The Pennsylvania State University
The J. Jeffrey and Ann Marie Fox Graduate School

THE GEOGRAPHY OF GENOCIDE: USING MACHINE LEARNING TO LOCATE UNDOCUMENTED MASS GRAVES OF THE HOLOCAUST

A Thesis in

Spatial Data Science

by

Kelly Schierman

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2024

The thesis of Kelly Schierman was reviewed and approved by the following:

Brandi Gaertner
Assistant Teaching Professor of Spatial Data Science
Thesis Advisor

Leanne Sulewski
Adjunct Instructor of Geography

Anthony Robinson
Professor of Geography
Director of Online Geospatial Education

ABSTRACT

Following their invasion of the Soviet Union in 1941, the Germans killed at least two million Jews in Eastern Europe by shooting them at close range and burying them in pits. Thousands of these mass graves are scattered across at least ten countries, and an unknown number of graves remain today forgotten and undocumented. The most commonly used methods of locating undocumented mass graves involve employing a combination of archival research, surface-based archaeological investigations, and, perhaps most critically, the testimony of eyewitnesses or survivors. However, the advanced age of most witnesses has created an imperative for developing new methods to support the discovery of undocumented mass graves. The maximum entropy (MaxEnt) machine learning algorithm uses a series of environmental variables to describe known presence points – in this case, the locations of mass graves – and that information is used to generate a geospatial model for predicting possible presence locations for undocumented mass graves. The objective of this study was to use MaxEnt to develop a viable model for predicting potential presence locations for undocumented mass graves as an assist to traditional methods. The most successful predictive model, using a combination of 13 geospatial variables, provided an area under the curve (AUC) value of 0.9417 and an omission rate of 0.0690, correctly identifying the locations for 207 out of a total of 229 known mass graves.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest thanks to Dr. Brandi Gaertner and Dr. Leanne Sulewski: Your encouragement, support, and unflagging patience transformed this lengthy journey into a delightful adventure. Thank you so much! I would also like to extend my deepest gratitude to Ms. Milda Jakulytė-Vasil, whose extraordinary work on both the Holocaust Atlas of Lithuania and on MACEVA's documentation of Jewish cemeteries in Lithuania has been an enormous boon to mankind's store of Holocaust-era knowledge; much of Ms. Jakulytė-Vasil's work forms the foundation of this research. Heather Ross and Tara L. Anthony, GIS specialists at the PSU libraries, answered a heroic number of questions and provided exemplary guidance early in my research; I am deeply indebted to their expertise in finding the unfindable. Likewise, Theresa Quill, map and spatial data librarian at the University of Indiana, went to an extraordinary amount of trouble to supply me with a large, and impressively complete, collection of Soviet-era topographic maps. I am deeply humbled by the number of extraordinarily talented people who have so generously shared their time and resources with me. Thank you.

And to my family – Gary, Jonathan, Timothy, and Benjamin – thank you so much for your sacrifices that give me the time, space, and opportunity to go on this adventure. Your encouragement and support have meant the world to me.

Chapter 1: Introduction and Background

Between 1933 and 1945, the Germans conducted mass killings of at least 6 million Jewish people in the territories they occupied, in addition to killing an additional 6 million Poles, Soviets, Roma, people with disabilities, and political dissenters (United States Holocaust Memorial Museum, 2023). While most victims from the western part of German territory died in concentration or extermination camps, in the eastern part of their territory, the Germans developed a new method of mass killings: The victims, mostly Jewish people, were concentrated into a ghetto within their home city or village, transported to a killing site, shot at close range, and the bodies were piled in pits and buried (Desbois, 2015; United States Holocaust Memorial Museum, n.d.; Yad Vashem, n.d.). It is estimated that, using these methods, the Germans killed between 2 million and 3.2 million victims (Silberklang, 2015; United States Holocaust Memorial Museum, 2021).

More than 3,000 of these mass graves have been documented across at least ten countries, stretching from the eastern half of Poland in the west, to the modern-day city of St. Petersburg in the north and east, reaching almost to the country of Georgia in the south. Figure 1 shows the distribution of documented mass graves across eastern Europe. These graves account for more than 1.5 million victims, most of them Jewish (*About Us: Yahad-in Unum*, n.d.). However, it is estimated that the locations and existence of most mass graves remain unknown (Sturdy Colls, 2015, p. 278; Silva & Burns, 2015).

There are many reasons these graves have been lost. The Germans tried to obscure the killing actions by choosing locations away from cities, villages, and major traffic routes (Kwiet, 1993), and they also often worked to hide the evidence of mass killings by digging up and cremating the remains after the fact (Desbois, 2010). Additionally, research into undocumented mass graves has long been a neglected area of Holocaust research, since most of the information

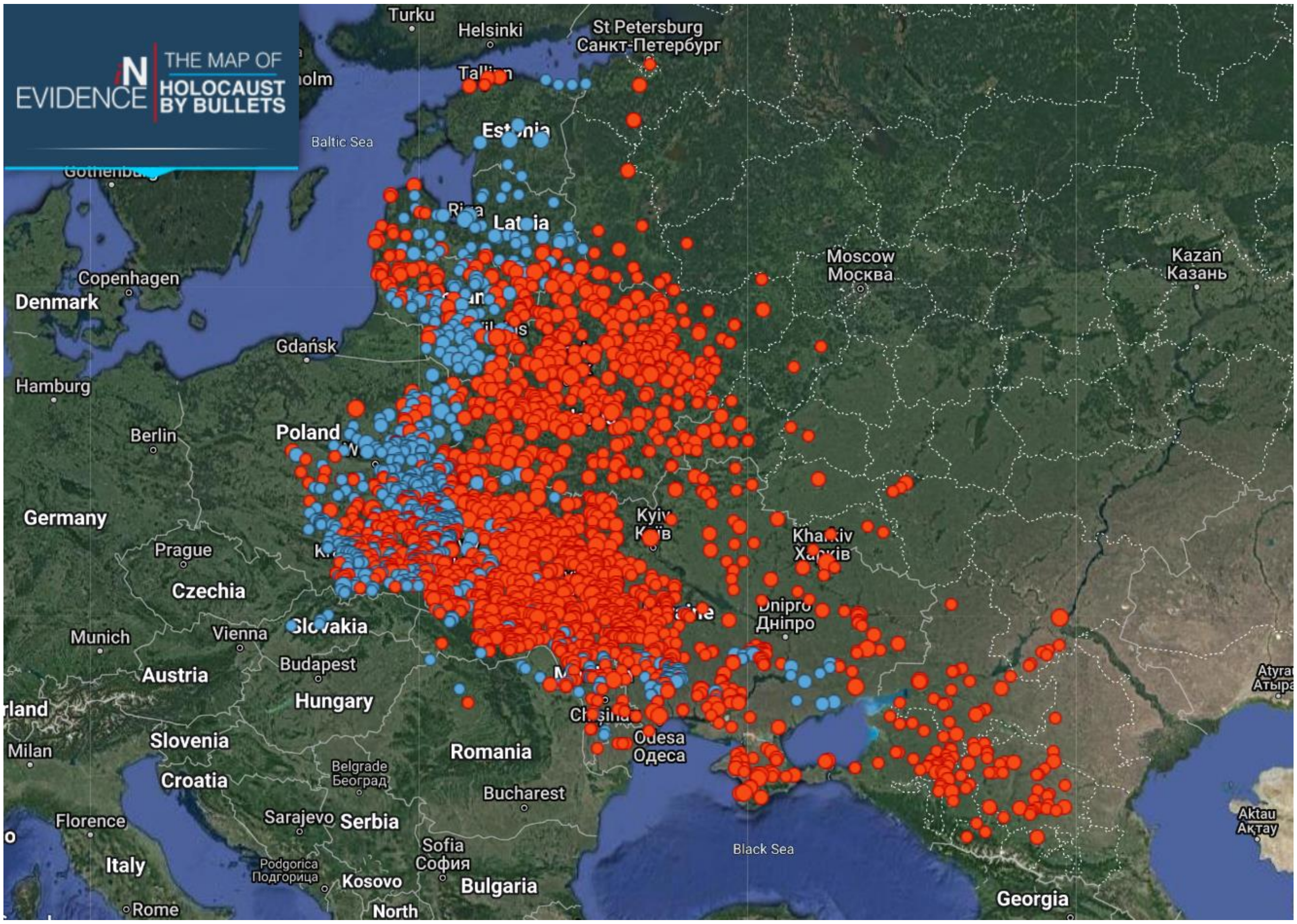

Figure 1: Yahad-in Unum map showing locations of mass killings and burials documented by eye-witness testimony. Red dots represent locations which have been fully documented online; blue dots represent locations for which full documentation has not yet been posted online. © Yahad-in Unum

was unavailable to researchers until after the Cold War (Walke, 2022).

Although there was not a standard procedure for selecting the site of a mass grave (Desbois, 2015; Sturdy Colls, 2015), many mass graves in Eastern Europe do often share certain characteristics. Whether the victims came from villages with a majority Jewish population or were from towns or villages with a minority of Jewish residents, some of the mass grave characteristics include:

- They were generally located at least a short distance away from populated areas, but with convenient access by road or footpath.
- Remote areas covered with trees – such as wooded parks, forests, or Jewish cemeteries – were often chosen for ease of concealment.
- Natural features, including low-lying areas such as marshes and ravines were used because of the relative convenience of burying bodies in these locations.
- Victims were often shot into rivers to avoid having to bury the bodies at all.
- In or near urban areas, Jewish residents were often killed within places like forts, prisons, or synagogues, and were buried nearby.

In most cases, the overarching goal was to choose killing locations that were convenient for the Germans; convenient to villages supplying conscripted labor; and convenient for the transport and management of large numbers of victims (Desbois, 2010, 2015; Pohl, 2015; Sturdy Colls, 2015).

The Germans' first mass shooting *Aktionen* ("actions") occurred in what is now the modern country of Lithuania (Porat, 1994), and during their occupation the Germans killed an estimated 90 to 95 percent of Lithuanian Jewry, amounting to as many as 250,000 people (Malinauskaite, 2013; United States Holocaust Memorial Museum, 2024). In fact, Lithuania had one of the highest victim rates in all of Europe (United States Holocaust Memorial Museum, 2024); thereby creating a concentrated source of information about the invading Germans' processes and procedures. However, even in Lithuania where nearly 10 percent[1] of the

[1] Before World War II, there were approximately 160,000 Jews in Lithuania, accounting for 7 percent of Lithuania's population. By the time Germany invaded in the summer of 1941, Jewish refugees from the west had swelled the number of Jews in the country to approximately 250,000, or 10 percent of the population of Lithuania (United States Holocaust Memorial Museum, 2024).

population was massacred (United States Holocaust Memorial Museum, 2024), it is not known how many grave locations remain unknown or undocumented. As recently as 2018, Fuerstenberg et al. (2018) identified a previously undocumented mass grave in northeastern Lithuania, and Reeder et al. (2024) confirmed the existence of undocumented mass grave locations in both Latvia and Lithuania in the summer of 2024.

Ignorance of the locations of additional mass graves combined with a vast search area creates an opportunity to develop models that can help researchers narrow the scope of their search for undocumented mass grave locations using bespoke geospatial analytic methods.

Lithuanian Jews, or Litvaks[2], have lived in Lithuania since the 14th century (Iwaskiw & Library of Congress, 1996). Consequently, there is a wealth of available historical data on Lithuanian Jewish history that can aid researchers in narrowing their searches for undocumented mass graves. The JewishGen website (JewishGen, n.d.-b) and the website of the Lithuania Jewish Special Interest Group (LitvakSIG, 2022) together provide detailed documentation of more than 300 Jewish villages that predate World War II. The Lithuanian non-profit organization MACEVA[3] has collected extensive data on the locations of pre-War Jewish cemeteries (*Maceva: Litvak Cemetery Catalogue*, n.d.). Additionally, the online Holocaust Atlas (*Holocaust Atlas of Lithuania*, n.d.; Jakulyte-Vasil, 2015) documents the locations of all known mass graves in the country.

This data is of the utmost importance as eyewitnesses to these atrocities slowly age. Historically, discover and positive identfication of mass graves has relied on interviews and

[2] "Litvak" is a commonly-used term, borrowed from Yiddish, for a Jew with ancestors from the area generally encompassing the historical Grand Duchy of Lithuania, which was considerably larger than the modern-day state of Lithuania (Nadler, 2010).
[3] "Maceva" is the Eastern European spelling of the Israeli Hebrew word *matsevá* (transliterated) which means "gravestone" (Katz, 2023).

testimonies of eyewitnesses (*About Us: Yahad-in Unum*, n.d.; Courcelle et al., 2023; Haglund et al., 2001; Pohl, 2015; Sturdy Colls, 2015; Walke, 2022). However, the passage of time is claiming the lives of these witnesses; the current average age of Holocaust survivors and witnesses is 86 (Claims Conference, 2024), making the application of alternative discovery methods even more important. Advanced statistical and machine learning models can help fill this looming gap in the search for unknown graves.

One such model is maximum entropy (MaxEnt), a machine learning tool that evaluates the intersection of multidimensional data to train the model on a range of predictor variables and datasets. It was originally developed to calculate potential species range using environmental variables; it requires only a defined study area, known presence points, and environmental variables in order to operate (Phillips et al., 2006). Since its development, MaxEnt has also been adapted for location prediction. Previous archaeological research has employed the MaxEnt model to predict unknown or undocumented historical and ancient locations with high accuracy: It has been used to locate ancient land use patterns (Galletti et al., 2013), ancient inscriptions (Gillespie et al., 2016), ancient archaeological sites (Rafuse, 2021), human settlement patterns (Wachtel et al., 2020), and clandestine graves (Congram et al., 2017; Molina et al., 2020). One of the great advantages of MaxEnt is that it only requires data about the presence of an entity, without requiring knowledge of locations where absence is confirmed (Phillips et al., 2006).

The goal of this research is to employ the maximum entropy model to predict the locations of unknown mass graves in Lithuania by using the locations of known mass graves and environmental variables which describe those locations. The environmental variables will include those factors which the Germans are known to have frequently employed in their selection of burial sites including: Road networks; distance to towns, Jewish communities, and

Jewish cemeteries; land cover; land use; distance to water features; and natural geographical features such as slope, aspect, and elevation (Desbois, 2010, 2015; Pohl, 2015; Sturdy Colls, 2015).

Chapter 2: Methodology

2.1 Research Area

This research is implemented in the modern-day country of Lithuania, which is the southernmost of the Baltic states. It is bordered by Latvia on the north; the Baltic Sea on the east; Poland and the Russian enclave of Kaliningrad on the south; and Belarus on the east. The country has an average elevation of 110 meters (360 feet). Highland glacial deposits in the northwest and southeast reach a maximum elevation of 293 meters (961 feet). These highlands are surrounded by lowlands covered with scattered lakes and marshes and a dense network of streams and rivers. Lithuania has a land area of 62,280 km^2 (Central Intelligence Agency, 2024).

Forestry and agriculture were critical parts of the pre-War Lithuanian economy. In 1939, 92 percent of Lithuania's land area was devoted to some form of agriculture or forestry (Central Intelligence Agency, 1958). At the time of the German invasion, most towns and villages had a small forest nearby which was used by the local residents for fuel and other resources (Central Intelligence Agency, 1958; Desbois, 2015).

Before World War II, almost all Lithuanian Jews were employed in commerce, banking, and industry, very few owned agricultural lands. Consequently, the greatest numbers of Jewish inhabitants were found in towns and cities: In 1923, 94.8 percent of all Lithuanian Jews lived in urban areas, and this figure remained largely unchanged until the German invasion in 1941 (Central Intelligence Agency, 1958). However, the largest number of Jewish communities were scattered across the countryside. Since Jews provided many non-agricultural services, outside of urban areas, their communities were typically located within or near rural, non-Jewish communities which depended on their services (Central Intelligence Agency, 1958; Desbois, 2015); thus, Jewish settlement patterns across Lithuania generally follow the settlement patterns of non-Jewish Lithuanians.

2.2 Study Area

The study area for this research is limited to the five northernmost counties (*apskritis*) in the country: Klaipėdos, Telšiu, Šiauliu, Panevėžio, and Utenos[4]. These counties were chosen as a representative sample of Lithuania as a whole, providing a subset of data for training. See Figure 2 for a map of Lithuania highlighting the study area.

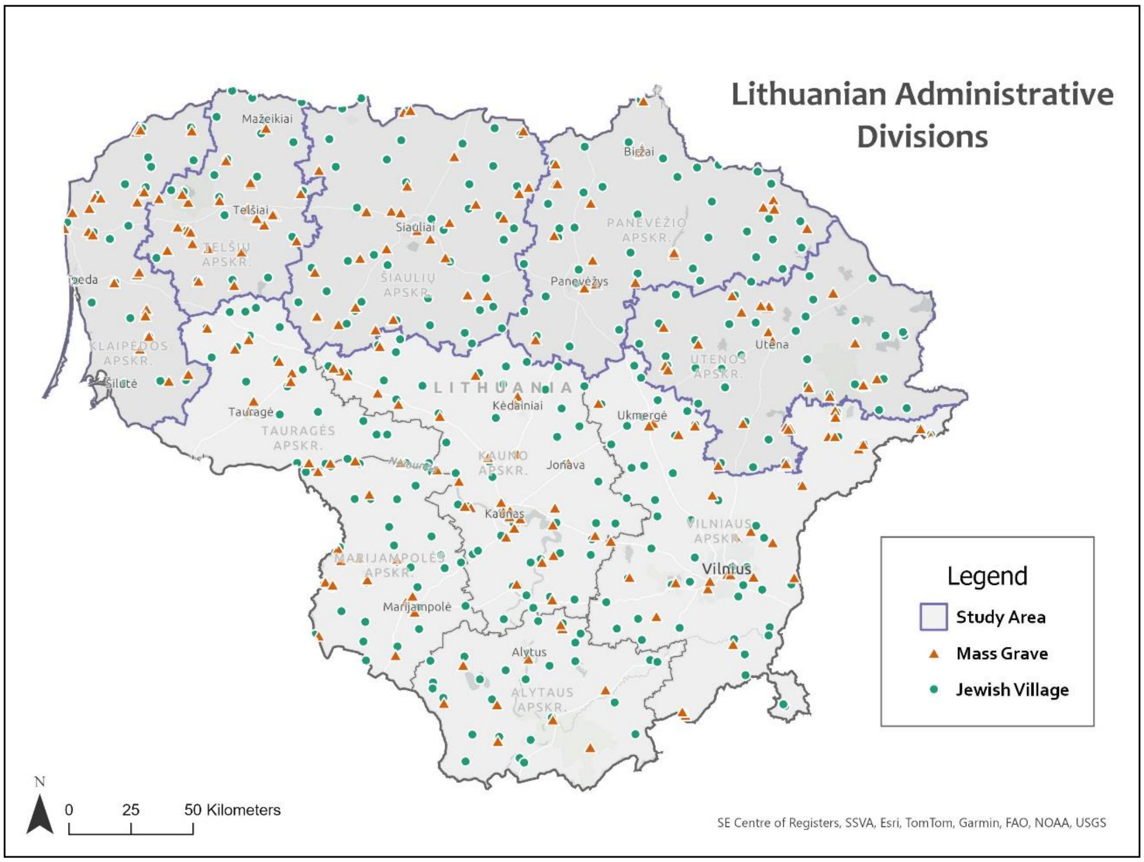

Figure 2: A map showing the administrative divisions of Lithuania and the locations of mass graves and Jewish villages throughout the country. "Apskritis," abbreviated "apskr," is the Lithuanian term for a county-level administrative area. The five counties used in the training data are highlighted and outlined in purple.

---

[4] The complexities of the varying grammatical forms in Lithuanian are far beyond the scope of this paper. The names of the counties are spelled to reflect their spellings on the accompanying map; city names follow conventional spellings as encountered in non-Lithuanian sources, and the author offers deepest apologies for the consequent misuse of the language.

This five-county region is bordered by Latvia in the north and by Belarus in the east; three of Lithuania's remaining five counties lie along its southern border. Klaipėdos *apskritis* contains Lithuania's entire coastline with the Baltic Sea and contains Lithuania's lowest elevation of 0.27 meters below sea level. Telšiu, Šiauliu, and Utenos are marked by glacial highlands which reach a maximum elevation of 288 meters in Utenos; the center of this region, the Panevėžio *apskritis*, is defined by a broad, flat, fertile plain with an elevation of 3.9 meters above sea level at its lowest point. Small farms and forest plots dominate this region, as they do in much of Lithuania. The major rivers in this region are the Venta, which flows north out of Šiauliu *apskritis*, and the Šventoji, which flows south out of Lake Samanis in Utenos *apskritis*. The rest of this region is watered by a network of smaller streams and rivers and contains numerous lakes and wetlands. Klaipėda, Šiauliai, and Panevėžys, eponymous cities of their respective counties, are three of Lithuania's six largest cities; Klaipėda is also Lithuania's major port.

This area contains 47.32 percent of all Jewish villages in Lithuania (168 out of a total of 355) and 51.52 percent of the mass graves in the country (116 out of a total of 229). Here, as in the rest of Lithuania, Jewish-majority villages broadly follow the same distribution as Jewish-minority villages: They are usually found near farmland and are often located along streams or rivers.

The largest mass graves in this region are found near the largest cities, which also contained the largest populations of Jewish inhabitants. The largest grave, which holds an estimated 8,000 victims (*Holocaust Atlas of Lithuania*, n.d.), is near the city of Šiauliai, which had an estimated pre-War population of 6,600 Jews (JewishGen, n.d.-a), plus several hundred more Jewish residents in areas surrounding the city. The largest pre-War Jewish population, approximately 6,800 residents (JewishGen, n.d.-a), was found in the city of Panevėžys; the largest grave associated with that city and its surrounding Jewish villages contains 7,523 victims (*Holocaust Atlas of Lithuania*, n.d.).

Smaller graves, some holding a dozen victims or fewer, are scattered across the countryside.

2.3 Software

For this study, all analysis, including MaxEnt modelling (known as Presence Only Prediction in ArcGIS Pro), was conducted using Esri's © ArcGIS Pro, version 3.3.1. Microsoft Excel © for Microsoft 365, version 2407, was also used for data preparation before importing it into ArcGIS Pro.

2.4 Data

Data for this research was drawn from multiple sources based on its completeness and relevance to the country of Lithuania as it existed at the time of the German occupation. All datasets used in this research are described below and are summarized in Table 1.

2.4.1 Data Source: Mass Graves

The online "Holocaust Atlas of Lithuania" (*Holocaust Atlas of Lithuania*, n.d.) contains locations of known mass graves within the country; each killing event is linked to a precise latitude and longitude. In November 2023, the database, with 279 grave locations, was downloaded into an Excel spreadsheet; those entries with incomplete location information were discarded. The final dataset contained 229 entries, or 82 percent, of the original data. The Excel data was imported into ArcGIS Pro as a point feature class.

This **Mass Graves** point feature class will be used as the known presence points for the MaxEnt model; all other datasets will be used as environmental variables.

2.4.2 Data Sources: Jewish Heritage

The Nazis often killed and buried their victims in existing Jewish cemeteries (Pohl, 2015), which were located near Jewish communities. Killing sites were always easily traversed from Jewish communities (Desbois, 2015), making the locations of both existing Jewish cemeteries and Jewish communities a valuable source of information about mass grave locations.

2.4.2.1 Jewish Cemeteries

The Lithuanian non-profit organization MACEVA provided a dataset of Jewish cemeteries (M. Jakulytė-Vasil, personal communication, June 22, 2024). It contained 415 entries, which included several entries for mass graves. All mass graves entries and any cemeteries with incomplete latitude/longitude information were removed from the dataset, leaving a total of 203 cemetery locations which were imported into ArcGIS Pro and converted to a point feature class.

2.4.2.2 Jewish Communities

The names and locations of 422 Jewish communities were downloaded from the LitvakSIG website (LitvakSIG, 2022) in June 2024. Additionally, the website JewishGen.org (JewishGen, n.d.-b) has documented the existence of more than 240 Litvak settlements from the Interwar period (1918-1939); this data was downloaded in December 2023. The LitvakSIG and JewishGen datasets were merged in a tabular format; duplicate entries and entries with incomplete geographic data were deleted, leaving a total of 355 records of Litvak communities with valid latitude/longitude coordinates. This file was imported into ArcGIS Pro and converted to a point feature class.

2.4.3 Data Sources: Natural Environment

Land use and land cover have changed since the German occupation of Lithuania. In 1939, two years before the German invasion, 73 percent of the country was devoted to agriculture; 19 percent was covered by forests; and the remaining 8 percent consisted of "other" areas, such as urban and industrial areas (Central Intelligence Agency, 1958). By contrast, in 2018, 44.8 percent of the country was devoted to agriculture; 34.6 percent of the country was covered by forests; and the remaining 20.6 percent consisted of "other" areas (Central Intelligence Agency, 2024). Despite these changes, visual comparison with Soviet maps from the 1930s (Indiana University Bloomington Libraries, 2010) reveals a strong correlation between agricultural and forest land use then and now. Based on these figures and observations, it is reasonable to conclude that most "other" modern land use is derived from the destruction of agricultural land for the expansion of urban and industrial areas, and that some agricultural land has been converted into forest. Thus, modern land use data can still be reasonably used to estimate the historical locations, extent, and type of land use, especially in rural, lightly populated areas where the majority of killing events took place (Desbois, 2015).

2.4.3.1 Elevation, Slope, and Aspect

Killing sites needed to be easily accessible on foot, by cart, or, rarely, by truck, and were often located in low-lying areas (Desbois, 2015). Therefore, slope, elevation, and aspect were extracted and implemented in this study.

Elevation data at 25-meter resolution produced in 2016 by the European Union's (EU's) Copernicus missions and housed on the OpenDEM website (*OpenDEM*, n.d.) was used for this study. The EU has archived this data and it is not directly available through the EU's Copernicus Land Monitoring Service (CLMS) portal (*European Digital Elevation Model (EU-DEM)*, n.d.). OpenDEM was chosen as a data provider since the data currently available on the CLMS portal is at a slightly larger 30-meter resolution instead of the 25-meter resolution available with the 2016 data. The DEM file was converted to **Slope** and **Aspect** with a 25-meter resolution.

2.4.3.2 Water

Killing sites were often located in or near marshes or swamps. In addition, low-lying areas, such as those often surrounding Lithuanian lakes and rivers – or the rivers themselves – were frequently chosen for killing sites (Desbois, 2015; *Holocaust Atlas of Lithuania*, n.d.; Sturdy Colls, 2015).

Data on **Lakes** and **Rivers** was downloaded from the Lithuanian geoportal as part of the "Cadastre of rivers, lakes and ponds" (*Lietuvos Respublikos upių, ežerų ir tvenkinių kadastras (UETK)*, 2023); records reflecting coastal waters and purely man-made constructions, such as artificial canals and industrial lakes and ponds, were removed. The original dataset has a resolution of ten meters.

Data on **Wetlands** was downloaded from the European Union's INSPIRE Geoportal from within the Annex I. Hydrography dataset for Lithuania. Data on the locations of wetlands, bogs, and marshes in the country was used in its entirety (*INSPIRE Geoportal*, n.d.). The original dataset has a resolution of ten meters.

2.4.3.3 Forests

Forests are a well-documented location for Nazi killings (Pohl, 2015; Desbois, 2015; *Holocaust Atlas of Lithuania*, n.d.).

The Lithuanian forest cadaster (*Miškų kadastro duomenys*, 2024) contained a vector polygon file of all forest plots in the country. This dataset, named **Forest: Plots**, was used in its entirety and was converted to raster format with a resolution of twenty-five meters.

The Lithuanian geoportal also hosted open data from the EU INSPIRE Geoportal containing a 10-meter resolution categorical raster of the "Dominant Leaf Type" of Lithuanian forests, which overlapped exactly in area with the vector polygon file of forest plots. The **Forest: Dominant Leaf Type** raster provided additional information, distinguishing between deciduous trees and conifers (*Open Data - Geoportal.Lt*, n.d.). Even though **Forest: Plots** and **Forest: Dominant Leaf Type** would appear to demonstrate collinearity, both contributed individually to the final model.

2.4.3.4 Land Use

The Copernicus Land Monitoring Service (CLMS) provided a 10-meter resolution categorical raster of ten land use categories within Lithuania, part of the "Corine Land Cover + Backbone" inventory (Copernicus Land Monitoring Service, 2024). Part of this **Land Use** data included categories for the type of forests, similar to that found in **Forest: Dominant Leaf Type**. However, far from introducing collinearity, both sets of data contributed significantly to the final model.

2.4.4 Data Sources: Built Environment

The Germans depended heavily on the infrastructure and transportation capabilities of villages and other locations which harbored Jews and Jewish communities, especially in rural areas which comprised most of the country at the time. They required roads to traverse the countryside from one village to the next. They required local villagers as conscripted labor both at the gravesites and within the villages – digging graves; transporting victims; ferrying supplies, such as bullets; cooking; producing tools, fencing, carts, and other equipment (Desbois, 2015). These activities required both the resources of the villages and quick and easy access between villages and killing sites.

Using modern data on the human-constructed environment to analyze events which took place more than eighty years ago presents some challenges: Urban and sub-urban development have changed settlement and farming patterns; infrastructure improvements have created roads and made others obsolete. While no modern datasets could fully recreate the built environment of World War II-era Lithuania, a comparison of modern data with Soviet maps from the 1930s (Indiana University Bloomington Libraries, 2010) reveals that there is still significant overlap between Lithuanian land use and infrastructure then and now, particularly in rural areas.

Since land use, roads, and settlement patterns were all so integral to the Nazis' operations, the following modern datasets have been included in this study even though they do not completely capture the nuances of Lithuanian infrastructure under German rule.

2.4.4.1 Modern Towns and Villages

Data on modern towns and villages in Lithuania was taken from Open Street Map (OSM) (*OpenStreetMap*, n.d.), which provided extensive information on inhabited areas in Lithuania. The OSM data was filtered to remove all locations with fewer than 50 inhabitants.

2.4.4.2 Roads

Data on the road network was also taken from Open Street Map (OSM) (*OpenStreetMap*, n.d.), which permitted a fine-grained approach to selecting and filtering records to more closely match the road network existing in Interwar Lithuania as seen in Soviet topographic maps from the late 1930s (Indiana University Bloomington Libraries, 2010). Based on the hyper-local nature of the Nazi killings in the East described by Desbois (2015), no map could adequately capture the dense web of foot and cart paths used for transport of the victims. However, the OSM road network can provide a fair approximation of accessible locations and of distances travelled in general. Roads were selected based on their OSM classification for those road categories which best correlated with the Soviet maps.

2.5 Data Pre-Processing

All data used within ArcGIS Pro was projected into the Lithuanian Coordinate System (LKS) 1994 geographic coordinate system and the Lithuania Transverse Mercator projected coordinate system (hereafter abbreviated LKS 94/Lithuania TM), which is specialized for use in Lithuania, and which matches the projected coordinate system of data provided by the Lithuanian government's geoportal. The Baltic 1986 vertical datum and vertical coordinate system were used for elevation.

The digital elevation model (DEM) from OpenDEM (*OpenDEM*, n.d.) had the largest ground sample distance at 25 meters, so this parameter was used for all rasters.
In the absence of background points, which this data did not contain, the Presence Only Prediction tool in ArcGIS Pro, which will be used for this study, requires the use of either continuous or categorical rasters as explanatory variable inputs; the tool then generates its own background points (see *Section 2.6 Data Analysis: MaxEnt Modelling*). All non-categorical rasters except **Slope**, **Elevation**, and **Aspect** were converted to distance accumulation raster datasets at 25-meter resolution. The two categorical variables – **Land Use** and **Forests: Dominant Leaf Type** – were already in an acceptable format. These two categorical rasters were resampled from their existing 10-meter resolution to a 25-meter resolution to match the pixel size values of the DEM file.

Table 1: Explanatory training data used for MaxEnt model. The columns provide the following information: Explanatory Variables: descriptive name for the dataset used throughout; Original Data Format: format in which the data was acquired or originally manipulated; Processed Format as Used in Model: method of data preparation for input into the MaxEnt model; Relative Dates of Data Source Origins: Identifies the era in which the data originated; Source: Online or other source for the dataset or its precursor. Where possible, links provide direct access to the metadata for the data set.

Explanatory Variables *(raster format)*	**Original Data Format**	**Processed Format as Used in Model**	**Relative Dates of Data Origins**	**Source**
Mass Graves[5]	Tabular; converted to Vector: Point	(Input point features; Vector: Point)	1941-1942 (digitized in the 2000s)	Holocaust Atlas of Lithuania
Jewish Cemeteries	Tabular; converted to Vector: Point	Distance Accumulation Raster	Pre-1941 (digitized in the 2000s)	MACEVA, Lithuanian Jewish non-profit[6]
Jewish Communities	Tabular; converted to Vector: Point	Distance Accumulation Raster	Pre-1941 (digitized in the 2000s)	LitvakSIG communities database; JewishGen communities database
Elevation[7]	Raster	Calculated from Digital Elevation Model (DEM)	2016	Copernicus mission/European Space Agency (ESA) (obtained from OpenDEM website)
Slope[7]	Raster	Calculated from DEM	2016	Copernicus mission/ESA (obtained from OpenDEM website)
Aspect[7]	Raster	Calculated from DEM	2016	Copernicus mission/ESA (obtained from OpenDEM website)
Lakes	Vector: Polygon	Distance Accumulation Raster	2023	Lithuanian Geoportal, Cadaster reference dataset
Rivers	Vector: Line	Distance Accumulation Raster	2023	Lithuanian Geoportal, Cadaster reference dataset

[5] All explanatory data for training has been converted to raster format; the Mass Graves dataset, which is also mentioned here, represents the vector point information of the input locations of all known mass graves. It is included here for completeness.
[6] This dataset is not available online. It was obtained through personal request (M. Jakulytė-Vasil, personal communication, June 22, 2024).
[7] Information derived from this dataset was produced using Copernicus data and information funded by the European Union.

Wetlands	Vector: Polygon	Distance Accumulation Raster	2024	EU INSPIRE Geoportal – Annex I. Hydrography: Lithuania
Forest: Plots	Vector: Polygon	Distance Accumulation Raster	2018	Lithuanian Geoportal, Forest cadaster dataset
Forest: Dominant leaf type	Raster	Categorical raster	2018	Lithuanian Geoportal, Open data
Land Use[8]	Raster	Categorical raster	2021	Copernicus Land Monitoring Service (CLMS)
Towns & Villages	Vector: Point	Distance Accumulation Raster	2024	OSM (Downloaded through GeoFabrik)[9]
Roads	Vector: Line	Distance Accumulation Raster	2024	Open Street Map (OSM) (Downloaded through GeoFabrik)[9]

[8] Information derived from this dataset has been prepared using the European Union's Copernicus Land Monitoring Service information with funding by the European Union.
[9] This layer has been created from OpenStreetMap data and is licensed under the Open Database 1.0 License. This file contains OpenStreetMap data as of 2024-07-26T20:21:00Z

2.6 Data Analysis: MaxEnt Modelling

The machine learning model MaxEnt, for evaluating maximum entropy, was implemented in this analysis. MaxEnt accepts a set of known presence points within a defined study area as inputs and compares the geospatial characteristics of those input points against the geospatial characteristics of background points in the study area, where the fact of presence is unknown. The geospatial characteristics of both the input points and the background points are compiled from a series of explanatory training variable rasters. When using MaxEnt, it is not necessary to know locations of absence (Merow et al., 2013). The MaxEnt model has shown its ability to discriminate between ancient and modern land use patterns, even in disturbed contexts (Galletti et al., 2013; Rafuse, 2021), and works well in archaeological contexts even with very small sample sizes – Gillespie et al. (2016) supplied only 29 data points, and Rafuse (2021) supplied only 66. These characteristics make it an ideal choice for this study: The Lithuanian mass graves data set is relatively small (229 records); absence points are not known with any certainty; and modern land use patterns have impinged on conditions which prevailed during World War II.

This study will use the point locations of **Mass Graves** as the known presence points; the remaining thirteen rasters represented in Table 1 will serve as the explanatory variable rasters for training and prediction.

2.6.1 Training: Parameters and Data

To prepare the data, the MaxEnt tool transforms the explanatory training rasters using the set of basis functions selected by the user. For this study, the following basis functions were chosen: Linear, Quadratic, Product, and Smoothed step (Hinge). The selection of multiple basis functions allows the tool to explore the maximum number of data combinations, permitting the most relevant combinations to stand out in the results (Fitzgibbon et al., 2022). The Hinge parameter separates the data into static segments of all zeroes or ones and increasing or decreasing segments defined by a linear function. These segments are separated by a threshold point, or knot. A forward hinge starts from a series of zeroes between the minimum value and the threshold (knot) value followed by an increasing linear function up to the maximum value. A reverse hinge starts with an increasing linear function from the minimum value to the threshold (knot) value, followed by a series of ones between the knot and the maximum value (*How Presence-Only Prediction (MaxEnt) Works*, n.d.). For this study, 10 threshold points were defined.

The study area was limited to the five northern counties (Klaipėdos, Telšiu, Šiauliu, Panevėžio, and Utenos). To reduce sampling bias, spatial thinning at a nearest neighbor distance of 500 meters was applied; the model was assigned 10 iterations to find the optimal thinning solution.

The **Mass Graves** presence point features were evaluated at a relative weight of 100 compared to the system-generated background points, meaning that the background points are not locations of confirmed absence but are locations where the fact of presence is not known (*How Presence-Only Prediction (MaxEnt) Works*, n.d.). And, since the location of the presence points (**Mass Graves**) never changes, the C-log-log transformation was chosen for converting outputs to a probability of presence.

The user-defined probability cutoff point was set at 0.15: Locations assigned a probability greater than 0.15 were classified as “presence” locations; locations assigned a probability lower than 0.15 were classified as “background” locations, i.e. locations where there was likely no presence. This value provided the best balance between an optimal AUC value and a low omission rate while also running the least risk of inadvertently missing potential presence points. This cutoff value also best balanced the number of background points classified as potential presence points. Table 5 presents detailed results of the trials for determining the optimum cutoff value.

Since spatial thinning was used, the tool was able to accept the input point file **Mass Graves** as the feature class which contained the locations where predictions will be made: The model was trained on the **Mass Graves** data in the five northern counties; predictions were made for the entire **Mass Graves** dataset.

For validation of the training model, the Presence-only Prediction tool uses a *k*-fold cross-validation scheme in which *k* number of random groups are formed from the training data, each group containing a roughly equal number of features. Each group (in this study, $k = 10$) is used as a validation subset while the remaining groups are used as a training subset. The resulting table of values can help in the evaluation of the model's quality.

These parameters are summarized in Table 2.

Table 2: MaxEnt model parameters, user-defined selections, and the rationale for those selections.

Parameter	User-defined selection	Rationale
Basis Functions	Linear	One of Esri's suggestions (Fitzgibbon et al., 2022) is to use multiple basis functions allowing the model to self-adjust so that all combinations of functions can be considered and the most important function(s) will most strongly influence the model.
	Quadratic	
	Product	
	Smoothed step (Hinge) *(10 knots)*	
Study Area Polygon	(5 northern counties)	Selecting a subset of data for training allows the model to be tested by making location predictions in new areas.

Spatial Thinning	500 meters	Given the hyper-local nature of the killings, this thinning distance provided a balance between reducing spatial bias and evaluating as many locations as possible.
Number of Thinning Iterations	10	The default value was sufficient for this data.
Relative Weight of Presence to Background	100	A value of 100 indicates that the provided presence points are the only source of information on presence locations.
Presence Probability Transformation	C-log-log	This function is recommended when presence points are stationary (as a plant or mass grave) rather than ambulatory (as a migrating species).
Presence Probability Cutoff	0.15	Provided both an optimal AUC value (0.9417) and omission rate (0.0690). See Table 5 for more detailed information.
Input Prediction Features	Mass Graves dataset	Permits predictions across the entire dataset.
Cross-validation	Random; 10 groups	Cross-validation provided an additional approach for evaluating the model; a maximum of 10 groups was permitted.

2.6.2 Model Validation

To evaluate model performance, the MaxEnt tool within ArcGIS Pro provides several metrics. Two of these metrics are values for the Area Under the Curve (AUC) and values for omission rates.

The AUC refers to the area under the Receiver Operating Characteristic (ROC) curve which displays model performance across all cutoff thresholds and provides a single metric for comparing multiple versions of the model (Phillips et al., 2006). An AUC value closer to 1 indicates optimal model performance (Liu, n.d.); that is, the model can perfectly distinguish between true presence points and true background, or random, points. An AUC of 0.5 indicates the model performs no better than random.

The omission rate refers to the proportion of true presence points mistakenly classified as absence. A high omission rate indicates that the tool is incorrectly identifying presence points as absence points. A low omission rate is preferred, since that indicates that more true presence points are being correctly identified as presence points (Liu, n.d.). The complement of the omission rate is reflected in the percentage of correctly identified presence points.

The percentage of background points classified as potential presence can provide additional insight into how precisely the model is identifying potential presence locations. A high percentage of background points classified as potential presence suggests that the model could be refined to more precisely target potential presence points. A lower percentage of background points classified as potential presence suggests that the variables training the model are permitting a more fine-grained evaluation of what constitutes presence.

Taken together, the measures for AUC, omission rate, percentage of correctly identified presence points, and the percentage of background points classified as potential presence can provide a firm context for evaluating the overall performance of the MaxEnt model.

One additional metric provided by the Presence Only Prediction tool in ArcGIS Pro is a list of variables and combinations of variables that were used in each of the basis functions and which also had an impact on the model. This list also includes the values of their regression coefficients. In the ArcGIS tool, unlike in standard statistical regression, an overall "coefficient budget" is implemented, and the coefficients of explanatory variables with little or no impact on the final outcome are reduced to zero and are eliminated entirely from the model. Some coefficients may be quite small, represented as +/-0.0000 in the results; however, the fact that the variable appears in the list at all indicates that it had an impact on the final model. Using a "coefficient budget" also helps address multicollinearity: The coefficients of variables exhibiting multicollinearity are more likely to be reduced to zero and be removed from the model (*How*

Presence-Only Prediction (MaxEnt) Works, n.d.). All variables listed in Table 1 and used in the models below appeared in the list of "impactful" variables for their respective models.

2.6.3 Model Runs

Four runs of the MaxEnt model were conducted, each using the same tool settings as detailed in *Section 2.6 Data Analysis: MaxEnt Modelling*; the only variation was in the use of different explanatory variable data sets. In all cases, the **Mass Graves** dataset was used as presence points. The results of each model run are described below. For the best-performing model, the final product, a probability distribution map, is also described. Table 3 contains a list of all models with the variables they contain.

Table 3: A list of the four models used in this study with a description of the variables each contains.

Model Number	Model Name	Variables Included in the Model
1	Contemporaneous Data	Slope Aspect Elevation Jewish Cemeteries Jewish Communities
2	Natural Water Features	***Model 1*** + Lakes Rivers Wetlands
3	Land Use and Land Cover	***Model 2*** + Forest: Plots Forest: Dominant Leaf Type Land Use
4	Towns & Villages; Roads	***Model 3*** + Towns & Villages Roads

2.6.3.1 Model 1: Contemporaneous Data

In Model 1 the model was trained, and predictions were made, using only datasets which could be considered contemporaneous with the Mass Graves dataset: **Slope**, **Aspect**, **Elevation**, **Jewish Cemeteries**, **Jewish Communities**. All these datasets contain known data which, while possibly incomplete, is unlikely to have changed since 1941.

2.6.3.2 Model 2: Natural Water Features

Model 2 included all data from Model 1 and added the three datasets containing **Lakes**, **Rivers**, and **Wetlands** – these are all natural features which may have been altered somewhat by humans in the past eight decades, but which are unlikely to have undergone complete revision, and so can still be considered largely contemporaneous with the **Mass Graves** dataset.

2.6.3.3 Model 3: Land Cover and Land Use

In Model 3, **Forest: Plots**, **Forest: Dominant Leaf Type**, and **Land Use** datasets were added – these land use patterns have all been influenced by human activity but share strong visual overlap with similar data from the 1930s. They cannot be considered fully contemporaneous with the **Mass Graves** dataset, but can provide valuable information, nonetheless.

2.6.3.4 Model 4: Towns & Villages; Roads.

Finally, in Model 4, data for **Towns & Villages** and **Roads** was added in. Since complete WWII-era information for these features is not available, and since they are wholly human constructs, they were considered potentially the least reliable in terms of re-creating conditions which prevailed in 1941. However, these variables had a significant impact on the results of the model.

Chapter 3: Results

3.1 Consideration of Multiple Models

The results for each of the four model runs are discussed below.

3.1.1 Model 1: Contemporaneous Data

Model 1 had an AUC value of 0.8328, indicating strong model performance, and an omission rate of 0.0431, suggesting that almost all presence points are being correctly identified; this is reflected in the percentage of correctly identified presence points, which was 95.69 percent. However, this model also identifies more than half (51.7 percent) of the background points as potential presence points, suggesting that the model has a very low threshold for determining what qualifies as a presence point.

The full results of each model run are described in Table 4.

3.1.2 Model 2: Natural Water Features

The addition of **Lakes**, **Rivers**, and **Wetlands** in Model 2 contributed the least to model performance. The AUC rose only slightly, to a value of 0.8429. The omission rate grew larger, to 0.0862, indicating that Model 2 was omitting actual presence points at a higher rate than Model 1. This is also seen in the lower percentage of correctly identified presence points: Model 2 achieved an accuracy of only 91.38 percent in this respect. However, the number of background points identified as potential presence dropped to 47.74 percent, suggesting that Model 2 can more finely discriminate between background and presence points. See Table 4.

3.1.3 Model 3: Land Cover and Land Use

Adding the land cover and land use datasets in Model 3 had a dramatic impact on the results. The AUC value jumped to 0.9178, indicating superlative model performance. The omission rate dropped to 0.0603; that is, 93.97 percent of all presence points were correctly identified. The percentage of background points classified as potential presence points also dropped to 27.61 percent, suggesting that Model 3 can finely discriminate between background points and potential presence points. See Table 4.

3.1.4 Model 4: Towns & Villages and Roads

In Model 4, the AUC rose to 0.9417, suggesting outstanding model performance. The omission rate also rose slightly to 0.0690; meaning that 93.1 percent of presence points were correctly identified. Model 4 also had the ability to very finely distinguish between background and potential presence points: Only 19.12 percent of the background points were identified as potential presence points.

3.2 Evaluating the Models

All four models performed well. The lowest AUC value of 0.8328 in Model 1 is well within an acceptable range for a well-performing model, and the omission rate of 0.0431 indicates that very few true presence points are being omitted. The highest AUC value of 0.9417, for Model 4, indicates extremely robust model performance, and the omission rate of 0.0690 is still within one standard deviation (+/- 0.0155) of the average of 0.0647 for all model runs.

The performance metrics for all four models are summarized in Table 4.

Table 4: Summary of performance metrics for the four models. Note that each model contains the explanatory variables for every model preceding it.

Model #	**Explanatory Variables**	**AUC**	**Omission rate**	**Percentage of correctly identified presence points**	**Percentage of background points classified as potential presence**
1	**Aspect** **Elevation** **Slope** **Jewish Communities** **Jewish Cemeteries**	0.8328	0.0431	95.69	51.7
2	*Model 1 Variables +* **Lakes** **Rivers** **Wetland**	0.8429	0.0862	91.38	47.74
3	*Model 2 Variables +* **Forest: Plots** **Forest: Dominant Leaf Type** **Land Use**	0.9178	0.0603	93.97	27.61
4	*Model 3 Variables +* **Towns & Villages** **Roads**	0.9417	0.0690	93.1	19.12
Averages of all model values:		***0.8838***	***0.0647***	***93.54***	***36.54***
Standard Deviation:		***+/- 0.0469***	***+/- 0.0155***	***+/- 1.55***	***+/- 13.59***

The best performance overall came from Model 4, which included all 13 explanatory variables. Model 4 had the highest AUC value, and the lowest percentage of background points classified as potential presence. Model 4's omission rate of 0.0690 meant that of the 116 mass graves in the training data, the model misclassified only 8. The AUC (Figure 3a) and omission rates (Figure 3b) for Model 4 can be seen graphically in Figure 3. Model 4 was also able to very finely distinguish between background points and potential presence points, particularly when Model 4 is compared with the other three models. Model 1 classifies more than half of background points as potential presence; the tradeoff for Model 1's very low omission rate (0.0431) is counterbalanced by the enormous amount additional search area that was identified. In contrast, Model 4 identified 19.12 percent of background points as potential presence points, greatly reducing the possible search area while still ensuring low omission rates. In comparison to Model 4, all other models also had lower AUC values, particularly Models 1 and 2; AUC values jumped significantly for the Model 3 results and jumped again for Model 4, suggesting that the added data contributed significantly to the overall model's prediction capabilities.

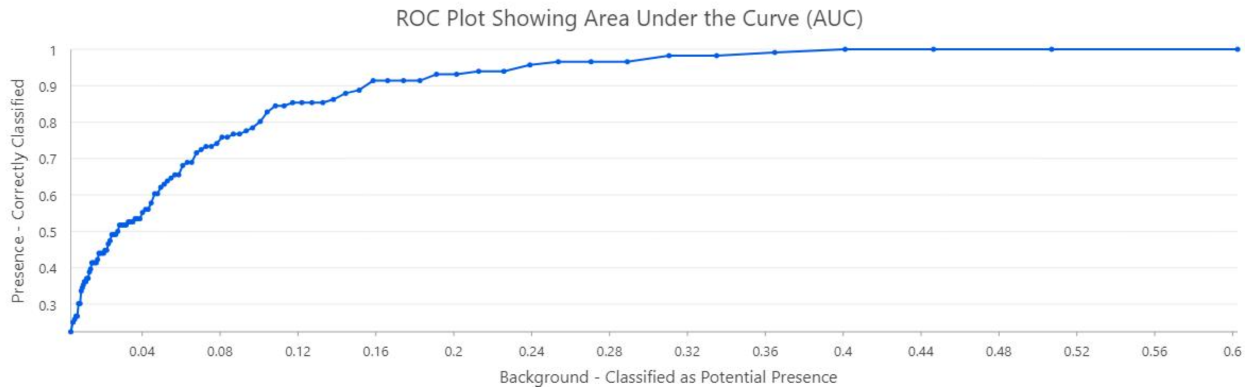

Figure 3a: AUC for Model 4.

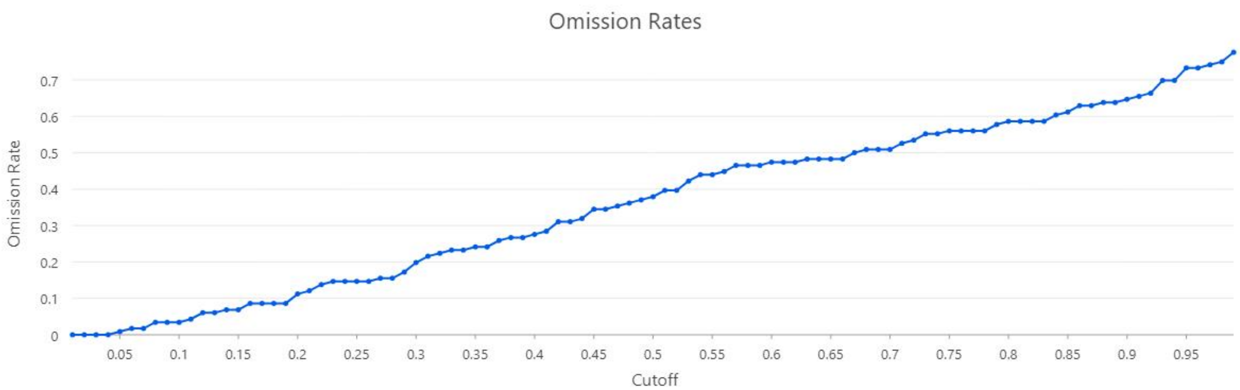

Figure 3b: Omission rates for Model 4.

Figure 3: AUC (Figure 3a) and omission rate plots (Figure 3b) for Model 4.

Table 5 shows the Area Under the Curve (AUC), omission rates, and the percentage of background points classified as potential presence for the trials evaluating various cutoff thresholds using the datasets for Model 4; the omission rates in this table can be compared with omission rates reported in Figure 3.

Table 5: Area Under the Curve (AUC), omission rates, and percentage of background points classified as potential presence for the trials evaluating various cutoff thresholds for Model 4. A cutoff threshold of 0.15 best balances all three indicators.

Presence Probability Cutoff Value	**Area Under the Curve (AUC)**	**Omission Rate**	**Percentage of Background points classified as potential presence**
0.5 *(default)*	0.9419	0.3793	4.89

0.4	0.9419	0.2931	6.93
0.25	0.9417	0.1466	12.20
0.15 ***(selected for the model)***	**0.9417**	**0.0690**	**19.12**
0.12	0.9437	0.0517	22.00
0.10	0.9419	0.0345	25.26

Although the regression coefficients represented in the Presence Only Prediction Tool do not conform to the same rules as standard regression coefficients in that all coefficients combined must conform to a single coefficient budget (*How Presence-Only Prediction (MaxEnt) Works*, n.d.), examining the relationship between the various reported coefficients can provide insight into which variables, or which combinations of variables, are contributing most to the model. Of the 55 regression coefficients reported in the results, most of them are reported to four decimal places as +/- 0.0000. However, several coefficients, or combinations of coefficients within different basis functions, reported values above +/-0.0000; variables with a value greater than +/-0.0500 and their potential contributions are summarized in Table 6. Please note that while not all variables appear in this list, all variables used in this model do contribute to the final results. The Appendix contains a complete list of all variables that are reflected in the final model, along with their accompanying regression coefficients.

Table 6: Regression coefficients above 0.05 for variables which contributed to Model 4, which contains all 13 variables.

Basis Function	Variable	Hinge Range OR Category Value (if applicable)	Coefficient
Linear	Slope		0.0898
Smoothed Step (Hinge)	Aspect	319.68 – 359.76 degrees (forward hinge)	-0.3131
Smoothed Step (Hinge)	Elevation	-6.30 – 25.83 m (reverse hinge)	1.0637

Smoothed Step (Hinge)	Slope	0.00 – 2.65 % (reverse hinge)	0.1432
Smoothed Step (Hinge)	Slope	0.00 – 5.29 % (reverse hinge)	0.0709
Smoothed Step (Hinge)	Jewish Cemeteries	0.00 – 3,040.35 m (reverse hinge)	-1.9778
Smoothed Step (Hinge)	Forest: Plots	2,048.25 – 2,304.28 m (forward hinge)	-3.6027
Smoothed Step (Hinge)	Towns & Villages	0.0 – 1,199.77 m (reverse hinge)	0.3645
Smoothed Step (Hinge)	Jewish Communities	0.0 – 2,462.75 m (reverse hinge)	-0.9108
Smoothed Step (Hinge)	Roads	0.0 – 310.65 m (reverse hinge)	-0.7071
Categorical	Forest: Dominant Leaf Type	Not forested	-0.2065
Categorical	Forest: Dominant Leaf Type	Coniferous	0.0535
Categorical	Land Use	Sealed	0.8221
Categorical	Land Use	Woody: Needle-leaved Trees	1.5923
Categorical	Land Use	Woody: Broadleaf Deciduous Trees	0.9309
Categorical	Land Use	Permanent Herbaceous	0.2111
Categorical	Land Use	Water	0.3330

Figure 4 is a visual representation of the partial response of each variable's effect on predicting presence across the full range of the variable. **Aspects** between 100 – 250 degrees have the highest probability of presence; the contributions of **Elevation** are greatest between approximately sea level and the country's average elevation of 110 meters. **Slopes** greater than approximately 12 percent approach a probability of presence of 1. Within a few hundred meters of **Jewish Cemeteries**, **Forest: Plots**, **Towns & Villages**, **Jewish Communities**, and **Roads**, the probability of presence generally ranges from 0.8 to 0.9, but drops off uniformly and precipitously with distance from these locations.

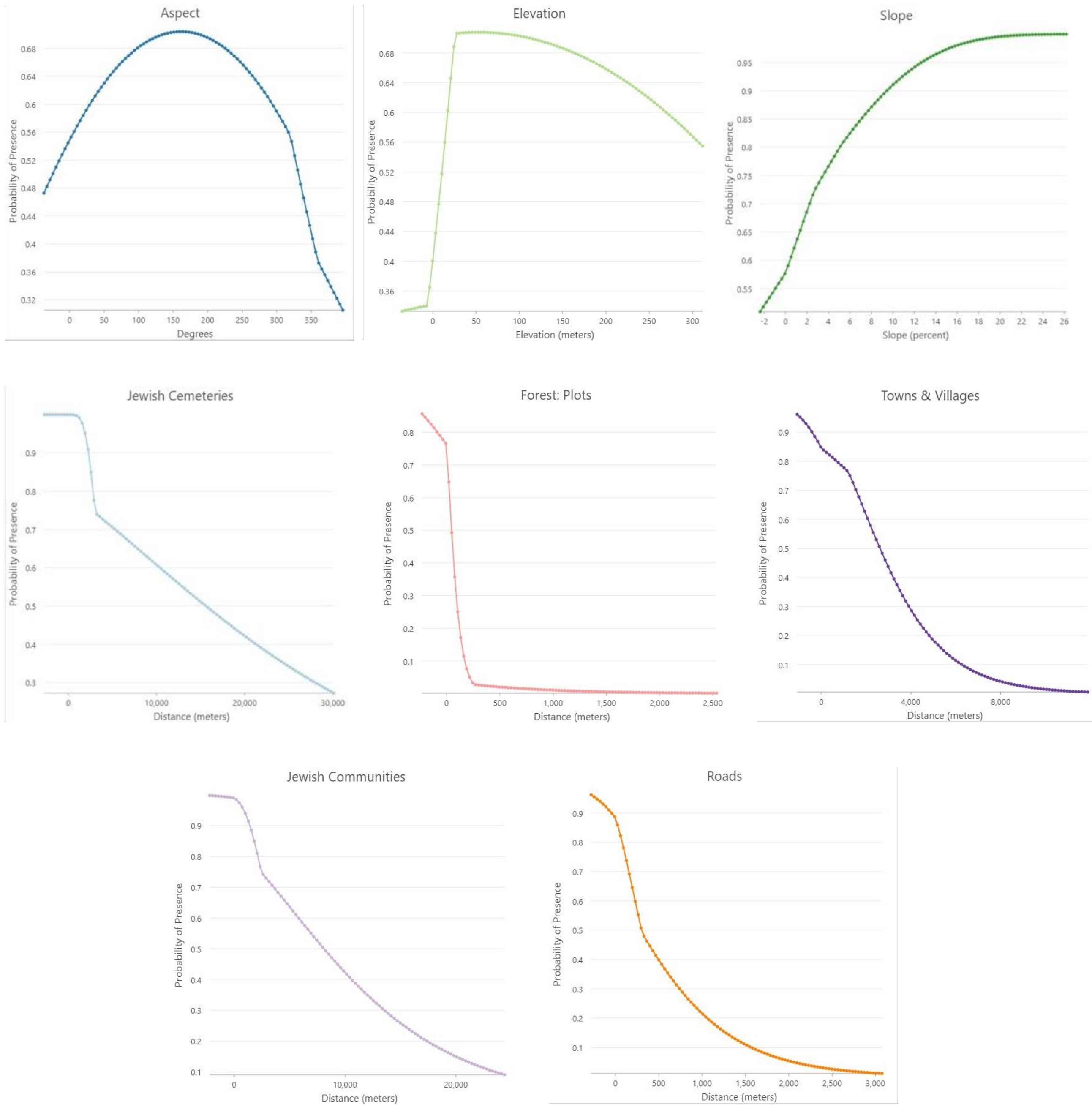

Figure 4: Partial response of continuous variables for those variables referenced in Table 4. Probability of presence drops off significantly with distance from cemeteries, forests, towns & villages, Jewish communities, and roads.

The probability distribution map for the training data is shown in Figure 5. Locations with a high probability of presence show a strong tendency towards clustering, visually aligning with the patterns of inhabited areas within the country; clustering along major water courses and roads and within forested areas is also evident. Probability distribution for predicted locations, illustrated in Figure 6, shows similar patterns.

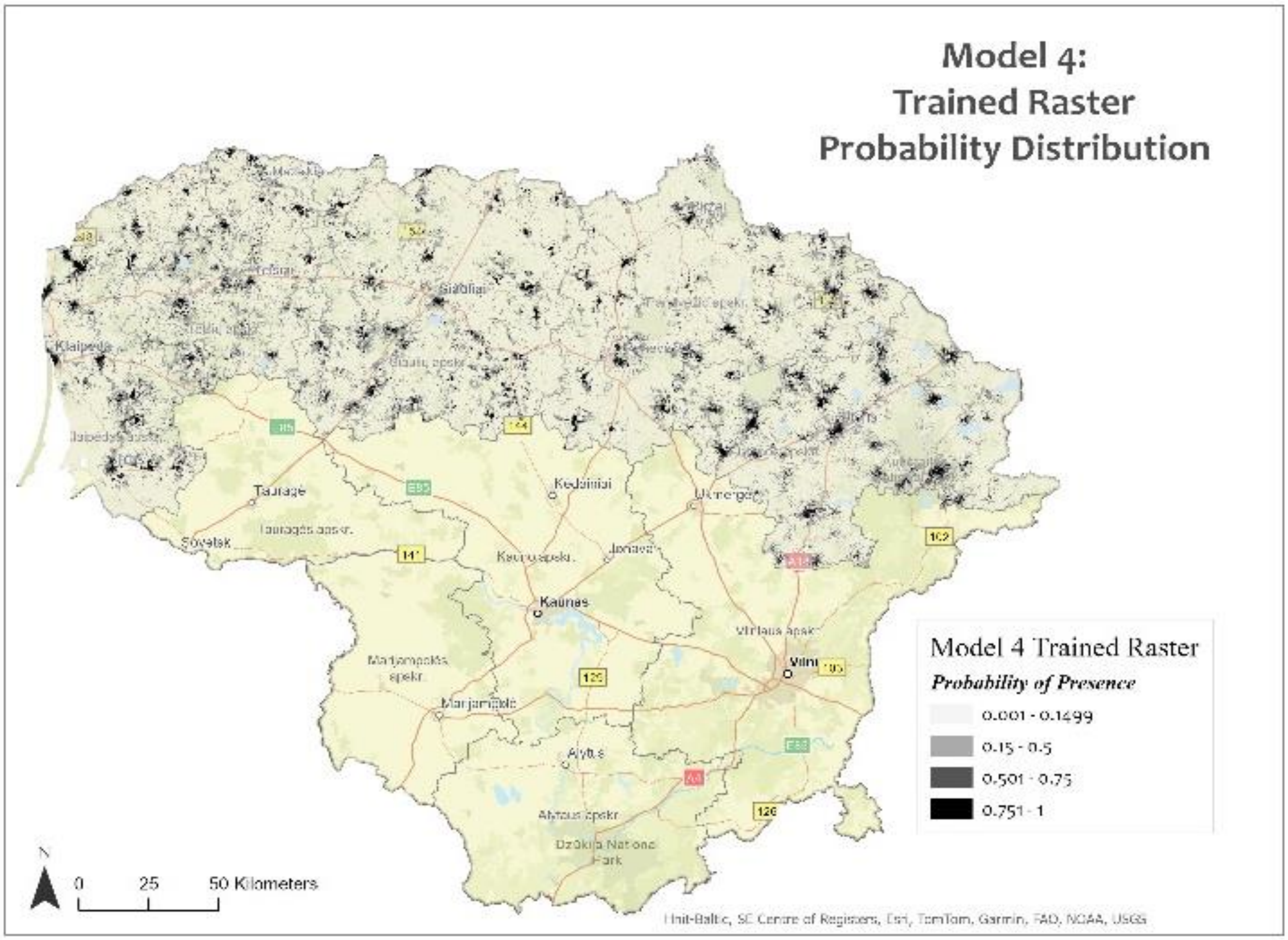

Figure 5: Probability distribution map showing the probability of presence for mass graves as determined in the training data. The locations with highest probability of presence are generally clustered around towns and villages, along watercoursesand roads, and within forested areas.

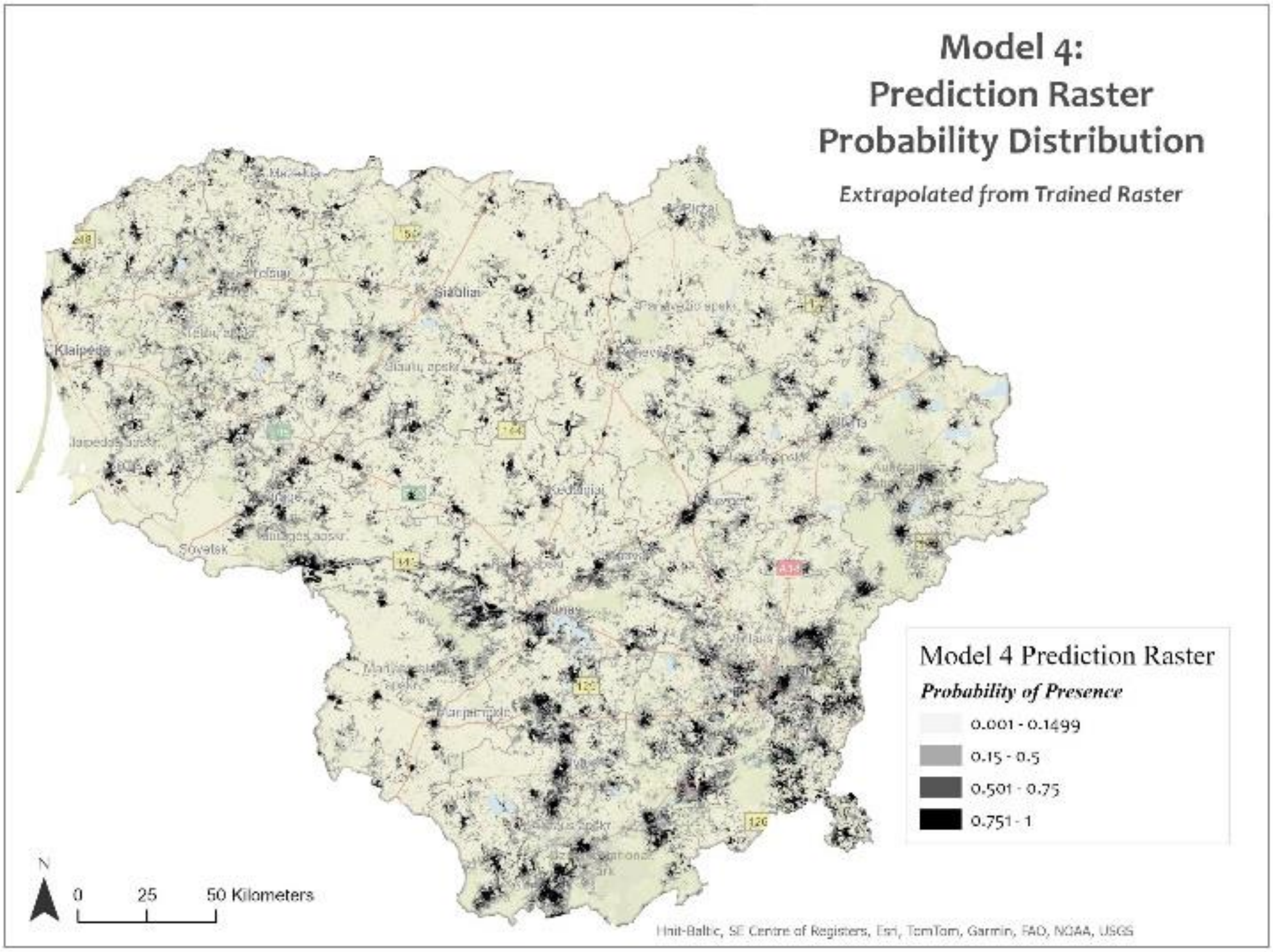

Figure 6: Probability distribution map showing the probability of presence for mass graves as determined in the prediction data. As in the training data (Figure 5), the locations with highest probability of presence are generally clustered around towns and villages, along watercourses and roads, and within forested areas.

Figure 7 shows the model prediction results as applied to actual mass grave locations; the model shows strong predictive capabilities throughout the country; failed predictions are scattered irregularly throughout. Figure 8 displays a histogram of this data, providing a numerical assessment of the map. Of the 229 known mass graves, and given a presence probability cutoff point of 0.15, the model correctly predicted the locations of 207 graves and failed to predict locations for 22 graves.

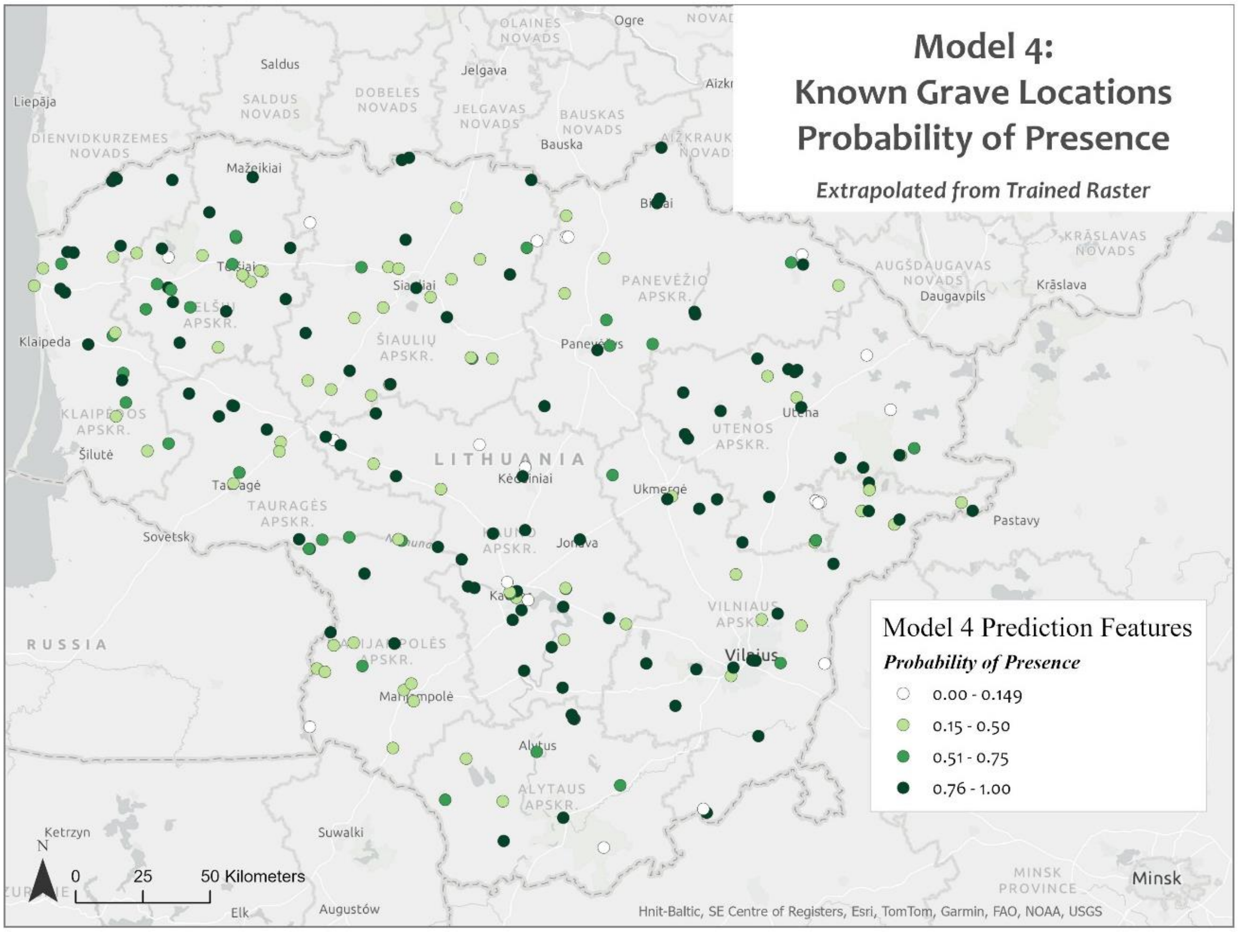

Figure 7: Prediction model applied to known mass graves (probability of presence cutoff point of 0.15). Correctly identified graves are found throughout the country; actual graves which were omitted from the predictions are also scattered throughout.

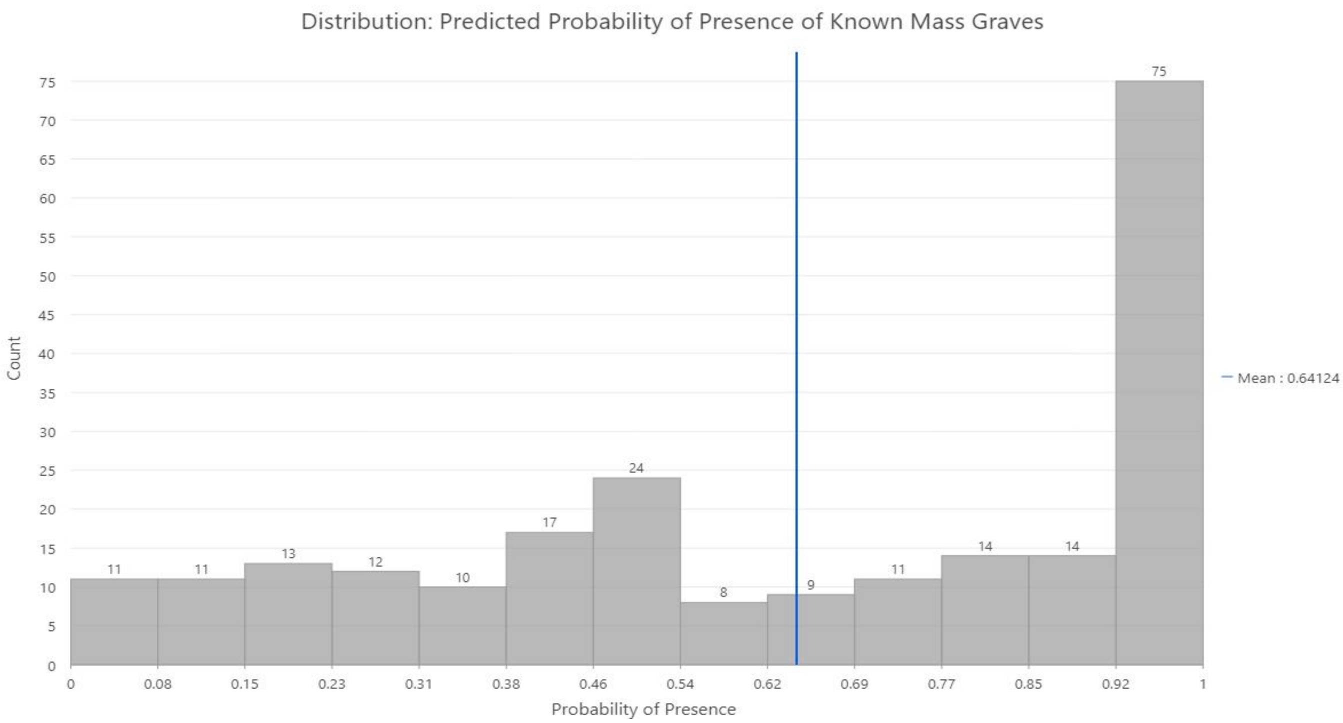

Figure 8: Histogram of the data represented in Figure 7. Of the 229 known mass graves, the model correctly predicted the locations of 207 graves and failed to predict locations for 22 graves (presence probability cutoff point of 0.15).

Chapter 4: Discussion

This research showed that the MaxEnt model's accuracy is highest in predicting mass graves when all 13 environmental variables were included (Model 4: Aspect, Elevation, Slope, Jewish Communities, Jewish Cemeteries, Lakes, Rivers, Wetlands, forest variables, Land Use, Towns & Villages, and Roads). This model correctly predicted the locations of 207 out of a total of 229 mass graves in the country of Lithuania; the AUC value of 0.9417 and omission rate of 0.0690 for this model are both indicative of an extremely robust, high-performing model. Model 4 also drew a sharp distinction between background points and potential presence points, identifying only 19.12 percent of background points as potential presence points. Although this does not guarantee that the model can always exhibit pinpoint accuracy, it can provide a slim profile of potential locations to investigate using more resource-intensive methods such as lidar, ground penetrating radar, or field surveys.

4.1 Recent mass grave discoveries in Lithuania

Since 2018, two previously undocumented graves, which were contained neither in the online Holocaust Atlas nor in the Yahad-in Unum archives (*Holocaust Atlas of Lithuania*, n.d.; *Yahad - in Unum Map*, n.d.), have been documented in Lithuania using ground penetrating radar. Fuerstenberg et al. (2018) documented a likely mass grave containing 28 victims in a forest near the town of Rokiškis in the Panevėžio *apskritis.* Reeder et al. (2024) documented a likely mass grave containing 40 to 60 victims in the Zaliakalnis Jewish Cemetery in the city of Kaunas. In both cases, researchers were led to the area based on survivor or witness testimony. The grave near Rokiškis was in a densely forested area and vegetation had to be cleared before surveys

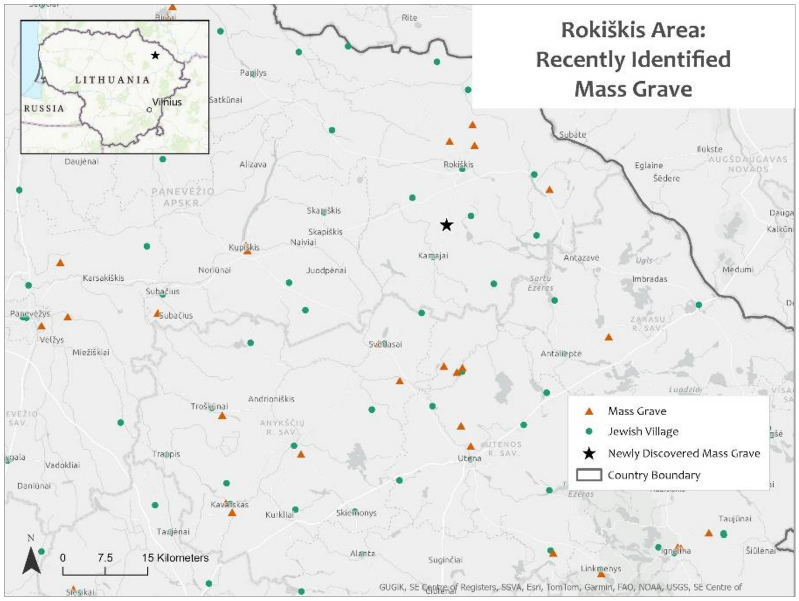

Figure 9a: Likely mass grave near Rokiškis.

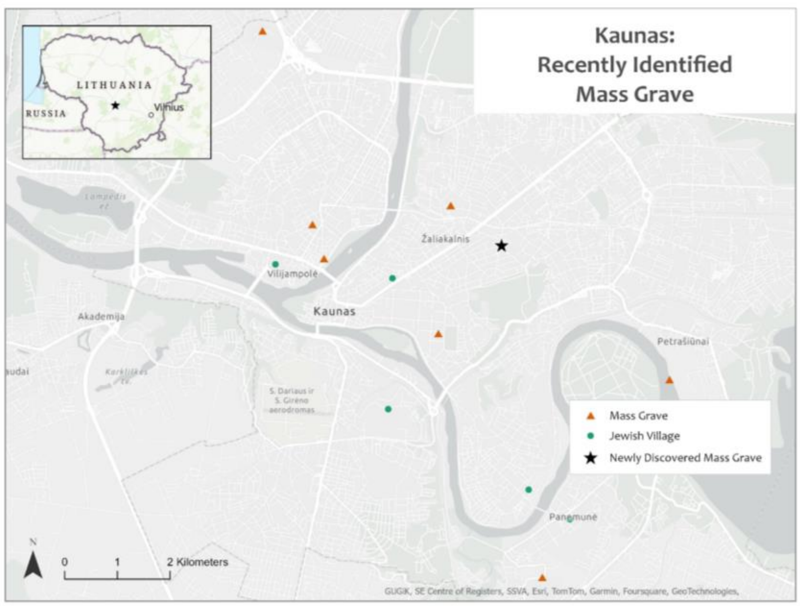

Figure 9b: Likely mass grave near Kaunas.

Figure 9: Locations of recently discovered likely mass graves in Rokiškis (Figure 9a), showing the regional distribution of mass graves and Jewish villages, and in Kaunas (Figure 9b), showing the local distribution of mass graves and Jewish villages. Please note the differences in scale.

could be undertaken (Beck et al., 2019); it is unlikely to have been found in a ground survey. Although Kaunas has many well-documented locations for killings and large mass graves, a mass burial in the Jewish cemetery was previously undocumented. However, both grave locations were accounted for in Model 4. See Figure 9 for a map of the area surrounding each likely mass grave.

The probability of presence of the grave near Rokiškis was above the cutoff point of 0.15, falling in the range between 0.15 – 0.5. The specific Rokiškis grave location had a probability of presence of 0.299159. The grave in Kaunas was in an area predicted to have a probability of presence between 0.75 – 1; the specific Kaunas grave location had a probability of presence of 0.991743.

4.2 Current obstacles to mass grave research

Current approaches to researching and discovering locations of mass graves of the Holocaust are not uniform. Primary sources on the Holocaust are scattered in archives and collections across Europe and the world. The European Holocaust Research Infrastructure (EHRI) is attempting to catalog and document all relevant archival holdings (*EHRI*, n.d.), and there have been efforts to integrate, elucidate, and standardize EHRI's collection of information to make it more accessible and transparent to researchers (Blanke et al., 2017; Blanke & Kristel, 2013), but challenges involving far-flung locations; poorly organized and faded records; and multiple languages, remain.

The horrors and trauma of the Holocaust are still part of active and living memory for survivors, collaborators, witnesses, family members, and others, likely preventing the development of Holocaust-era archaeology as a well-defined sub-discipline of conflict archaeology (Sturdy Colls, 2015), and most recent archaeological explorations of mass graves are conducted ad hoc based on foreknowledge of possible locations (Burds et al., 2018; Fuerstenberg et al., 2018; Reeder et al., 2024; Sturdy Colls, 2015). As a result, there are neither standards of investigation nor systematic record keeping. Additionally, Jewish burial customs and the Jewish practice of honoring the dead generally prohibit disinterment or excavation of any bodies except in extraordinary circumstances (Kohler, n.d.), justly limiting traditional archaeological practices.

There is likewise no systematic geospatial approach to Holocaust-era research. Beorn and Knowles (2014) note the lack of, and argue the need for, specialized scholarship and datasets documenting the mass murders in Eastern Europe, and they propose a spatially based method for evaluating and documenting these atrocities. The work of Yahad-in Unum in documenting locations of mass graves using eyewitness testimonies and interviews has produced a trove of data and has enabled entirely new approaches to Holocaust research (Birch, 2021; Courcelle et al., 2023; Mykhalchuk, 2023; Shliakhtych, 2019; Zilinskyi, 2019). However, there has been no systematic approach that uses investigative GIS tools to locate mass graves in Eastern Europe.

The machine learning model proposed in this research could viably assist at every step in the effort to locate mass graves, whether researchers are searching in archives, speaking to eyewitnesses, or proposing an archaeological investigation. Yahad-in Unum investigators often discover witness to and locations of mass killings by seeking out elderly village residents (Desbois, 2015). As the final eyewitnesses to this chapter of history continue to age (Claims Conference, 2024), identifying geospatial tools, such as a prediction model, that help focus this research, are becoming increasingly urgent. Archaeologists commonly depend on archival research and personal or professional connections to locate possible grave sites (Beck et al., 2019; Brethour & Davidson, 2021; Burds et al., 2018; Reeder et al., 2024); in these cases, too, reference to a prediction model could spur memory and speed research. In all cases, a prediction model can help focus a search which otherwise stretches across almost a dozen countries in areas which often still have limited infrastructure (Desbois, 2015).

4.3 Limitations

4.3.1 Changes Through Time

There are several limitations posed by the data used in this study. Foremost is the issue of using 21st century data to evaluate events which happened more than 80 years ago. The MaxEnt algorithm has shown acceptable results for detecting ancient settlement patterns, even when modern land use practices are part of the data (Galletti et al., 2013; Rafuse, 2021), and the prediction results from Model 4 indicate that the modern datasets do contribute to the overall success of the model. However, given the time disparity, it is possible to call into question the validity of that success.

4.3.2 Genocide in Lithuania

Although the locations of various mass graves throughout Eastern Europe undoubtedly share several geospatial characteristics, that relationship may not always be clearly represented in the Lithuanian data. Lithuania was the first place in which German killing squads began organizing the mass murder of Jews by shooting *Aktionen* ("actions"). Additionally, large numbers of the civilian population in Lithuania participated in or instigated killings, even when no Germans were present (Holocaust Atlas of Lithuania, n.d.; Porat, 1994, p. 159), and local police battalions and partisans participated in *Aktionen* with the Germans throughout the country (Rhodes, 2002).

Given that so many different types of participants were involved in the killings of the Jews of Lithuania, it is very likely that many killing and burial sites were selected, not by Germans seeking the most rational and convenient burial site, but by locals, who knew the area well; or they were chosen more haphazardly, on the spur of the moment by a hastily assembled killing squad. Therefore, it is probable that there are many burial sites which, even in 1941, did not fit into the geospatial pattern the Germans were beginning to establish for their *Aktionen* and which this model is attempting to chart.

4.3.3 Modern Lithuanian Borders

This study only examined datasets situated wholly within the borders of the modern country of Lithuania, but in 1941, the Germans did not constrain themselves to borders; their only criteria for conducting shooting actions were that there were victims available in an area they controlled militarily. It is very likely that there are Jewish settlements and related Jewish mass graves which are now separated by an international border which were all part of the Third Reich at the time of the murders. The existence of these anomalies could further hinder model performance.

4.4 Future Research

Future development of a prediction model could benefit from several additional considerations. Although digital scans and paper copies of contemporary maps exist (Frackowiak, 2024; Indiana University Bloomington Libraries, 2010; LithuanianMaps.com, n.d.; Mapstor.com, n.d.; Powell, 2022), there are no digitized versions of these maps available for use in a GIS mapping program; accurate digitized data on locations of roads, forests, and villages could only improve prediction results.

Additionally, this study did not consider the chemical analysis of soil or characteristics of vegetation health and growth. Both soil and vegetation have been shown to be affected by anthropogenic disturbances and by burials of organic matter in an archeological context (Agapiou et al., 2012; Asare et al., 2020; Cholewa et al., 2022; Fiedler et al., 2009; Fuldain González & Varón Hernández, 2019; Harrison & Donnelly, 2009; Imen et al., 2024; Smejda et al., 2017); some effects have been shown to persist for centuries (Smejda et al., 2017). Likewise, thermal and infrared imaging have demonstrated their usefulness to archaeologists in revealing subsurface features and disturbances (Casana et al., 2017; Collaro et al., 2023; Verhoeven, 2012). Given the gross disturbances of nature inherent in the construction of mass graves, multifaceted analyses of soil and vegetation could positively contribute to this area of research.

And, although German killing squads operated across a vast swath of Eastern Europe and Russia, this study focused only on Lithuania. Similar studies in other countries containing mass graves should be compared with the Lithuanian data to evaluate the overall viability of the model.

Chapter 5: Conclusions

The results of this study demonstrate that MaxEnt machine modelling can provide a reliable framework for predicting the locations of unknown and undocumented Holocaust-era mass graves based on the characteristics of known mass graves. Both modern and World War II-era datasets were successfully used to create a robust and highly predictive model for Lithuania, which could help future Holocaust researchers to narrow their field of research, and which could provide data on useful foci for field surveys, thereby conserving scarce resources and taking best advantage of a rapidly ageing population of eyewitnesses.

Hundreds of known mass graves of the Holocaust are scattered across nearly a dozen Eastern European countries, including Russia. Despite the large number of known graves and killing sites, this region has not yet been subjected to a rigorous geospatial approach for locating the hundreds or thousands of mass graves which remain undocumented. Adapting this model for use across multiple countries could be used to fill a yawning gap, greatly contributing to researchers' ability to find killing sites which have been lost and forgotten for more than 80 years. A systematic, geospatially based study of Holocaust-era graves in this region could bring some measure of closure to survivors and family members while offering dignity to the dead.

Beyond the borders of Eastern Europe are mass graves of other conflicts – in Bosnia, Cambodia, Rwanda, and Sudan, to name a few – and researchers and family members of victims of these more recent massacres could potentially benefit from a wider application of this model. But even beyond its use in finding victims of conflict and massacre, this model could be applied to predict hotspots for human trafficking or undocumented migration, thereby aiding current humanitarian and law enforcement efforts. Similarly, on a much smaller scale, it could be used as an assist in the search for lost or missing persons, including historical cold cases, and lost or unmarked cemeteries, including those of victims of enslavement or persecution. Displaced indigenous people could benefit from the application of this model to rediscover lost historical, cultural, and sacred sites; it could also be applied to find locations of resistance or underground movements, such as waystations of the Underground Railroad in the United States. Especially when applied across multiple disciplines and arenas of human life, this model can be a powerful geospatial tool for finding locations humanity has lost and which should not be forgotten.

References

About us: Yahad-in Unum. (n.d.). Retrieved June 7, 2024, from https://www.yiu.ngo/en/about-us/ourmission

Agapiou, A., Hadjimitsis, D. G., & Alexakis, D. D. (2012). Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks. *Remote Sensing*, *4*(12), Article 12. https://doi.org/10.3390/rs4123892

Asare, M. O., Šmejda, L., Horák, J., Holodňák, P., Černý, M., Pavlů, V., & Hejcman, M. (2020). Human burials can affect soil elemental composition for millennia—Analysis of necrosols from the Corded Ware Culture graveyard in the Czech Republic. *Archaeological and Anthropological Sciences*, *12*(11), 255. https://doi.org/10.1007/s12520-020-01211-1

Beck, J. D., Schneider, S. G., Fuerstenberg, M. M., & Kofman, C. C. (2019). *Investigating the Trakas Holocaust Mass Execution Site: Subsurface Imaging*. http://digital.library.wisc.edu/1793/79612

Birch, A. (2021). 'They were drinking, singing, and shooting': Singing and the Holocaust in the USSR. *Journal of Perpetrator Research*, *4*(1), 171–191. https://doi.org/10.21039/85

Blanke, T., Bryant, M., Frankl, M., Kristel, C., Speck, R., Daelen, V. V., & Horik, R. V. (2017). The European Holocaust Research Infrastructure Portal. *Journal on Computing and Cultural Heritage*, *10*(1), 1:1-1:18. https://doi.org/10.1145/3004457

Blanke, T., & Kristel, C. (2013). Integrating Holocaust research. *International Journal of Humanities and Arts Computing*, *7*(1–2), 41–57. https://doi.org/10.3366/ijhac.2013.0080

Brethour, M., & Davidson, F. (2021). Visualizing intimate geographies of genocide: A spatial analysis of the Holocaust in Węgrów County, Poland (1942–1944). *Journal of Historical Geography*, *72*, 40–52. https://doi.org/10.1016/j.jhg.2021.02.003

Burds, L. T., Beck, J. D., Mataitis, R. J., Jol, H. M., Freund, R. A., McClymont, A. F., & Bauman, P. (2018). *Holocaust archaeology: Using ground penetrating radar to locate a Jewish mass grave in Kaunas, Lithuania*. 1–4. https://doi.org/10.1109/ICGPR.2018.8441590

Casana, J., Wiewel, A., Cool, A., Hill, A. C., Fisher, K. D., & Laugier, E. J. (2017). Archaeological aerial thermography in theory and practice. *Advances in Archaeological Practice*, *5*(4), 310–327. https://doi.org/10.1017/aap.2017.23

Central Intelligence Agency. (1958). *A history of the Lithuanian forests and forestry* (Declassified Intelligence Report CIA-RDP81-01043R001800230002-1). Central Intelligence Agency. https://www.cia.gov/readingroom/document/cia-rdp81-01043r001800230002-1

Central Intelligence Agency. (2024). Lithuania. In *The World Factbook*. Central Intelligence Agency. https://www.cia.gov/the-world-factbook/countries/lithuania/

Cholewa, M., Bonar, M., & Kadej, M. (2022). Can plants indicate where a corpse is buried? Effects of buried animal tissues on plant chemistry: Preliminary study. *Forensic Science International*, *333*, 111208. https://doi.org/10.1016/j.forsciint.2022.111208

Claims Conference. (2024, January 19). *Holocaust survivor demographic reports*. Claims Conference. https://www.claimscon.org/demographics/

Collaro, C., Enríquez-Muñoz, C., López, A., Enríquez, C., & Jurado, J. M. (2023). Detection of landscape features with visible and thermal imaging at the Castle of Puerta Arenas. *Archaeological and Anthropological Sciences*, *15*(152). https://doi.org/10.1007/s12520-023-01831-3

Congram, D., Kenyhercz, M., & Green, A. G. (2017). Grave mapping in support of the search for missing persons in conflict contexts. *Forensic Science International*, *278*, 260–268. https://doi.org/10.1016/j.forsciint.2017.07.021

Copernicus Land Monitoring Service. (2024). *CLC+Backbone 2021 (raster 10 m), Europe, 3-yearly*. EEA Geospatial Data Catalogue. https://sdi.eea.europa.eu/catalogue/srv/api/records/71fc9d1b-479f-4da1-aa66-662a2fff2cf7

Courcelle, T., Fijalkow, Y., & Victor, S. (2023). Introducing research through oral surveys: French students meet witnesses of the Holocaust by Bullets. *Eastern European Holocaust Studies*, *2*(1), 95–110. https://doi.org/10.1515/eehs-2023-0015

Desbois, P. (2010). *The Holocaust by bullets: A priest's journey to uncover the truth behind the murder of 1.5 million Jews*. Palgrave Macmillan.

Desbois, P. (2015). *In broad daylight: The secret procedures behind the Holocaust by Bullets*. Arcade.

European digital elevation model (EU-DEM). (n.d.). Retrieved August 19, 2024, from https://www.eea.europa.eu/en/datahub/datahubitem-view/d08852bc-7b5f-4835-a776-08362e2fbf4b

European Holocaust research infrastructure. (n.d.). [Text]. Retrieved September 14, 2024, from https://www.ehri-project.eu/

Fiedler, S., Berger, J., Stahr, K., & Graw, M. (2009). Localisation of a mass grave from the Nazi era: A case study. In K. Ritz, L. Dawson, & D. Miller (Eds.), *Criminal and Environmental Soil Forensics* (pp. 303–314). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9204-6_19

Fitzgibbon, A., Bakshi, A., & Nieto, A. (2022, August 9). *Spatial statistics: Making predictions* [Video recording]. 2022 Esri User Conference. https://mediaspace.esri.com/media/t/1_5de20htz

Frackowiak, R. (2024). *Research guides: Lithuanian collections at the Library of Congress: Geography and maps* [Research guide]. https://guides.loc.gov/lithuanian-collections-library-of-congress/geography-maps

Fuerstenberg, M. M., Jol, H. M., Freund, R. A., Jarockis, R., Kujelis, G., Reeder, P., Beck, J. D., Kofman, C. C., & Schneider, S. G. (2018, November 4). *Ground penetrating radar imaging of the Trakas Holocaust mass grave, Lithuania*. GSA Annual Meeting in Indianapolis, Indiana, USA - 2018. https://gsa.confex.com/gsa/2018AM/webprogram/Paper324710.html

Fuldain González, J. J., & Varón Hernández, F. R. (2019). NDVI identification and survey of a Roman road in the northern Spanish province of Álava. *Remote Sensing*, *11*, Article 6. https://doi.org/10.3390/rs11060725

Galletti, C. S., Ridder, E., Falconer, S. E., & Fall, P. L. (2013). Maxent modeling of ancient and modern agricultural terraces in the Troodos foothills, Cyprus. *Applied Geography*, *39*, 46–56. https://doi.org/10.1016/j.apgeog.2012.11.020

Gillespie, T. W., Smith, M. L., Barron, S., Kalra, K., & Rovzar, C. (2016). Predictive modelling for archaeological sites: Ashokan edicts from the Indian subcontinent. *Current Science (00113891)*, *110*(10), 1916–1921. https://doi.org/10.18520/cs/v110/i10/1916-1921

Haglund, W. D., Connor, M., & Scott, D. D. (2001). The archaeology of contemporary mass graves. *Historical Archaeology*, *35*(1), 57–69.

Harrison, M., & Donnelly, L. J. (2009). Locating concealed homicide victims: Developing the role of geoforensics. In K. Ritz, L. Dawson, & D. Miller (Eds.), *Criminal and environmental soil forensics* (pp. 197–219). Springer. https://www.webofscience.com/api/gateway?GWVersion=2&SrcAuth=DynamicDOIConfProc&SrcApp=WOS&KeyAID=10.1007%2F978-1-4020-9204-6_13&DestApp=DOI&SrcAppSID=USW2EC0C69TVsR7vF86IgTKRJxs0v&SrcJTitle=CRIMINAL+AND+ENVIRONMENTAL+SOIL+FORENSICS&DestDOIRegistrantName=Springer-Verlag

Holocaust atlas of Lithuania. (n.d.). Retrieved June 6, 2024, from http://holocaustatlas.lt/EN/

How presence-only prediction (MaxEnt) works. (n.d.). ArcGIS Pro. Retrieved August 18, 2024, from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-presence-only-prediction-works.htm#

Imen, G., Halima, G., Ayoub, K., & Djamel, A. (2024). Utilizing the MaxEnt machine learning model to forecast urban heritage sites in the desert regions of southwestern Algeria: A case study in the Saoura region. *Archaeological Prospection*, *31*(1), 23–35. https://doi.org/10.1002/arp.1923

Indiana University Bloomington Libraries. (2010, November 20). *Russian military topographic maps*. https://libraries.indiana.edu/cyrillic-index-cyrillic-topographic-maps

INSPIRE geoportal. (n.d.). Retrieved August 19, 2024, from https://inspire-geoportal.ec.europa.eu/srv/eng/catalog.search#/datasetdetails?country=lt&view=thematicOverview&theme=none&resourceId=27b07913-8bb2-47e7-b11c-2c8a8ed5b4f5

Iwaskiw, W. R. & Library of Congress (Eds.). (1996). *Estonia, Latvia & Lithuania: Country studies* (1st ed). Federal Research Division, Library of Congress: For sale by the Supt. of Docs., U.S. G.P.O. https://www.loc.gov/resource/frdcstdy.estonialatvialit00iwas_0

Jakulyte-Vasil, M. (2015). Holocaust atlas of Lithuania. In T. Lutz, D. Silberklang, P. Trojański, J. Wetzel, M. Bistrović, A. Angrick, & International Holocaust Remembrance Alliance (Eds.), *Killing sites: Research and remembrance*. Metropol Verlag.

Jakulytė-Vasil, M. (2024, June 22). *RE: Seeking access to Lithuanian Jewish Cemetery Data for Master's Thesis* [Personal communication].

JewishGen. (n.d.-a). *Communities tree*. JewishGen. Retrieved July 1, 2024, from https://www.jewishgen.org/Communities/Trees/

JewishGen. (n.d.-b). *JewishGen: The home of Jewish genealogy*. Retrieved June 19, 2024, from https://www.jewishgen.org/

Katz, D. (2023). *The language of Litvak gravestones: A cultural dictionary*. Defending History. https://defendinghistory.com/the-language-of-litvak-gravestones-a-dictionary-by-dovid-katz

Kohler, K. (n.d.). *Disinterment*. Jewish Encyclopedia. Retrieved September 14, 2024, from https://jewishencyclopedia.com/articles/5222-disinterment

Kwiet, K. (1993). From the diary of a killing unit. In J. Milfull (Ed.), *Why Germany? : National Socialist anti-Semitism and the European context* (pp. 75–90). Providence : Berg. https://archive.org/details/whygermanynation0000unse/mode/2up

Lietuvos Respublikos upių, ežerų ir tvenkinių kadastras (UETK). (2023). LEIP metaduomenų aplankas. https://www.geoportal.lt/metadata-catalog/catalog/search/resource/details.page?uuid=%7B43007570-30B8-43AA-818F-B18F125E8B5F%7D

LithuanianMaps.com. (n.d.). *Maps1940-45*. Retrieved July 2, 2024, from http://www.lithuanianmaps.com/Maps1940-45.html

LitvakSIG. (2022, August). *Shtetl map*. LitvakSIG Lithuanian-Jewish Special Interest Group. https://www.litvaksig.org/research/map

Liu, J. (n.d.). Presence-only prediction (MaxEnt) 101: Using GIS to model species distribution. *ArcGIS Blog*. Retrieved August 20, 2024, from https://www.esri.com/arcgis-blog/products/arcgis-pro/analytics/presence-only-prediction-maxent-101-using-gis-to-model-species-distribution/

Maceva: Litvak cemetery catalogue. (n.d.). Maceva. Retrieved June 30, 2024, from http://www.litvak-cemetery.info/

Malinauskaite, G. (2013, November 8). *Holocaust memory and Antisemitism in Lithuania: Reversed memories of the Second World War*. Antisemitism in Europe Today: The Phenomena, the Conflicts. https://www.jmberlin.de/sites/default/files/media/documents/antisemitism-in-europe-today_15-malinauskaite.pdf

Mapstor.com. (n.d.). *About*. Retrieved June 29, 2024, from https://mapstor.com/product.html

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Miškų kadastro duomenys. (2024). LEIP metaduomenų aplankas. https://www.geoportal.lt/metadata-catalog/catalog/search/resource/details.page?uuid=%7BBF626474-9B36-BFE6-2E43-EAEC396FCFE9%7D

Molina, C. M., Wisniewski, K. D., Drake, J., Baena, A., Guatame, A., & Pringle, J. K. (2020). Testing application of geographical information systems, forensic geomorphology and electrical resistivity tomography to investigate clandestine grave sites in Colombia, South America. *Journal of Forensic Sciences*, *65*(1), 266–273. https://doi.org/10.1111/1556-4029.14168

Mykhalchuk, R. Y. (2023). Holocaust oral history sources in the Yahad-In Unum archive collection (Based on Volhynia-Podolia General District materials). *Colloquia Humanistica*, *12*. https://doi.org/10.11649/ch.2908

Nadler, A. (2010, February 12). *Litvak*. The YIVO Encyclopedia of Jews in Eastern Europe. https://yivoencyclopedia.org/article.aspx/Litvak

Open data—Geoportal.lt. (n.d.). Retrieved August 19, 2024, from https://www.geoportal.lt/geoportal/en/web/en/open-data#kiti_ad

OpenDEM. (n.d.). Retrieved August 19, 2024, from https://www.opendem.info/

OpenStreetMap. (n.d.). OpenStreetMap. Retrieved August 10, 2024, from https://www.openstreetmap.org/

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Pohl, D. (2015). Historiography and Nazi killing sites. In *Killing sites: Research and remembrance* (Vol. 1, pp. 31–46). Metropol Verlag.

Porat, D. (1994). *The Final Solution: Origins and implementation* (D. Cesarani, Ed.). Routledge. https://ebookcentral.proquest.com/lib/pensu/detail.action?docID=168847

Powell, S. (2022, December 14). *Library guides: German World War II captured maps: The maps*. https://guides.lib.berkeley.edu/c.php?g=1055588&p=7668284

Rafuse, D. J. (2021). A Maxent predictive model for hunter-gatherer sites in the Southern Pampas, Argentina. *Open Quaternary*, *7*(1), 1–21. https://doi.org/10.5334/oq.97

Reeder, P., Jol, H., McClymont, A., Bauman, P., & Barrow, M. (2024). Using geophysics to locate Holocaust era mass graves in Jewish cemeteries: Examples from Latvia and Lithuania. *Heritage*, *7*(7), 3766–3798.

Shliakhtych, R. (2019). Holocaust in countryside of Dnipropetrovska Oblast (by testimony in the Yahad-In Unum archive). *Roxolania Historĭca*, *2*(17), 188–198. http://dx.doi.org/10.15421/30190212

Silberklang, D. (2015). Introduction to the conference and IHRA perspective. In *Killing sites: Research and remembrance* (Vol. 1). Metropol Verlag.

Silva, M., & Burns, A. (2015). Preface. In *Killing sites: Research and remembrance* (Vol. 1, p. 13). Metropol Verlag.

Smejda, L., Horak, J., Alexandrova, J., & Hejcman, M. (2017). *Ashes to ashes, dust to dust: Geochemical signals of mortuary areas*. 5068. EGU General Assembly Conference Abstracts.

Sturdy Colls, C. (2015). *Holocaust archaeologies: Approaches and future directions*. Springer International Publishing. https://doi.org/10.1007/978-3-319-10641-0

United States Holocaust Memorial Museum. (n.d.). *Einsatzgruppen (mobile killing units)*. Holocaust Encyclopedia. Retrieved June 6, 2024, from https://encyclopedia.ushmm.org/content/en/article/einsatzgruppen

United States Holocaust Memorial Museum. (2021, August 31). *Mass shootings of Jews during the Holocaust*. Holocaust Encyclopedia.

https://encyclopedia.ushmm.org/content/en/article/mass-shootings-of-jews-during-the-holocaust

United States Holocaust Memorial Museum. (2023, September 26). *How many people did the Nazis murder?* Holocaust Encyclopedia. https://encyclopedia.ushmm.org/content/en/article/documenting-numbers-of-victims-of-the-holocaust-and-nazi-persecution

United States Holocaust Memorial Museum. (2024, April 9). *Lithuania*. Holocaust Encyclopedia. https://encyclopedia.ushmm.org/content/en/article/lithuania

Verhoeven, G. J. (2012). Near-infrared aerial crop mark archaeology: From its historical use to current digital implementations. *Journal of Archaeological Method and Theory*, *19*(1), 132–160. https://doi.org/10.1007/s10816-011-9104-5

Wachtel, I., Zidon, R., & Shelach-Lavi, G. (2020). Using the maximal entropy modeling approach to analyze the evolution of sedentary agricultural societies in northeast China. *Entropy*, *22*(3), Article 3. https://doi.org/10.3390/e22030307

Walke, A. (2022). Testimony in place: Witnessing the Holocaust in Belarus. *East European Jewish Affairs*, *52*(1), 80–104. https://doi.org/10.1080/13501674.2022.2131284

Yad Vashem. (n.d.). *Murder of the Jews of the Baltic States*. Retrieved June 6, 2024, from https://www.yadvashem.org/holocaust/about/final-solution-beginning/baltic-states.html

Yahad—In Unum map. (n.d.). Retrieved June 7, 2024, from https://www.yahadmap.org/#map/

Zilinskyi, V. (2019). They were killed on the spot: The destruction of Jews during the Nazi occupation in the territory of the Lviv region. *Bulletin of Lviv University*, *Special Issue*, 340–360. https://doi.org/10.30970/his.2019.0.9855

Appendix

Regression Coefficients of All Variables Contributing to the Model

Basis Function	Variable(s)	Hinge Range OR Category Value (if applicable)	Coefficient
Linear	Aspect		0.0054
Linear	Elevation		-0.0058
Linear	Slope		0.0898
Linear	Jewish Cemeteries		-0.0001
Linear	Lakes		-0.0000
Linear	Rivers		-0.0001
Linear	Wetlands		-0.0001
Linear	Towns & Villages		-0.0006
Linear	Jewish Communities		-0.0001
Quadratic	Aspect, Aspect		-0.0000
Quadratic	Elevation, Elevation		-0.0000
Quadratic	Wetlands, Wetlands		-0.0000
Product	Aspect, Elevation		-0.0000
Product	Aspect, Jewish Cemeteries		0.0000
Product	Aspect, Lakes		0.0000
Product	Aspect, Rivers		0.0000
Product	Aspect, Wetlands		-0.0000

Product	Aspect, Forest: Plots		-0.0000
Product	Aspect, Towns & Villages		-0.0000
Product	Aspect, Jewish Communities		0.0000
Product	Elevation, Slope		-0.0002
Product	Elevation, Jewish Cemeteries		0.0000
Product	Elevation, Lakes		0.0000
Product	Elevation, Rivers		-0.0000
Product	Elevation, Wetlands		0.0000
Product	Elevation, Towns & Villages		0.0000
Product	Slope, Jewish Communities		0.0000
Product	Slope, Roads		0.0000
Product	Jewish Cemeteries, Lakes		0.0000
Product	Jewish Cemeteries, Wetlands		0.0000
Product	Jewish Cemeteries, Towns & Villages		-0.0000
Product	Jewish Cemeteries, Jewish Communities		-0.0000
Product	Jewish Cemeteries, Roads		-0.0000
Product	Lakes, Jewish Communities		0.0000
Product	Rivers, Wetlands		0.0000
Product	Rivers, Towns & Villages		0.0000
Product	Rivers, Roads		-0.0000
Product	Wetlands, Towns & Villages		0.0000
Product	Jewish Communities, Roads		-0.0000

Smoothed Step (Hinge)	Aspect	319.68 – 359.76 degrees	-0.3131
Smoothed Step (Hinge)	Elevation	-6.30 – 25.83 meters	1.0637
Smoothed Step (Hinge)	Slope	0.00 – 2.65 %	0.1432
Smoothed Step (Hinge)	Slope	0.00 – 5.29 %	0.0709
Smoothed Step (Hinge)	Jewish Cemeteries	0.00 – 3,040.35 meters	-1.9778
Smoothed Step (Hinge)	Forest: Plots	0.00 – 256.03 meters	-3.6027
Smoothed Step (Hinge)	Towns & Villages	0.00 – 1,199.77 meters	0.3645
Smoothed Step (Hinge)	Jewish Communities	0.00 – 2,462.72 meters	-0.9108
Smoothed Step (Hinge)	Roads	0.00 – 310.65 meters	-0.7071
Categorical	Forest: Dominant Leaf Type	Non-forested	-0.2065
Categorical	Forest: Dominant Leaf Type	Coniferous	0.0535
Categorical	Land Use	Sealed	0.8221
Categorical	Land Use	Woody: Needle-leaved Trees	1.5923
Categorical	Land Use	Woody: Broadleaf Deciduous Trees	0.9309
Categorical	Land Use	Permanent Herbaceous	0.2111
Categorical	Land Use	Water	0.3330