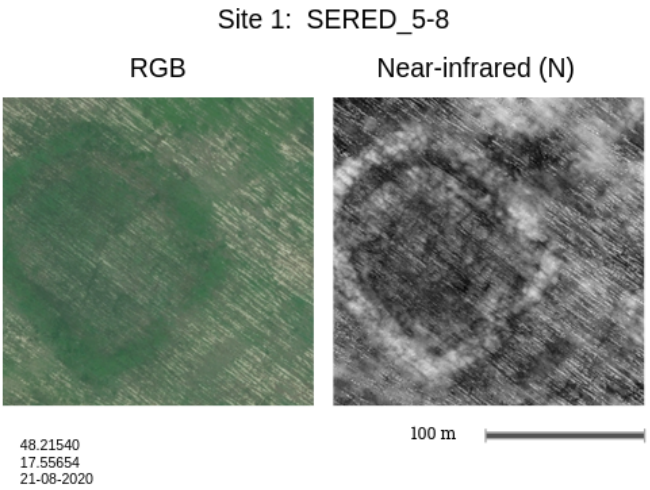

Figure 5.17: 1st suspected rondel.

This site shows a clear rondel-like pattern in the 2nd cycle and 2010 orthophoto images. The 2nd and 1st cycles' images were taken during the same month and season, but the circular ring is not apparent in the 1st cycle unlike in 2010's data set. On a digital terrain map there is a ring-like depression on this location, so it is likely this site contains a rondel.

Figure 5.18: Site 1 on the 1st and 2nd orthophotomosaic.

Figure 5.19: Site 1 on the 2010 and 1950 orthophotomosaic.

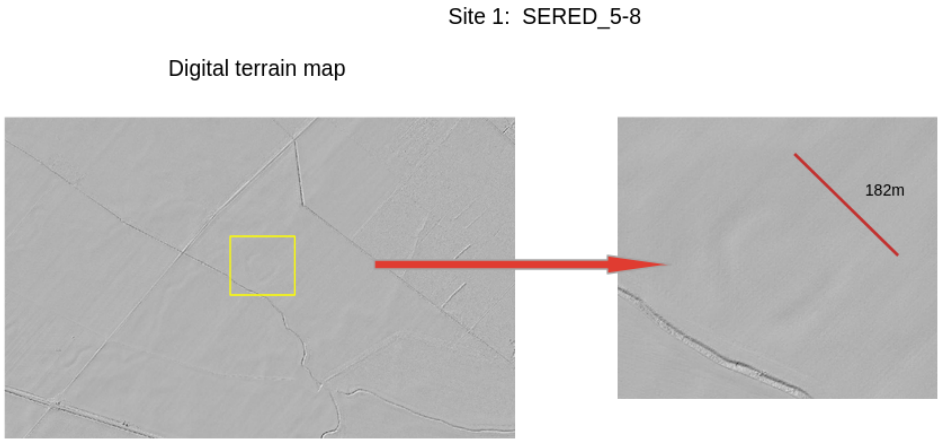

Figure 5.20: Site 1 digital terrain map.

5.5.2. Site 2: 47.94624, 17.69755 DUNAJSKA_STREDA_2-3

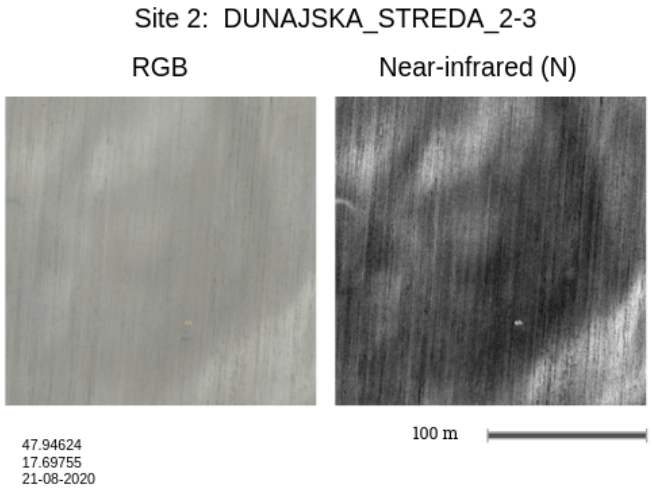

Figure 5.21: 2nd suspected rondel.

Upon zooming out from the concentric circular pattern on this site, the pattern is revealed to be merely a geographical feature resulting from an old part of a river's abandoned channels with vegetation growing above. Although the ResNet model has classified a circular rondel-like pattern correctly, this site likely does not contain a rondel after inspecting the patch's surroundings at a large scale.

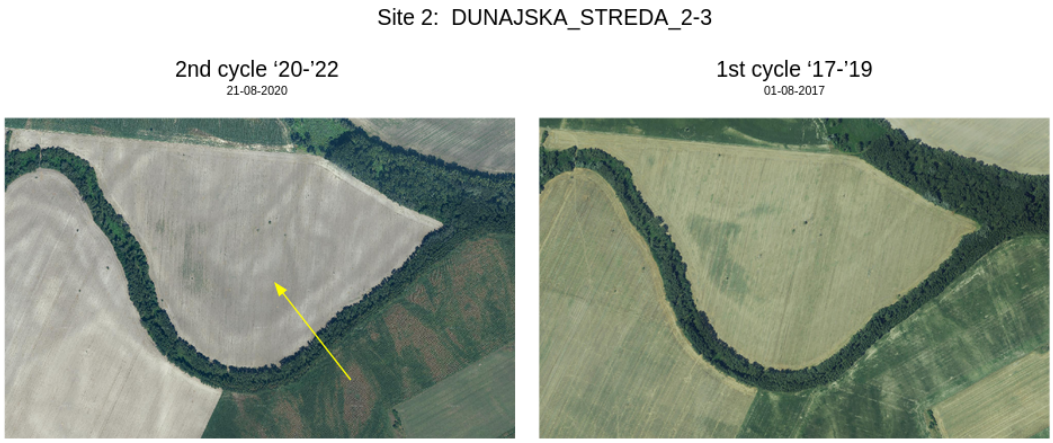

Figure 5.22: Site 2 on the 1st and 2nd orthophotomosaic.

5.5.3. Site 3: 48.22847, 17.84106 MOCENOK_708

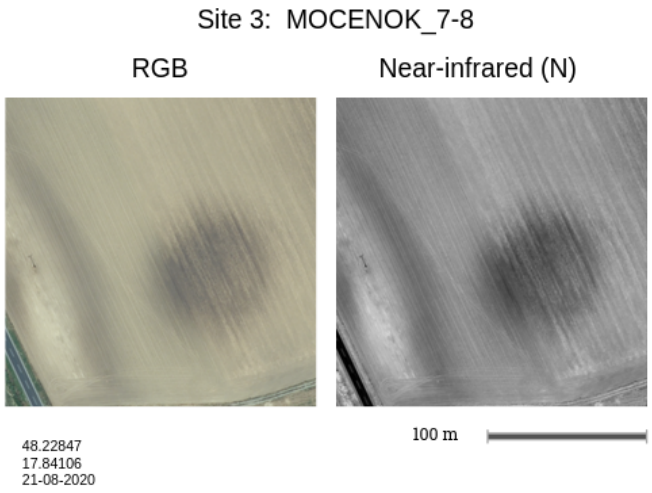

Figure 5.23: 3rd suspected rondel.

In all 4 orthophotos, a dark circular pattern is visible. The circle does not show a ring-like pattern in all but the first cycle orthophotomosaic. On digital terrain, it's location has a circular depression and could be a remain of a ditch. Once zoomed in on the first cycle's image, a slightly darker ringed pattern becomes more apparent. This site likely contains a rondel.

Figure 5.24: Site 3 on the 1st and 2nd orthophotomosaic.

Figure 5.25: Site 3 on the 2010 and 1950 orthophotomosaic.

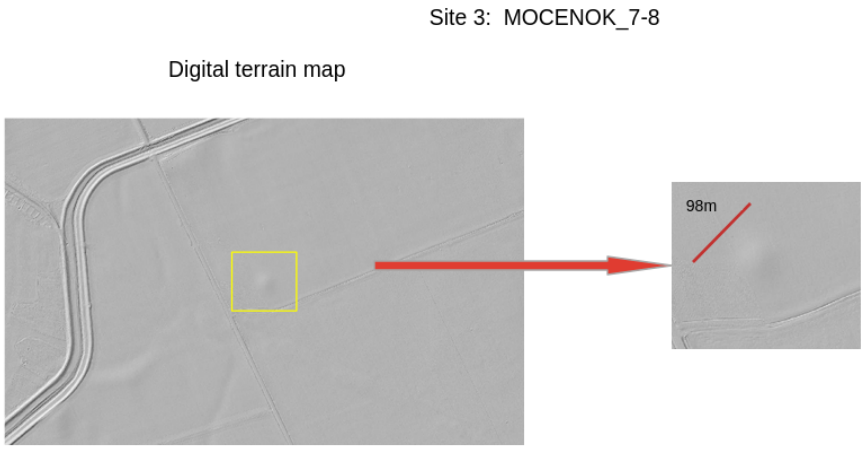

Figure 5.26: Site 3 on a digital terrain map.

Figure 5.27: The ringed ditch zoomed in on the first orthophotomosaic cycle.

5.5.4. Site 4: 48.33399, 17.78244 MOCENOK_8-2

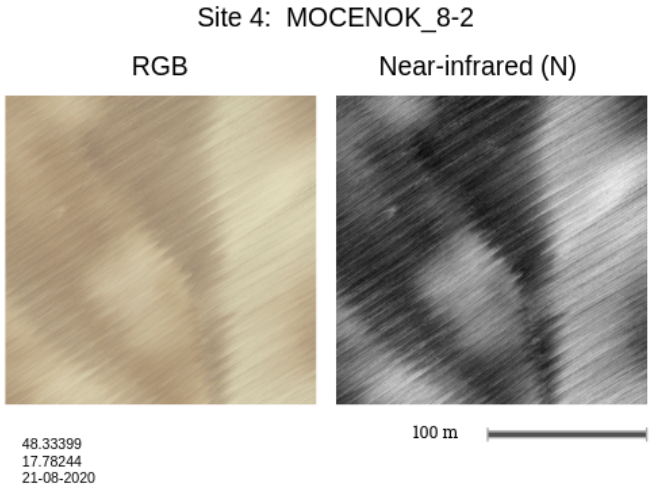

Figure 5.28: 4th suspected rondel.

Similar to site 1, the circular pattern on site 4 is likely to be a pattern resulting from an old part of a river's abandoned channels with vegetation growing above. This site likely does not contain a rondel despite a rondel-like pattern on the orthophoto being classified by the model as a rondel.

Figure 5.29: Site 4 on the 1st and 2nd orthophotomosaic.

5.5.5. Site 5: 48.11326, 18.83470 PLASTOVCE_8-7

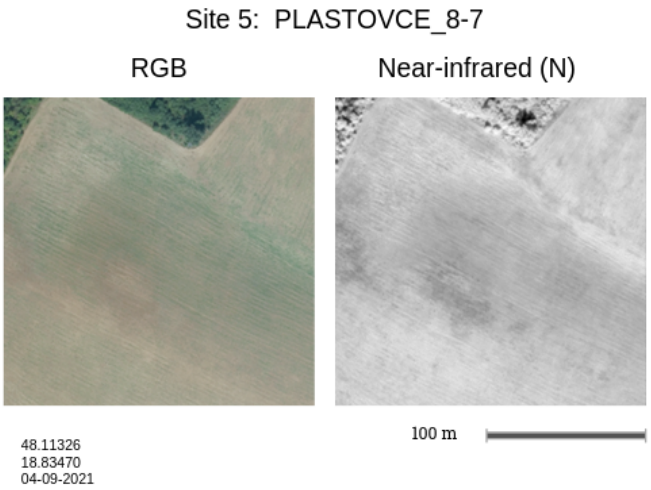

Figure 5.30: 5th suspected rondel.

A darker circular pattern is vaguely visible in the two latest orthophotomosaics. On the digital terrain map, this circle is revealed to be part of a large depression next to a hill which could have been an abandoned river channel, rather than a rondel's remains.

Figure 5.31: Site 5 on the 1st and 2nd orthophotomosaic.

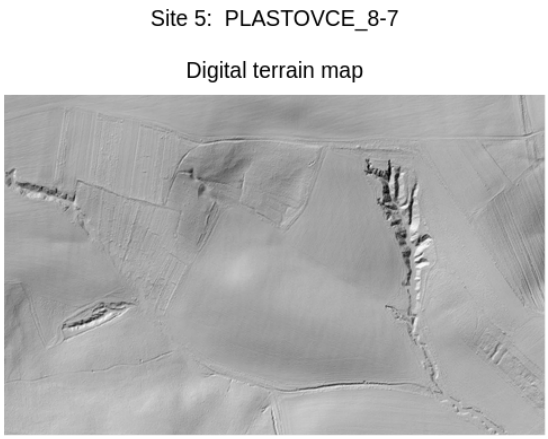

Figure 5.32: Site 5 on a digital terrain map.

5.5.6. Site 6: 48.55702, 17.70455 SENICA_0-9

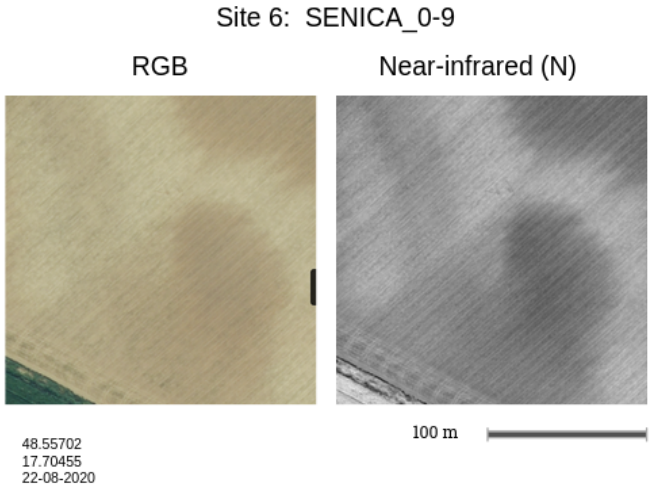

Figure 5.33: 6th suspected rondel.

A circular pattern appears in all 4 orthophotomosaics, however, there is no ditch-like pattern on the digital terrain map. It is unlikely this site contains a rondel despite the existence and classification of a rondel pattern. The digital terrain map shows no clear evidence of the site being an abandoned river channel either.

Figure 5.34: Site 6 on the 1st and 2nd orthophotomosaic.

Figure 5.35: Site 6 on the 2010 and 1950 orthophotomosaic.

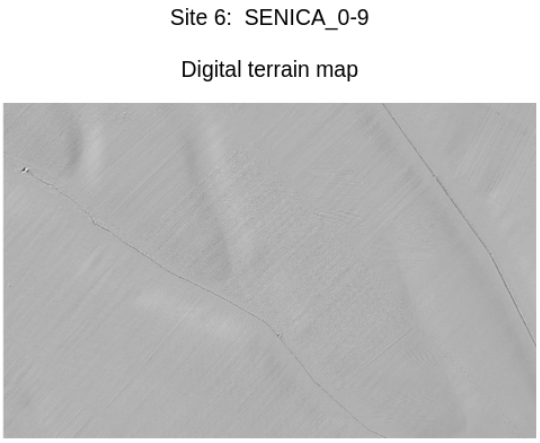

Figure 5.36: Site 6 on a digital terrain map.

5.5.7. Site 7: 48.08548, 18.35974 SURANY_2-7

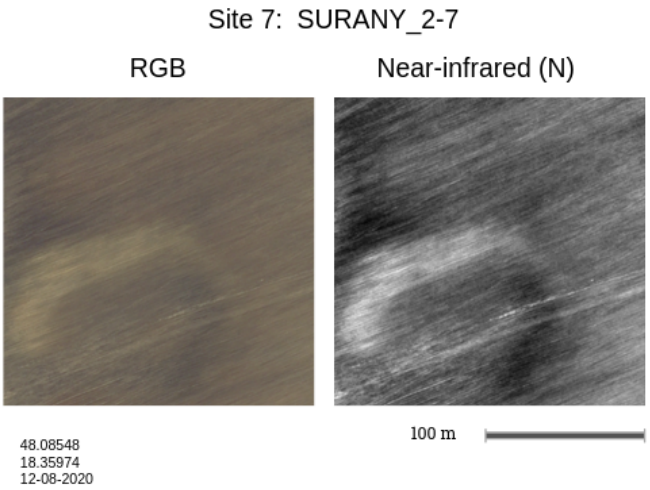

Figure 5.37: 7th suspected rondel.

Only the two latest orthophotos shows a darker circular pattern on the field. This pattern is not visible in orthophotos from 2010 and 1950. In the first cycle's image, nearby lines indicate that the circular pattern could be traces of an old abandoned river channel rather than a rondel. This is confirmed to be the case on the digital terrain map.

Figure 5.38: Site 7 on the 1st and 2nd orthophotomosaic.

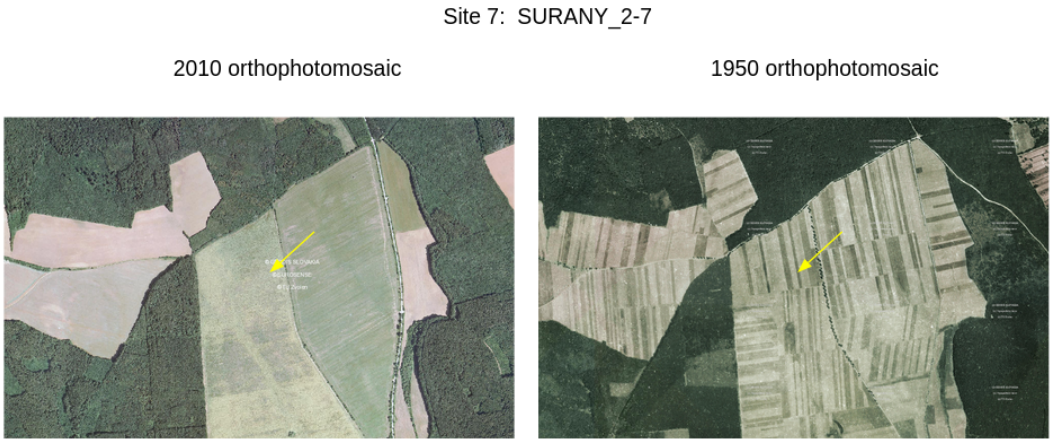

Figure 5.39: Site 7 on the 2010 and 1950 orthophotomosaic.

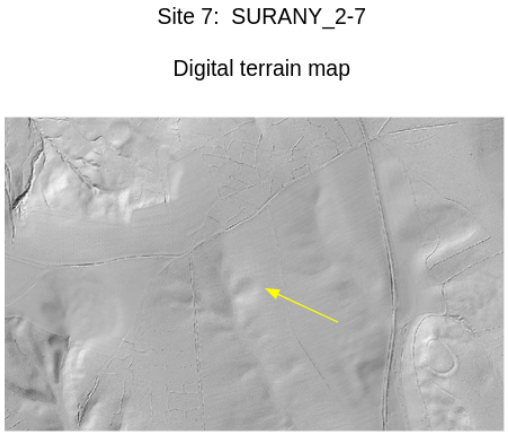

Figure 5.40: Site 7 on a digital terrain map.

6

Conclusion and future work

We'll conclude this thesis by first discussing which method(s) were effective (or otherwise), then by answering our main and sub research questions, and finally by suggesting alternative methods or data sets which could be used to find more new rondels in future projects.

6.1. What worked?

A more complex model isn't always the best model to use in this project's case. A generated data set is required for a better trained model since validation loss is much lower than when the model is trained on a much smaller but real data set consisting of real rondel images. Judging from validation loss graphs, the simple ResNet-34 model with less parameters showed lower loss values and outperformed a more complex ConvNeXt model with 28 million parameters due to a lower false positive rate on images containing no rondels of the 'not a rondel' class. It is possible that the ConvNeXt has identified even more empty agricultural land cover, or its random darker patterns, as rondels than the ResNet. Additionally, a simpler model with less parameters takes less time to train per epoch. We may want to also try a different simple model in future studies or attempt to reduce overfitting at later epochs by introducing learning rate cycling [45].

Regardless of the high false positive count, we were still able to manually and visually narrow down two possible rondels out of 7 potential rondel sites on the 2nd orthophoto cycle with the help of other orthophotomosaic cycles (1st, 2010, and 1950) and a digital terrain map for verification [37]. The locations of the 2 potential rondels are shown in figure 6.2 on Google maps in relation the locations of previously reported rondels and on the Slovak orthophotomosaic in figure 6.1.

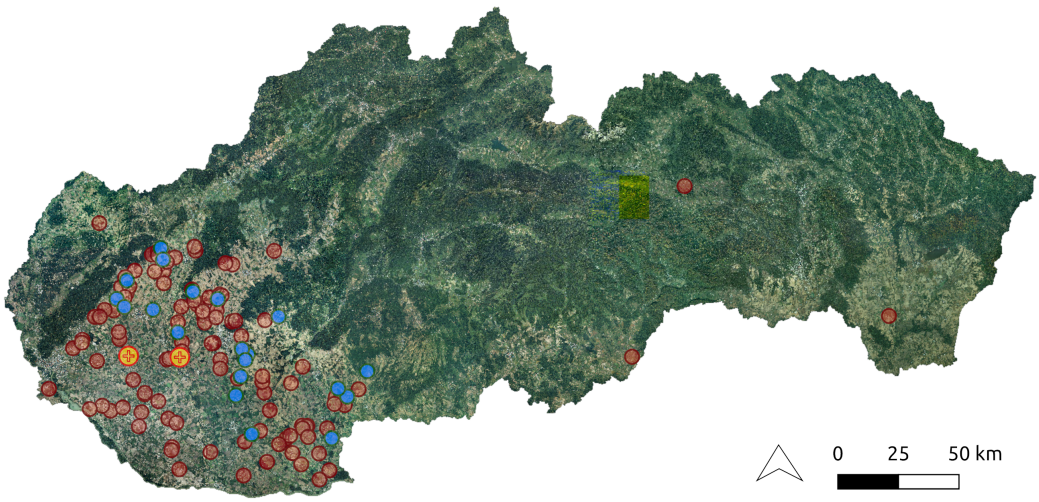

Figure 6.1: 2 new potential rondel sites (sites 1 and 3). Both marked in yellow with a plus sign located on the Slovak orthophotomosaic.

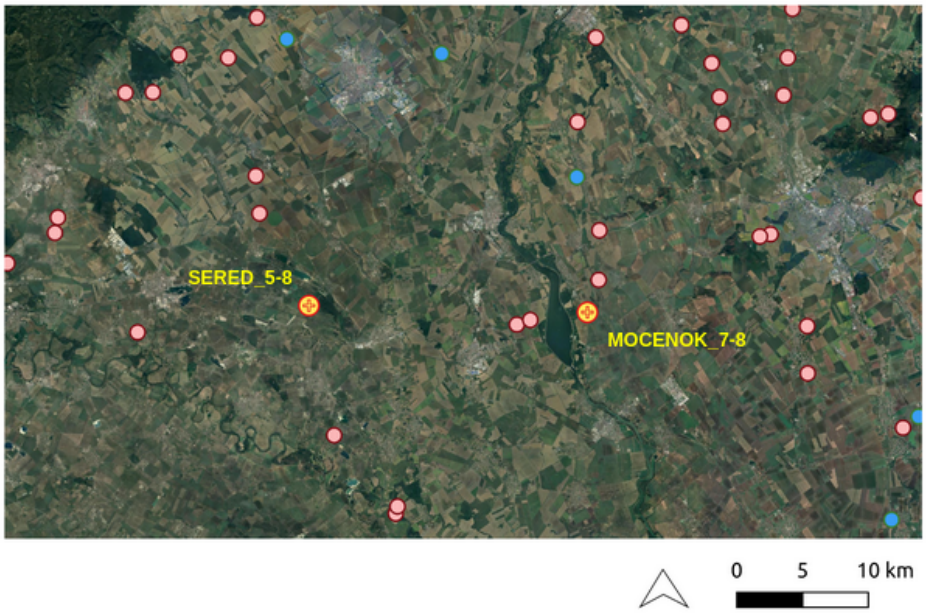

Figure 6.2: 2 new potential rondel sites (yellow plus signs) on Google earth, zoomed in, relative to nearby sites.

6.2. What did not work?

It was time consuming to search for rondels with the current high false positive count. Even though the model has classified multiple patches which contain rondel-like patterns correctly, we are left with the task of further checking to see if the tiles' features actually have the characteristics of a rondel in other orthophotomosaic data sets and in a digital terrain map. Minimizing time searching for rare rondels is the goal of this project, and we attempted to reduce search time by reducing false positive classifications. We cannot confidently say that a tile that contains a rondel-like pattern without the pattern in other data sets are truly false positives, due to the classifier correctly still identifying the pattern as belonging to the 'rondel' class. Further archaeological expertise is needed to study possible identified rondel sites in detail and distinguish their similarly circular patterns from geographical features.

Nevertheless, we encountered several tiles with up to 40 to 60 patches of false positive 'rondel' classes. There is a known uneven and extreme distribution of softmax percentages in the test set in figure 5.12. Images classified as 'not rondels' have a near-zero probability, while images classified as rondels almost always have high softmax probabilities. This proves to be an issue with false positive classifications when we are trying to cut down time manually searching for rondels.

Some tiles (a few examples in figure 6.3) contain large fields with darker incoherent patterns, which resemble few of our test set rondel images, interestingly, yet less circular. Site 2, 4, 6, and 7, our potential sites, are clear examples of this. Up close, a patch may appear to contain a circular patten- upon zooming out to the larger area, the patterns are a part of the natural geography, they could be an old, abandoned river or they could also be arbitrary crop patterns over a field.

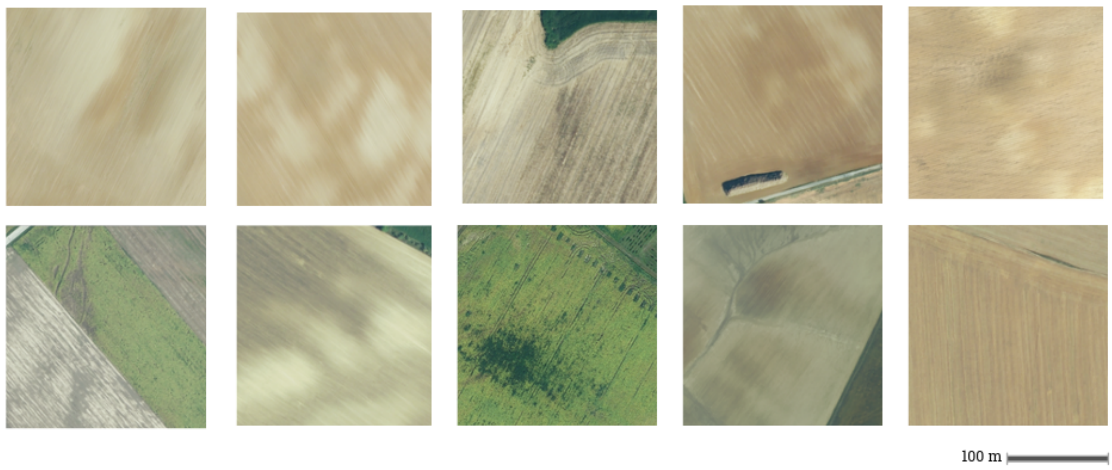

Figure 6.3: Examples of patches on fields classified as rondels where no further evidence of rondels are found. Fields contain somewhat circular and darker patterns (some may be old rivers).

Tiles with multiple circular to oval puddles or pools were classified as rondels, and this may include inherent colored patterns on fields the puddles were on. These circular examples likely were from imagery during a period of high precipitation and moisture and are probably not rondels. Buried archaeological structures are more apparent during drier periods of the year [27] and what we are seeing are puddles.

Figure 6.4: False positive patch examples of puddles and pooling water.

We also encounter several rare tiles, confidently classified as a rondel (even with a probability of 1) where the image definitely does not have a chance of containing a rondel. These are from villages, urbanized areas, and bodies of water where the CORINE 2018 land cover map may have incorrectly indicated as an agricultural land cover. Some bodies of water show a small darker circular pattern and few urbanized areas contain fields. This may explain their classification results, but it is not certain.

Figure 6.5: False positive patch examples on urbanized areas and water bodies.

Another part of the project which did not work was using orthophotomosaic timestamped metadata to explain and visualize seasonal variability in crop pattern visibility or growth. It turns out that for both the

first and the second image cycles, most aerial images were taken within the same months of the year, and differences were at most 2 to 3 weeks apart [21]. We will require additional timestamped imagery from other times of the year [27].

Furthermore, undetected rondels may be located between two orthophoto tiles, causing the sliding window search process to miss out on more possible rondel sites. We do not have enough local or remote computing power to merge the search area tiles into one large .TIF file in either GDAL or QGIS [38], therefore the tiles must be pre-processed then searched through individually. We also ran into an issue with patterns resembling rondels, only once zoomed in. The ResNet classifier does classify such patterns as rondels.

6.3. Rondel visibility

Starting from the second cycle, 4 channels, RGBN, are made available for the orthophotomosaic [21]. Some rondels (Konsky jarok and site 1) are more clearly visible in the near-infrared channels than in RGBN imagery. It is not a uniform requirement either that their ringed ditches should appear darker or brighter in the NIR channel as they do in the RGB channels [42]. This occurs despite literature stating that the ditches will lead to visual differences in the fields from moisture retention and different growth [46] [3], especially in the R,G, and N channels [8]. One of the most obvious rondels (Zadne pole 2) in RGB channels is barely visible in the near-infrared channel, which could have affected classification results and caused it to be classified as empty land.

Most orthophotomosaic tiles were processed from aerial images taken from the same months (with a few weeks of difference) for both the first and the second cycle, therefore the imagery are all from the same season, summer, around August [21]. We do not have orthophotomosaic data of Slovakia, or at least their collection date metadata, from other seasons outside of summer and early fall. For older orthophotomosaic data sets with a resolution of 50 centimeters per pixel, such as 2010 or 1950 RGB aerial images, the data is for now only available without date or scale metadata through an online map service provided by TU Zvoden [37]. Digital terrain imagery from processed LiDAR data is also provided on the same online service. Our figures of possible rondel sites (such as figures 5.25 and 5.19) have shown that even at a 50 cm resolution, we are still able to discern at least a few clear rondel structures and their layouts from 2010 and 1950 orthophoto images, therefore these older data sets served as useful references for the most obvious circular ditches. We cannot easily determine an accurate rondel layout using only digital terrain data from LiDAR due to it's lower resolution at 1 meter per pixel. Older data sets may even display rondels which are later destroyed by urbanization in more recent years.

6.4. Answers to research questions

Main question: How can we detect new rondels in RGBN aerial images with limited positive examples?

We trained a ResNet convolutional neural network classifier on synthetic images of augmented rondel sketches replicating characteristics of rondel remains in central and eastern Europe. Additional synthetic rondel images resolved the class imbalance problems by increasing data available from the rare rondel class. Otherwise, models trained on data with a large class imbalance perform worse [30]. The images were created using rondel layout sketches from archaeological publications and from agricultural land cover extracted out of the orthophoto. Multiple models with varying hyperparameters were trained for 20 epochs and the most promising model was picked and trained further to 200 epochs. This model has the lowest false positive count on our test set and a low validation loss value. Next, a sliding window creates small patches at each orthophotomosaic tile in our search area to run the patches through our model. We then inspect patches classified as 'rondels' for a clear circular rondel structure.

Compared to a model trained with only a few real examples of rondels, the model trained on a generated training set has resulted in lower validation losses and is therefore more effective. We can say that using a generated training set is required for this study.

Sub-question 1: Which configurations and parameters can we use for the training data set and the convolutional neural network models that will result in an effective model for finding new rondels?

A good CNN model for the project's purpose has a low validation loss and a low false positive count. The model should also be able to generalize to new, unseen data.

We are using the cross-entropy loss function on the trained CNN's validation set at each epoch to determine how much a classified image's class probability strays from it's true class . This value is ideally minimized as the function's output increases with a higher probability difference from the true class, therefore better CNN models have low validation set loss values [19].

A low false positive count is visualized using a confusion matrix on our test set containing 20 real rondel examples and 625 images without rondels. Our project goal is to reduce time spent manually finding new rondels in western Slovakia and spending more time should be penalized [49]. A high false positive classification count would cost us more time to sift through possible rondel images which do not contain rondels. We have trained a modified ResNet-34 model that should classify less false positives but also less true positives, meaning that it misses some rondels. We are still successful if we were to miss few barely visible unreported rondels, but detect one or two obviously visible rondel structures.

Sub-question 2: Can we differentiate any seasonal vegetation changes that would affect rondel appearances in the orthophotomosaic data set?

Before we can conclude a seasonal trend, we may have to also account for the influence of weather over multiple years, especially precipitation in the area which affects crop growth [27] [3]. With our current orthophotomosaic data set and their imagery metadata, this is not possible. A majority of aerial images used in this project may be collected during different years and weather conditions, but they are mostly taken on dates that are at most 2-3 weeks apart, likely to intentionally reduce seasonal differences between the cycles. There are also date overlaps in both cycles. Most images in western Slovakia's 1st and 2nd cycles were taken during August, a typically dry month where buried archaeological features become more apparent under crops [27]. We will expand on this in section 7.2 under 'future work'.

Sub-question 3: What additional value do high-resolution multispectral images bring when it comes to rondel detection? How useful are they compared to LiDAR data?

With multispectral imagery, we were able to find 2 unreported and possible rondels for further examination by archaeologists, whereas a previous study utilizing processed LiDAR data was able to find 32 possible rondels within a smaller search area [36]. Many rondels were reported to be built on hillslopes [51], we cannot confirm elevation differences in a rondel's vicinity using solely multispectral imagery since processed LiDAR contains terrain information but multispectral images do not. A drawback LiDAR and their processed digital terrain maps both have is that while their lower resolutions (1 meter)

may be enough to detect a circular ditch-like depression, once the rondel pattern is found, it's exact blueprint cannot be easily drawn or determined. Higher resolution imagery with a lower ground sampling distance (20 cm) can resolve finer details of the rondel's structure for archaeologists to study, such as its openings or its number of rings, due to more details and pixels representing a smaller area [34].

6.5. Future work

6.5.1. Data level improvements

As of mid-July 2024, the third orthophotomosaic starting from 2023 was made available for the western part of Slovakia. This data set contains all RGBN channels at a 15 centimeter resolution [21]. Rondels are may or may not be visible on a particular orthophotomosaic, depending on weather conditions when the aerial images were taken. This data set can be used to perform another search for undetected rondels in the second cycle.

We suggest acquiring high-resolution commercial multispectral satellite data, such as imagery from Planet's SkySat with a 50 centimeter resolution. This data set starts from 2020 to the present date. Satellite images are taken during orbits and have a shorter (daily) revisit frequency than the Slovak cadaster's aerial surveys every 3-4 years [14]. A high revisit frequency equates to the availability of imagery during most seasons of the year, depending on cloud coverage percent [32]. A 50 centimeter resolution from this data set is also enough to resolve a rondel's structural details, as we could see from the 2010 Slovak orthophotomosaic with a 50 cm resolution as well.

To account for rondel-like patterns which may not be rondels at a larger scale (such as sites 2, 4, 5, 6, and 7), we may want to use an image pyramid during our preprocessing steps for the sliding window method to find rondel patterns at multiple different scales of an image, such as at different zoom levels of an orthophotomosaic tile [18] [12]. This method increases file reading and computing time, and therefore takes longer than simply using a uniform-sized sliding window over each tile.

6.5.2. Model level improvements

Due to hardware limitations during the computationally expensive sliding window search [12], we may want to get exchange the sliding window inference for a quicker, and less computationally expensive classification method over larger images. This involves using convolutional layers in a neural network, as they already contain sliding feature detectors before a fully connected layer with an output class. The CNN's fully connected layer is converted to a fully convolutional layer with global average pooling resulting in lower memory usage and similar classification results where features of interest can be detected using a large input area without multiple sliding windows [40] [43].

Finally, on the model level, we may want to try using few-shot image classification models based on existing model backbones, such as the ResNet, commonly used when there are limited annotated examples. Less manually annotated data is required for supervision and training. Methods for few-shot classification are most effective using transformer models, similar to the ConvNeXt model, however ResNet is also a common choice [50]. Boosting algorithms for weight update penalization from the majority class can also be used to insure that more weights are updated based on the minority rondel class [30].

References

[1] Agapiou et al. “Evaluating the Potentials of Sentinel-2 for Archaeological Perspective”. In: *remote sensing* (2014).

[2] Guyot et al. “Combined Detection and Segmentation of Archeological Structures from LiDAR Data Using a Deep Learning Approach”. In: *Journal of Computer Applications in Archaeology* 4 (2021), pp. 1–19.

[3] Kalayci et al. “Multispectral Contrast of Archaeological Features: A Quantitative Evaluation”. In: *remote sensing* 913 (2019), pp. 1–23.

[4] Kuzma et al. “študijné zvesti archeologického ústavu SAV”. In: *Archeologický ústav SAV* 41 (2007).

[5] Ridky et al. *Big Men or Chiefs? Rondel Builders of Neolithic Europe*. Oxbow, 2018.

[6] Ronneberger et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: (2015).

[7] Zhang et al. *Dive into Deep Learning*. 1st ed. Cambridge: Cambridge University Press, 2023.

[8] Syed Aqdus, Jane Drummond, and William Hanson. “Discovering Archaeological Cropmarks: A Hyperspectral Approach”. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2008). URL: https://api.semanticscholar.org/CorpusID:7417751.

[9] Aharon Azulay and Yair Weiss. *Why do deep convolutional networks generalize so poorly to small image transformations?* 2019. arXiv: 1805.12177 [cs.CV]. URL: https://arxiv.org/abs/1805.12177.

[10] Randall Balestriero, Ishan Misra, and Yann LeCun. “A Data-Augmentation Is Worth A Thousand Samples: Exact Quantification From Analytical Augmented Sample Moments”. In: *arXiv* (2022). DOI: 10.48550/ARXIV.2202.08325. URL: https://arxiv.org/abs/2202.08325.

[11] Ivan Belic and Cody Styker. *What is learning rate in machine learning?* en. Nov. 2024. URL: https://www.ibm.com/think/topics/learning-rate.

[12] Christopher Bishop and Hugh Bishop. *Deep Learning Foundations and Concepts*. Springer, 2023. ISBN: 978-3-031-45468-4. URL: https://www.bishopbook.com/.

[13] Gian Carlo Cardarilli et al. “A pseudo-softmax function for hardware-based high speed image classification”. en. In: *Scientific Reports* 11.1 (July 2021), p. 15307. ISSN: 2045-2322. DOI: 10.1038/s41598-021-94691-7. URL: https://www.nature.com/articles/s41598-021-94691-7.

[14] Mike Davis. *SkySat 50cm FAQ*. 2020. URL: https://support.planet.com/hc/en-us/articles/360010635857-SkySat-50cm-FAQ.

[15] Google Developers. *Overfitting*. 2024. URL: https://developers.google.com/machine-learning/crash-course/overfitting/overfitting.

[16] Google Developers. *Overfitting: Interpreting loss curves*. 2024. URL: https://developers.google.com/machine-learning/crash-course/overfitting/interpreting-loss-curves.

[17] USGS EROS. *USGS EROS Archive - Aerial Photography - High Resolution Orthoimagery (HRO)*. 2019. URL: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-aerial-photography-high-resolution-orthoimagery-hro (visited on 2024).

[18] Sanja Fidler. *Object Detection Sliding Windows*. URL: https://www.cs.utoronto.ca/~fidler/slides/CSC420/lecture17.pdf.

[19] Brendan Fortuner. URL: https://ml-cheatsheet.readthedocs.io/en/latest/loss_functions.html.

[20] Ruth Fraňková and Štěpán Sedláček. *Archaeologists in Prague uncover ancient Neolithic structure*. en. 2022. URL: https://english.radio.cz/archaeologists-prague-uncover-ancient-neolithic-structure-8760696.

[21] Geoportal. *Orthophotomosaic of Slovakia*. 2023. URL: https://www.geoportal.sk/en/zbgis/orthophotomosaic/.

[22] Afshin Gholamy, Vladik Kreinovich, and Olga Kosheleva. *Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation*. UTEP-CS-18-09. University of Texas at El Paso, Dec. 2018, p. 7. URL: https://scholarworks.utep.edu/cs_techrep/1209/.

[23] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV]. URL: https://arxiv.org/abs/1512.03385.

[24] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: 2015, pp. 448–456. URL: http://jmlr.org/proceedings/papers/v37/ioffe15.pdf.

[25] Justin M. Johnson and Taghi M. Khoshgoftaar. "Survey on deep learning with class imbalance". en. In: *Journal of Big Data* 6.1 (Dec. 2019), p. 27. ISSN: 2196-1115. DOI: 10.1186/s40537-019-0192-5. URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0192-5.

[26] V. Roshan Joseph. "Optimal ratio for data splitting". In: *Statistical Analysis and Data Mining* (2022), pp. 531–538. ISSN: 1932-1864, 1932-1872. DOI: 10.1002/sam.11583. URL: https://onlinelibrary.wiley.com/doi/10.1002/sam.11583.

[27] Dimitrios Kaimaris, Petros Patias, and Maria Tsakiri. "Best period for high spatial resolution satellite images for the detection of marks of buried structures". en. In: *The Egyptian Journal of Remote Sensing and Space Science* 15.1 (June 2012), pp. 9–18. DOI: 10.1016/j.ejrs.2011.12.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1110982311000366.

[28] Zoumana Keita. *An Introduction to Convolutional Neural Networks (CNNs)*. 2023. URL: https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns/ (visited on 2024).

[29] Sung Kim and Jenny Kang. 2017. URL: https://pytorch.org/tutorials/beginner/blitz/data_parallel_tutorial.html.

[30] Bartosz Krawczyk. "Learning from imbalanced data: open challenges and future directions". en. In: *Progress in Artificial Intelligence* 5.4 (Nov. 2016), pp. 221–232. ISSN: 2192-6352, 2192-6360. DOI: 10.1007/s13748-016-0094-0. URL: http://link.springer.com/10.1007/s13748-016-0094-0.

[31] Miroslav Kubat, Robert C. Holte, and Stan Matwin. "Machine Learning for the Detection of Oil Spills in Satellite Radar Images". In: *Machine Learning* 30.2/3 (1998), pp. 195–215. ISSN: 08856125. DOI: 10.1023/A:1007452223027. URL: http://link.springer.com/10.1023/A:1007452223027.

[32] Planet Labs. *Usable Data in Planet imagery*. 2024. URL: https://developers.planet.com/docs/planetschool/usable-data-in-planet-imagery/.

[33] Zhuang Liu et al. "A ConvNet for the 2020s". In: *Computer Vision and Pattern Recognition (cs.CV)* (2022). DOI: 10.48550/ARXIV.2201.03545. URL: https://arxiv.org/abs/2201.03545.

[34] Emily Loosli. *Ground sample distance explained and why it matters*. 2024. URL: https://wingtra.com/surveying-gis/ground-sample-distance/.

[35] Ilya Loshchilov and Frank Hutter. *Decoupled Weight Decay Regularization*. 2019. arXiv: 1711.05101 [cs.LG]. URL: https://arxiv.org/abs/1711.05101.

[36] Yuqi Meng. *Identification of Neolithic Circular Enclosures through Aerial Imagery: A study about pattern recognition and deep learning techniques*. TU Delft, Cosine BV, 2022.

[37] GIS portál. *Historická ortofomapa Slovenska*. 2022. URL: https://gis.tuzvo.sk/?page_id=1226.

[38] QGIS Project. *17.15. Clipping and merging raster layers*. 2025. URL: https://docs.qgis.org/3.34/en/docs/training_manual/processing/cutting_merging.html.

[39] PyTorch. *Optimizing Model Parameters*. Jan. 2024. URL: https://pytorch.org/tutorials/beginner/basics/optimization_tutorial.html.

[40] Zhongchao Qian et al. “Do We Need Fully Connected Output Layers in Convolutional Networks?” In: *preprint* (2020). DOI: 10.48550/ARXIV.2004.13587. URL: https://arxiv.org/abs/2004.13587.

[41] Masato Sakai et al. “AI-accelerated Nazca survey nearly doubles the number of known figurative geoglyphs and sheds light on their purpose”. en. In: *Proceedings of the National Academy of Sciences* 121.40 (Oct. 2024). ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.2407652121. URL: https://pnas.org/doi/10.1073/pnas.2407652121.

[42] Max Sap. *Returning the gaze of our ancestors: A research in the applicability of different visualisation techniques to help with automated learning to better prospect neolithic circular enclosures within Central Europe*. Leiden university, Cosine BV, 2023.

[43] Grigory Serebryakov and Satya Mallick. *Fully Convolutional Network For Image Classification on Arbitrary Sized Image*. 2020. URL: https://learnopencv.com/fully-convolutional-image-classification-on-arbitrary-sized-image/.

[44] Copernicus Land Monitoring Service. *CORINE Land Cover 2018 (vector)*. en. 2019. DOI: 10.2909/71C95A07-E296-44FC-B22B-415F42ACFDF0. URL: https://sdi.eea.europa.eu/catalogue/copernicus/api/records/71c95a07-e296-44fc-b22b-415f42acfdf0?language=all.

[45] Leslie N. Smith. “Cyclical Learning Rates for Training Neural Networks”. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017, pp. 464–472. DOI: 10.1109/WACV.2017.58.

[46] Geert Julien Verhoeven. “Near-Infrared Aerial Crop Mark Archaeology: From its Historical Use to Current Digital Implementations”. In: *J Archaeol Method Theory* 19 (2011). DOI: https://doi.org/10.1007/s10816-011-9104-5.

[47] Chien Vu. *Do and don’t when using transformation to improve CNN deep learning model*. 2020. URL: https://towardsdatascience.com/improves-cnn-performance-by-applying-data-transformation-bf86b3f4cef4.

[48] Mario Wallner et al. “Interdisciplinary Investigations of the Neolithic Circular Ditch Enclosure of Velm (Lower Austria)”. en. In: *Remote Sensing* 14.11 (Jan. 2022), p. 2657. ISSN: 2072-4292. DOI: 10.3390/rs14112657. URL: https://www.mdpi.com/2072-4292/14/11/2657.

[49] Gary M. Weiss. “Mining with rarity: a unifying framework”. In: *SIGKDD Explor.* 6 (2004), pp. 7–19. URL: https://api.semanticscholar.org/CorpusID:1212431.

[50] Han-Jia Ye et al. *Few-Shot Learning via Embedding Adaptation with Set-to-Set Functions*. 2021. arXiv: 1812.03664 [cs.LG]. URL: https://arxiv.org/abs/1812.03664.

[51] Georg Zotti and Wolfgang Neubauer. “Astronomical and Topographical Orientation of Kreisgrabenanlagen in Lower Austria”. In: (2011), pp. 188–193. URL: https://astrosim.univie.ac.at/media/2011SEAC_Zotti-Neubauer_preprint_WEB.pdf.