*Article*

# Artificial Interpretation: An Investigation into the Feasibility of Archaeologically Focused Seismic Interpretation via Machine Learning

Andrew Iain Fraser [1], Jürgen Landauer [2], Vincent Gaffney [1,*] and Elizabeth Zieschang [1]

1 Submerged Landscape Research Centre, School of Archaeological and Forensic Sciences, University of Bradford, Richmond Road, Bradford BD7 1DP, UK; a.i.fraser1@bradford.ac.uk (A.I.F.); ejzieschang@gmail.com (E.Z.)
2 Landauer Research, 71642 Ludwigsburg, Germany; juergenlandauer@gmx.de
* Correspondence: v.gaffney@bradford.ac.uk

**Abstract:** The value of artificial intelligence and machine learning applications for use in heritage research is increasingly appreciated. In specific areas, notably remote sensing, datasets have increased in extent and resolution to the point that manual interpretation is problematic and the availability of skilled interpreters to undertake such work is limited. Interpretation of the geophysical datasets associated with prehistoric submerged landscapes is particularly challenging. Following the Last Glacial Maximum, sea levels rose by 120 m globally, and vast, habitable landscapes were lost to the sea. These landscapes were inaccessible until extensive remote sensing datasets were provided by the offshore energy sector. In this paper, we provide the results of a research programme centred on AI applications using data from the southern North Sea. Here, an area of c. 188,000 km$^2$ of habitable terrestrial land was inundated between c. 20,000 BP and 7000 BP, along with the cultural heritage it contained. As part of this project, machine learning tools were applied to detect and interpret features with potential archaeological significance from shallow seismic data. The output provides a proof-of-concept model demonstrating verifiable results and the potential for a further, more complex, leveraging of AI interpretation for the study of submarine palaeolandscapes.

**Keywords:** AI; semantic segmentation; deep learning; CNN; archaeology; submerged landscape; marine geophysics

## 1. Introduction

Artificial intelligence (AI) has demonstrated efficient, consistent, and high-quality solutions in many industries [1] where datasets are too large to be processed manually due to high resolution data acquisition, the vast geographic scales involved, or where resources are limited and investigation is costly and\or time dependent [2]. The investigation of submerged landscapes suffers many of these problems. Areas of potential interest are frequently inaccessible and often beyond the scope of diver investigation. Locations of interest may only measure metres across yet are obscured by modern sediments that may be tens of metres thick and many kilometres in extent. Traditional tools used to locate and explore points of interest (coring, dredging, and survey) require expensive boat time, and successful results are contingent on external factors including weather. The context of archaeological exploration in marine environments is also critical. Academic researchers, undertaking long-term research in marine prospection, have benefitted massively through access to data generated by the offshore energy sector. When analysis is required in mitigation of development by the commercial sector, archaeologists may be involved in rapid and detailed studies of extensive marine surveys. As survey extent and data resolution increases, the challenges of undertaking all these activities have mounted exponentially.

This is significant when it is appreciated that marine palaeolandscapes are currently among the least understood of cultural landscapes, despite their extent and archaeological significance. Sea levels have risen globally by up to 120 m since the Last Glacial Maximum (LGM), resulting in the loss of c. 20 million km$^2$ of territory worldwide and potentially 3 million km$^2$ of habitable land around Europe [3]. Until recently, interpretation of the long-term human occupation of these areas was largely reliant on comparative data provided from terrestrial sites, coastal or nearshore evidence. It would be comforting to believe that these areas were of little significance, but where archaeological sites have been identified in coastal or nearshore regions, such as those off the Danish and Swedish Baltic coasts, we glimpse a rich heritage of continual occupation spanning thousands of years [4].

In deeper waters, cultural data are almost entirely lacking, despite the southern North Sea being one of the most researched and mapped areas of submerged, prehistoric landscapes in the world. Essentially, the evidence base comprises chance finds dredged from the seabed or deposited on beaches via sand extraction vessels [5,6]. Even where cultural material is directly recovered, the value of these finds is frequently compromised through a lack of archaeological context. Currently, neither extant occupation sites nor substantive in situ finds of cultural material have yet been recovered at depth and beyond 12 nautical miles of a modern coastline across northwestern Europe [7].

One area in which archaeology has made significant progress is through the application of remote sensing technologies to provide topographic mapping of inundated landscapes. These maps can now be used to direct fieldwork to recover sediment samples that may contain environmental proxies, supporting our understanding of climate and landscape change [8]. While such sampling is extremely unlikely to recover in situ archaeological material, it may provide proxy environmental evidence, including sedimentary aDNA, and datable material that can be used to build a better understanding of the evolution of the landscapes that were home to prehistoric communities [9]. The geophysical surveys underpinning such work may cover areas of hundreds, if not thousands, of kilometres, and the available datasets now exceed the interpretative capacity of available geoarchaeological analysts.

Given the emerging challenges of undertaking research in such a data-rich environment, the University of Bradford funded a 12-month research project to investigate how AI could be employed to detect and interpret features of archaeological significance from shallow seismic data. The objective of the fellowship was to design and implement a machine learning solution that could identify sub-seafloor features and the sediments within these features.

*Research Aims*

This research was undertaken alongside the University of Bradford's *Taken at the Flood* AHRC funded project, utilising data provided by windfarm developers to generate new processes and workflows that might identify specific areas of human activity beneath the sea. This project targets locations that fall within a 'Goldilocks Zone', i.e., an area likely to be attractive to Mesolithic hunter gatherers (c. 10,000–5000 BC), with potential for preservation of archaeological material and at a depth accessible to current investigative methods such as coring and dredging [10]. This is not to suggest that archaeological evidence does not exist at greater depths or is not associated with earlier periods. However, archaeological researchers rarely have access to rigs providing deeper cores and, for the purposes of study of the later prehistoric communities in the North Sea, smaller cores, dredges, and grabs are probably adequate to recover archaeological evidence if such localities can be identified.

The machine learning solution implemented in the research presented here was therefore required to adhere to the Goldilocks model, and the outputs of this study recorded the machine learning methods considered initially as well as the rationale supporting the final choice of a convolutional neural network (CNN). The CNN was required to provide verifiable, reproduceable, and efficient results in identifying archaeologically significant deposits within Holocene sediments and at a depth that was still accessible to prospection

below the current seabed. The final output, while not a total solution, was intended to act as a proof-of-concept model for future development and to support further AI applications within submarine palaeolandscape research.

The project utilised survey data from the Brown Bank area of the southern North Sea (Figure 1). The Brown Bank study area was selected for a number of reasons. The area has a history of chance recovery of archaeological material [6,11,12]. Although the source of such material is not precisely known (Figure 1), the presence of such material suggests that this area possesses a high archaeological potential. The study area also benefits from recent survey using the latest marine geophysical technologies in both extensive, regional survey lines and tight, local grid survey [7]. This work provides us with a detailed understanding of the landscape and identifies locations where we might expect evidence for human activity, for instance areas adjacent to rivers, lakes, and estuaries. The Brown Bank region also contains many known peat beds that are of Holocene date [7]. Peat exhibits excellent preservation potential while also possessing a distinctive amplitude signal in shallow seismic surveys that can be identified easily via manual interpretation techniques [8,13,14]. These Holocene peats occur at depths that are often close to, or even exposed at the seabed [15]. Whether these peat layers are the source of archaeological material in the area remains to be proven but these conditions make them accessible in a number of areas within the region.
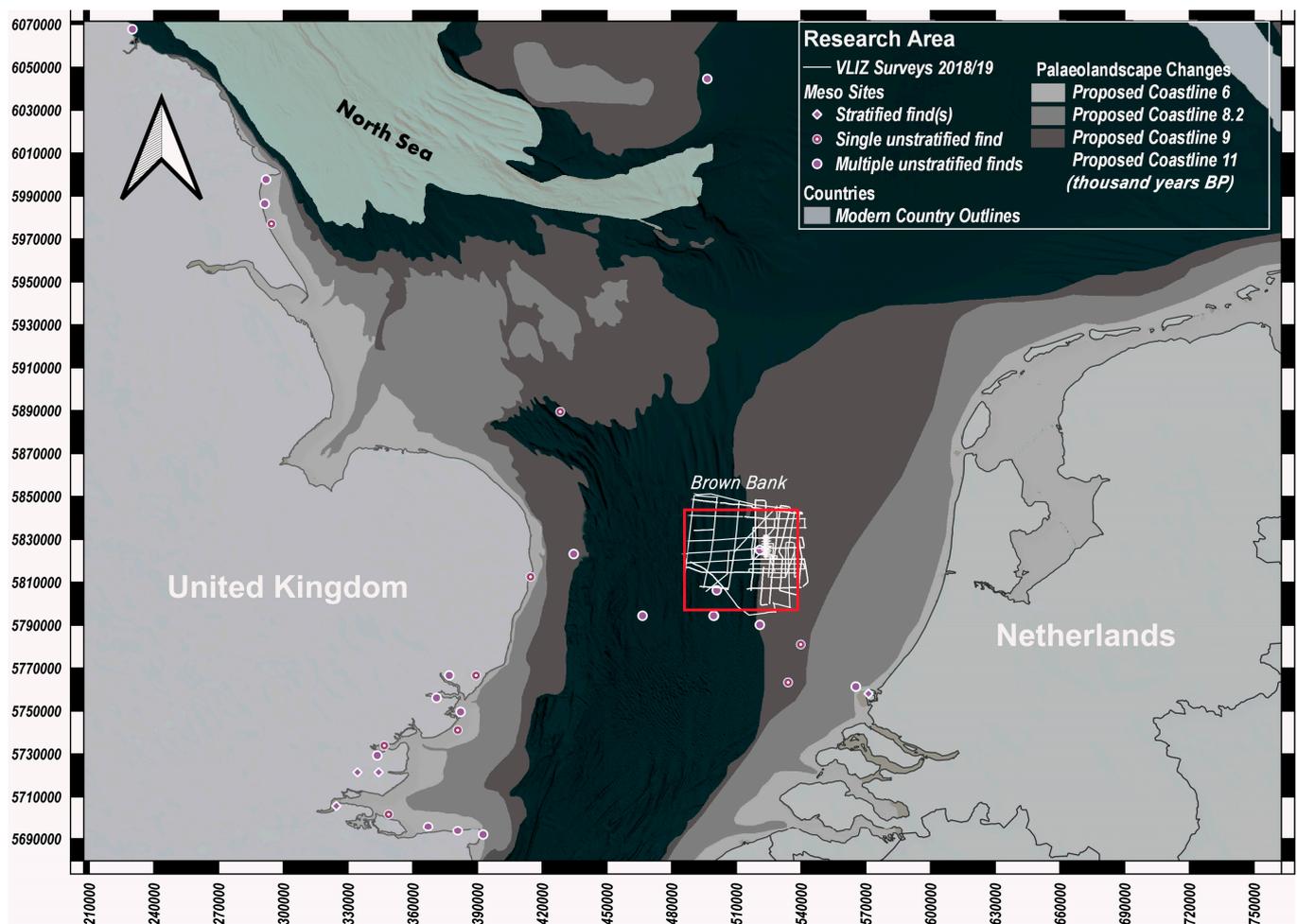


**Figure 1.** Extent of survey data acquired around the Brown Bank area of the southern North Sea. Approximate extents of prehistoric coastlines are shown along with locations of archaeological material recovered from the area (data source: SplashCOS [16]).

The aims of the research can be summarized as follows:

- The creation a proof-of-concept AI tool to aid in the identification of potential archaeologically significant targets by means of shallow marine geophysical interpretation.
- The model must be guided by the principal of the 'Goldilocks Zone', as described above. It attempts to identify terrestrial surface features containing specific sediment depositions that are accessible to sampling using current archaeological methods.
- Targets within the initial AI model are restricted to the Brown Bank study area and aim at identifying organic sediments that display distinctive negative amplitude signals.
- The proof-of-concept model is initially confined to the analysis of potential Holocene land surfaces, especially where features such as rivers, lakes, and estuaries are evident in the data.
- The AI Tool is flexible enough to allow for further layers of complexity to be added at a later date.

Figure 1 shows the predicted location of coastlines throughout the early Holocene and the survey lines used in the study. These surveys were carried out in 2018 and 2019 by the Flanders Marine Institute (VLIZ), partnered by Europe's Lost Frontiers, an ERC-funded project at the University of Bradford. The surveys employed a multi-transducer parametric echosounder arranged in a single beam array to enhance output. Using a frequency of 8–10 kHz, this configuration provided decimetre-level resolution up to 15 m below the seafloor and excellent images of shallow features [7]. The interpretation software used was IHS Kingdom Suite (2020). Other applications and tools were employed to test methods of data input into the machine learning model, and these are described below. Training and testing of the datasets were undertaken at Bradford, and AI programming was executed by Jürgen Landauer of Landauer Research.

## 2. Materials and Methods

Before any AI workflow could be built, it was necessary to identify appropriate parameters for detailed study. For this purpose, precise definitions of key terms were also required. In this case, the definition of 'features of archaeological significance', was problematic. While the southern North Sea is one of the best understood submerged palaeolandscapes in the world [2,6–8], the lack of formal sites, in comparison with the terrestrial record, is stark. This is not entirely surprising given the nature of hunter-gatherer archaeology. Archaeological significance is frequently difficult to ascribe even when considering terrestrial archaeological evidence. Even in terrestrial archaeology, what constitutes a 'site' is often contentious when applied to prehistoric hunter-gatherer activities. There is little or no guidance to assist the characterisation of such locations in submarine environments. For this reason, this research was oriented less towards the direct identification of human activity and instead emphasised the location of organic deposits suitable for environmental archaeological analysis. While the presence of organic deposits does not directly indicate archaeological activity, such sediments do indicate an increased chance of preservation of organic archaeological materials, including bone tools, as well as palaeoenvironmental materials, such as charcoal, pollen, and wood, and these can provide important information related to the environment, dating, and cultural affinity. The extensive layers of peat throughout the study area, and the distinctive high-negative-amplitude signal they generate in geophysical survey, therefore provide an ideal target for a proof-of-concept AI tool.

The goal of the research, in line with the 'Goldilocks Zone' criteria, was therefore to create an AI tool that could classify organic sediment, namely peat, at a depth that would be accessible to vibrocore prospection (i.e., a maximum of 6 m below the seabed) and would provide sufficient data for the testing of such a tool.

The geophysical data acquired were in the form of 2D seismic profiles. During survey, sound waves are transmitted into the seabed from a ship-based source. The sound wave passes through layers of sub-bottom sediment and any reflected waves are picked up by receivers. The waveforms recorded vary in frequency and amplitude intensity at each material change in the sediment. Therefore, multiple parallel waveforms measured together

on a time axis give a visual representation of sedimentary units and their interfaces. A single survey line may be made up of many thousands of these parallel waves, or shotpoints, and the 2D profile of the survey line are interpreted manually and the high amplitude responses are picked and interpolated to provide continuous surfaces between sedimentary units.

*2.1. Data Input Methods*

Three different methods were considered to provide the most useful and efficient means of inputting data into the machine or deep learning model. Each method utilised a different pre-processing path based on the raw seismic data and the manual interpretation required to train it. These methods were Vector Array, Sound Wave Prominence, and Image Classification.

2.1.1. Vector Array

Manual interpretation of seismic features is usually performed using hand-drawn lines, either on paper or digital profiles. These lines demonstrate linear movement on both X and Y axes and can be displayed as a grid array to provide vector information in much the same way that online handwriting recognition AI reads pen strokes [17]. To implement this method, manual seismic interpretations were exported from IHS Kingdom Suite as flat text files with X and Y co-ordinates. The proposed model utilised edge detection and slope analysis to classify the line paths, employing a machine learning workflow such as K nearest neighbour (K-NN) pre-processing. Such approaches have previously demonstrated excellent recognition results for line interpretation [18,19].

This method was not pursued because the design relied on hand-drawn interpretations as data input and these shapes alone held no geolocation markers or other reference to the position along the seismic survey line. Training data would therefore remain independent from the native data. This method would have restricted the AI to only working with previously interpreted data and would severely limit any future functionality to work with direct, raw seismic data. Additional tools would also be required to translate the AI output back onto the original seismic profile to provide depth, amplitude, and geo-location information.

2.1.2. Sound Wave Prominence

Seismic data comprise acoustic signals reflected back to a receiver or receivers from material changes in the sub-seafloor geomorphology. These reflections are visualised as amplitude lines, possessing peaks and troughs defined by periodicity [20]. It is possible to extract individual waveforms by means of trace analysis in the IHS Kingdom Suite and, by using a precise depth model, to align multiple traces by amplitude prominence at coherent depths.

Flat text files containing amplitude strength at precise depths were extracted from the source seismic data using IHS Kingdom Suite. A number of different classification algorithms were tested in Matlab, using Seis Pick and Waveform Suite to test the coherence of prominent amplitudes at precise time signatures. The benefit of this method as a potential pre-processing of data for input into an AI routine was that it directly processed the source seismic profiles, including location, depth and amplitude, so classification could be made directly by using amplitude and depth from survey data.

The creation of a sizeable training dataset required for this method was extremely intensive and proved problematic with the resources available to the project. Each survey might contain thousands of traces, each requiring manual extraction and processing before creating individual flat text files. A second problem was that although the method was able to use the seismic source data, it could not process manual interpretations and would have required additional AI pre-processing steps in the workflow. A final issue was that the flat file format was not readable by a human interpreter, thus poor input data could not be identified or investigated. For these reasons, this method was not pursued.

2.1.3. Image Classification

An AI solution to identifying individual parts of an image on a pixel-by-pixel basis was required. The tool needed the ability to classify parts of the sub-seafloor and identify potential horizons and features as well as any deposits that lay within the features. Images were processed in greyscale with alternating white and black amplitude lines (negative and positive). This meant that the final CNN could use a simplified greyscale input classifier rather than a colour palette.

The issues faced with image classification were twofold. First, the CNN required a large training dataset. To provide this, a simplified interpretation technique was applied. This limited manual analysis to the identification of a specific palaeoland horizon, which was picked in a single colour (green). Landscape features associated with this horizon, for example river channels, were colour-filled at a transparency of 50% in order not to obscure the underlying source data. Then, a single dashed line was used to mark any pertinent deposition within the feature (Figure 2). The horizon and deposition lines were coloured in order to contrast with the greyscale seismic image and the AI interpretation likewise employed a contrasting colou:
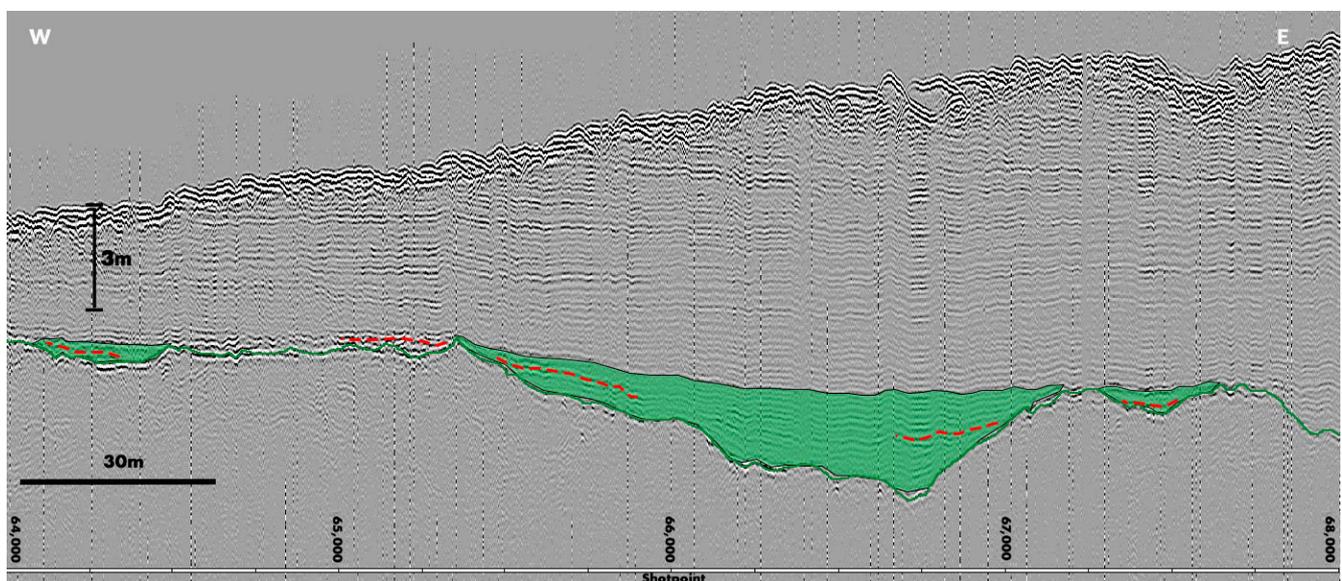


**Figure 2.** Training data comprised of simplified interpretations over a seismic profile, here showing a marked horizon and landscape feature (green). They also show significant sediment deposition (red dashed line). Each used image comprised 4000 shotpoint traces at a resolution of $1600 \times 600$ px.

Simplification of the manual interpretation enabled efficient creation of a training image catalogue by removing all the horizons not deemed pertinent to the study. Modern deposition and deeper geological formations were therefore ignored. The focus of interpretation was on verifiable horizons and landscape features associated with a terrestrial surface. In accordance with the 'Goldilocks Zone,' as defined by the *Taken at the Flood* project, these horizons needed to be located at depths accessible to archaeological investigation, and the features needed to contain high amplitude waveforms that might indicate peats or other organic sediments [21]. A single dashed line was used to identify such sediments and was considered an important development for future iterations of the AI involving more complex analyses. For an initial proof-of-concept model, it was decided to restrict the focus to the identification of features.

The second issue was one of image resolution versus computer memory restrictions. By default, most image classification models use small image sizes and resolutions to support efficient processing. To manually interpret a seismic profile, large, high-resolution images are required. To address this, it was decided to employ semantic segmentation.

Semantic segmentation refers to the process of partitioning a digital image into several parts where pixels can be said to be of the same class and share logical relationships, known as semantics. Today's AI technologies, especially CNNs, are considered to be the state-of-the-art approach for such processes [22]. AI-based image segmentation has been used extensively in many fields, including archaeology [23]. Energy companies already employ deep learning CNNs to interpret seismic and well data (see [24]). Unfortunately, access to these proprietary models is restricted and the horizons they interpret tend to cover many kilometres in area and at great depth, though with less resolution. In submerged landscape archaeology, the focus is on seismic data with very high (decimetre scale) resolution and at very shallow (60 m maximum) depths. Therefore, a new solution was required.

### 2.2. Materials: Dataset Creation

A process was established for the creation of verifiable, repeatable datasets for use in training the CNN. The survey lines were viewed in IHS Kingdom Suite and set to fixed axes of 120 shotpoint traces/cm on a horizontal axis and 1000 cm/s on a vertical axis. This allowed for a uniform data capture across all images without a loss of data resolution, as per Historic England's guidance for such surveys [25]. For improved visualisation, each trace was normalised to a mean amplitude to give a uniform greyscale representation across multiple traces. A standard low bandpass filter was employed to eradicate background noise. This was set at thresholds of 5% and 95% of the total amplitude range. The results were exported as jpeg images at $1600 \times 600$ pixels and upscaled to 96 dpi. This set a uniform window of 4000 shotpoint traces for each image. To support an improved segmentation of the images, it was also decided to remove all superimposed user interface artefacts at this stage, including automated annotation and scale bars.

Once this method had been established, processing was implemented following interpretation training, and the results verified by an experienced seismic interpreter.

Exported images were interpreted in GIMP (GNU Image Manipulation Program) version 2.10 and Adobe Photoshop 2018 and 2020. The simplified interpretation set out above was applied and then saved as a jpeg file at the highest possible setting. Each file was labelled with the seismic survey name, shotpoint range, and appended with a label of 'v1' so that each section of the seismic data was saved as both source (uninterpreted) data and interpreted data. In this way, a training dataset of 551 interpreted features was created, with an additional 2100 images without manual interpretation. These included examples containing palaeo features and examples without any notable features or sediment deposition. Images that were deemed to be of poor quality or which contained problematic data—such as potential gas blanking—were removed, as these were considered detrimental to the training of the CNN. It was noted, however, that the identification of gas pockets would be an important addition to later iterations of the AI tool.

Each of the 551 images were split into smaller squares, or patches, by means of semantic segmentation to ensure each patch still held sufficient information to interpret continuous horizons, topographic features, or sediment deposition within the seismic data. It was found that a patch size of $256 \times 256$ pixels yielded better AI training results than smaller patches. The dataset was further augmented by splitting the large images into a second grid with an offset of 50%. This resulted in additional patches with different information but similar semantic relationships to the initial patches. This was used to ensure better standardization during training.

### 2.3. Method—Deep Learning Model Training

The resulting dataset was then used to train a CNN for semantic segmentation. We used a state-of-the-art DeepLabV3+ architecture [26] with a ResNet-34 backbone based on the Fast.ai library with PyTorch. The training dataset was split into training/validation subsets with an 80/20 ratio. A variety of data augmentation techniques were applied, ranging from randomly flipping training images horizontally to slight changes in their scale. Of particular benefit was the random erasing data technique [27], where a few

randomly chosen patches of an image are replaced with random noise, thus contributing to better regularization. Transfer learning, a technique where the neural network is pre-trained on another task (ImageNet in our case), was also employed. It was observed that this reduced training time greatly, despite our images being greyscale unlike the full colour of the ImageNet data. Empirically, we found that focal loss [28] boosted prediction results as a loss function for the CNN training (Figure 3). This setup allowed for training of the model for 35 epochs (or iterations) without overfitting.
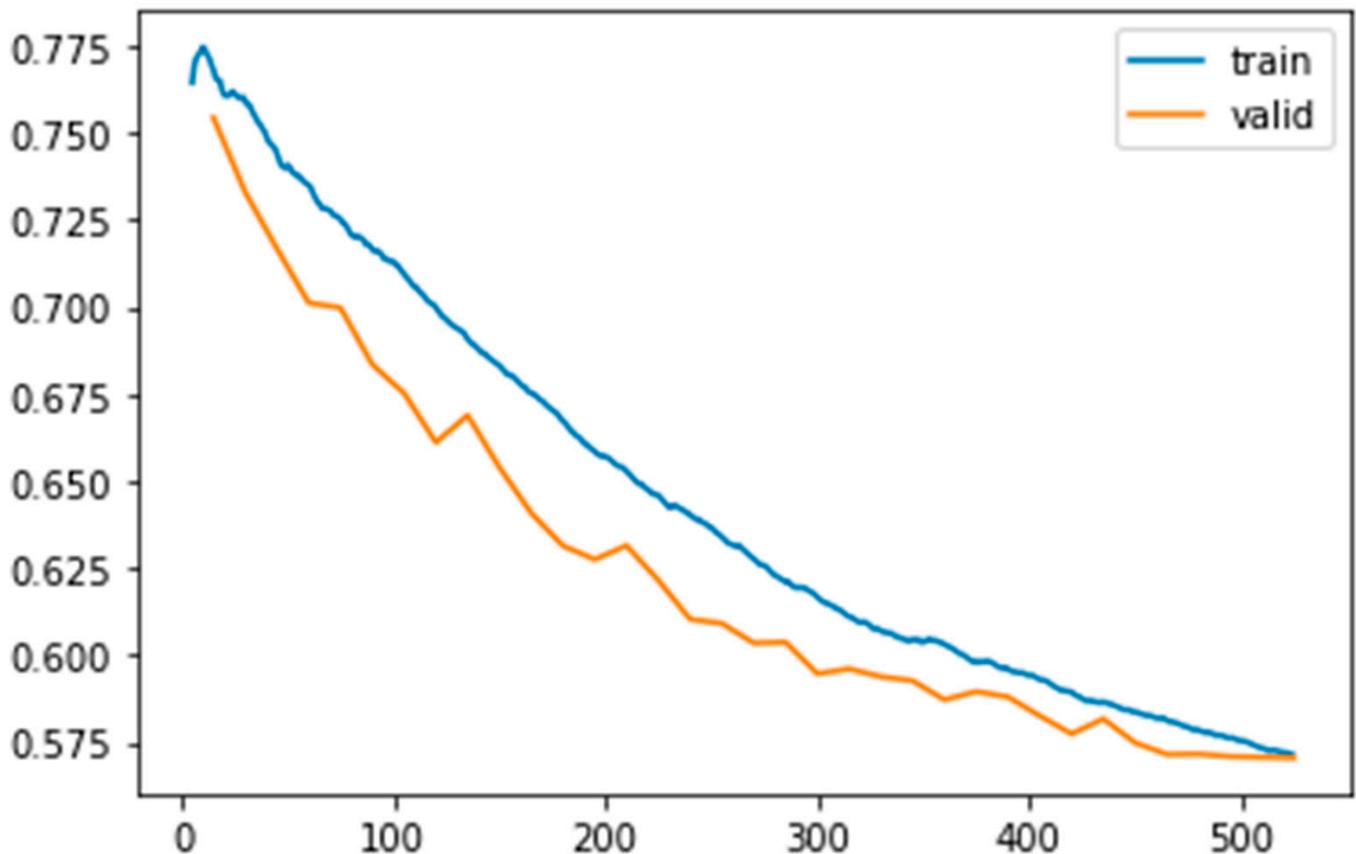


**Figure 3.** Demonstration of the improved loss through 35 epochs of the deep learning model. Each CNN was trained over up to 35 epochs (*x* axis shows batches, *y* axis is loss, data in blue are training loss, orange are validation loss).

The key benchmark tracked during training was IoU (Intersection-over-Union), a metric often used as a default for semantic segmentation tasks. With the validation dataset, the model reached an IoU of 0.89, an acceptable value considering that the results on imagery outside the training data without a validation dataset were significantly lower.

Over a period of nine months, a total of eleven versions of the CNN were created and trained, each employing additional refinements. All of the code, training and testing catalogues, and reiterative results were held in a shared repository on Google Colab.

This training process resulted in a semantic segmentation model that used image patches of $256 \times 256$ pixels as input and was able to compute a corresponding segmentation mask as an output. For each pixel in each patch, the model reported a confidence score (between 0% and 100%) for classifications of interpreted topographic features containing sediment deposition at an appropriate depth. Typically, scores around 50% or better were classed as being part of the 'feature class'.

Once this had been achieved, the patches needed to be reassembled back to the original image size. To accomplish this, a mosaicking technique was used, which resulted in a continuous set of $256 \times 256$-pixel patches. The predictions for each patch were once

again merged into one image. Figure 4A–C provides an example of the output: Figure 4A shows the full-size input image in greyscale. This was split into patches for training the CNN. These AI predictions were reassembled using the mosaicking technique (Figure 4B) and visualized as greyscale images where darker pixels represented regions of lower or zero confidence for the 'feature class', whereas brighter pixels represent predicted high confidence in the classification. To ease human interpretation, this greyscale image was then filtered using the 40% threshold mentioned below and overlaid onto the original input image (Figure 4C).
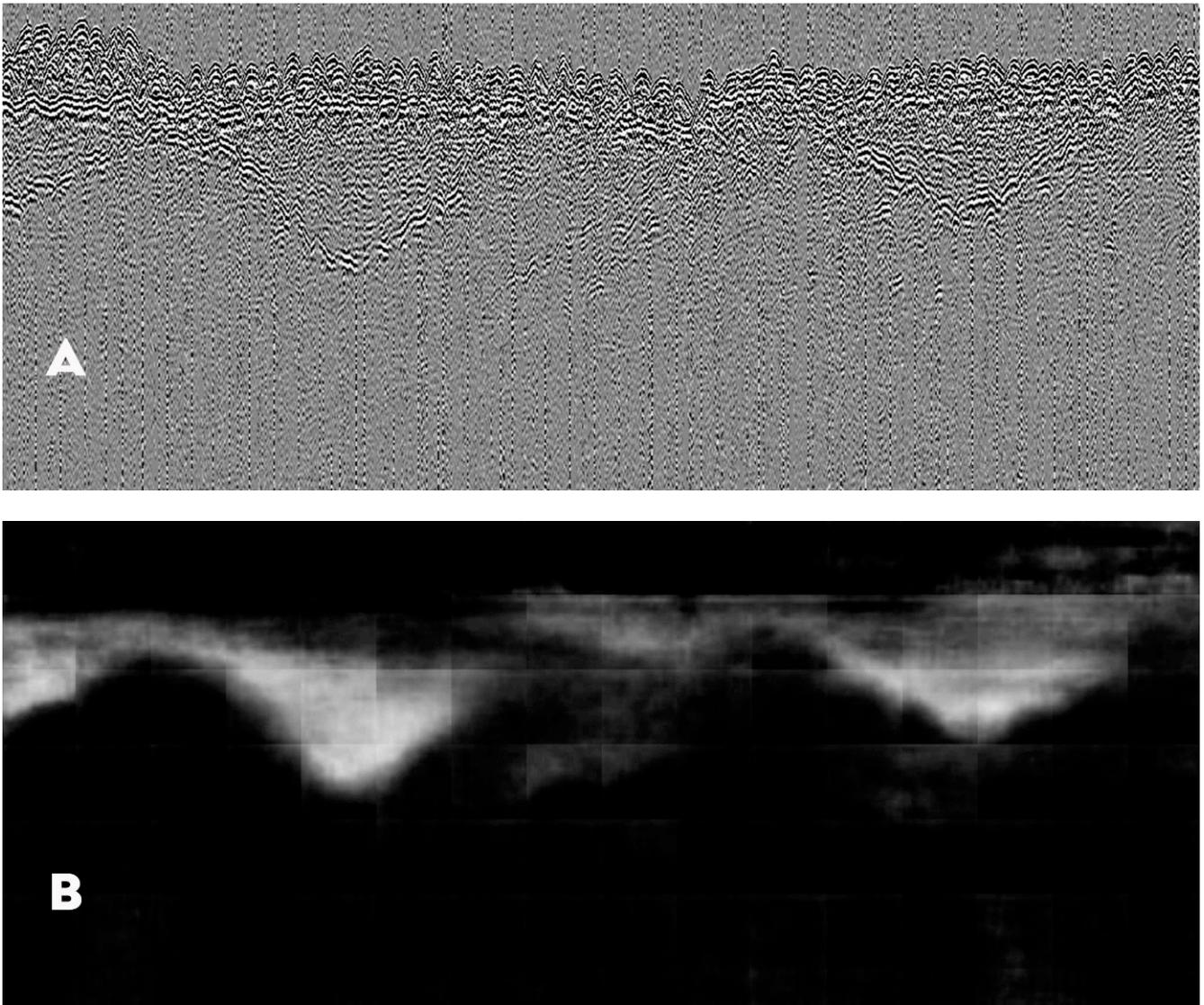


**Figure 4.** *Cont.*
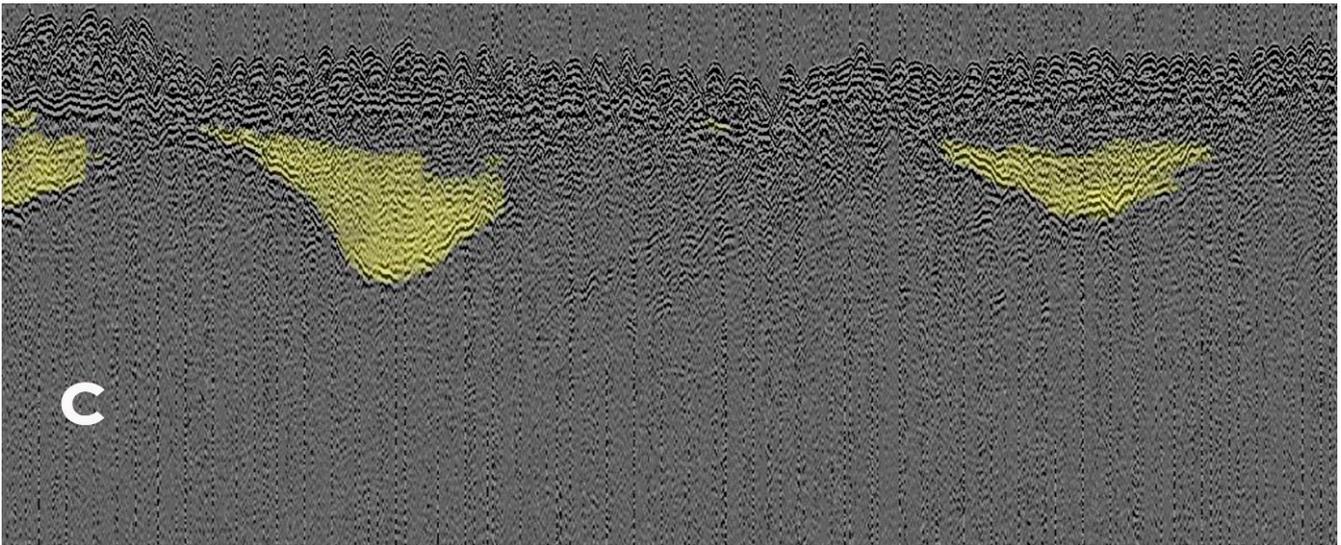
**Figure 4.** (**A**) source input image; (**B**) prediction, obtained after splitting into patches, then re-assembling results; (**C**) overlay of prediction over source image.

### 3. Evaluation of Results

After training, a randomly selected number of images were run to test the model. These were non-contiguous and were selected from a number of different surveys so that the CNN could not make inferences from one image to the next. Initially, the results were mixed with a high occurrence of both positive and false positive classifications. To overcome this, a process commonly referred to as model calibration was employed. Confidence thresholds were changed to identify the best possible detection rate or threshold at an acceptable number of false positives. The test results were then analysed by an experienced interpreter and categorized by successful identification. In Figure 5, correctly classified features are ringed in green, incorrectly classified (or partially classified) features in yellow, and false positives are blue. Red rings denote features that were missed by the CNN. No single threshold gave perfect predictions, but setting a 40% (0.4) confidence threshold returned a significantly larger proportion of positive detections to false positives. Repeated iterations of training and testing using this threshold level were found to improve the predictions still further.
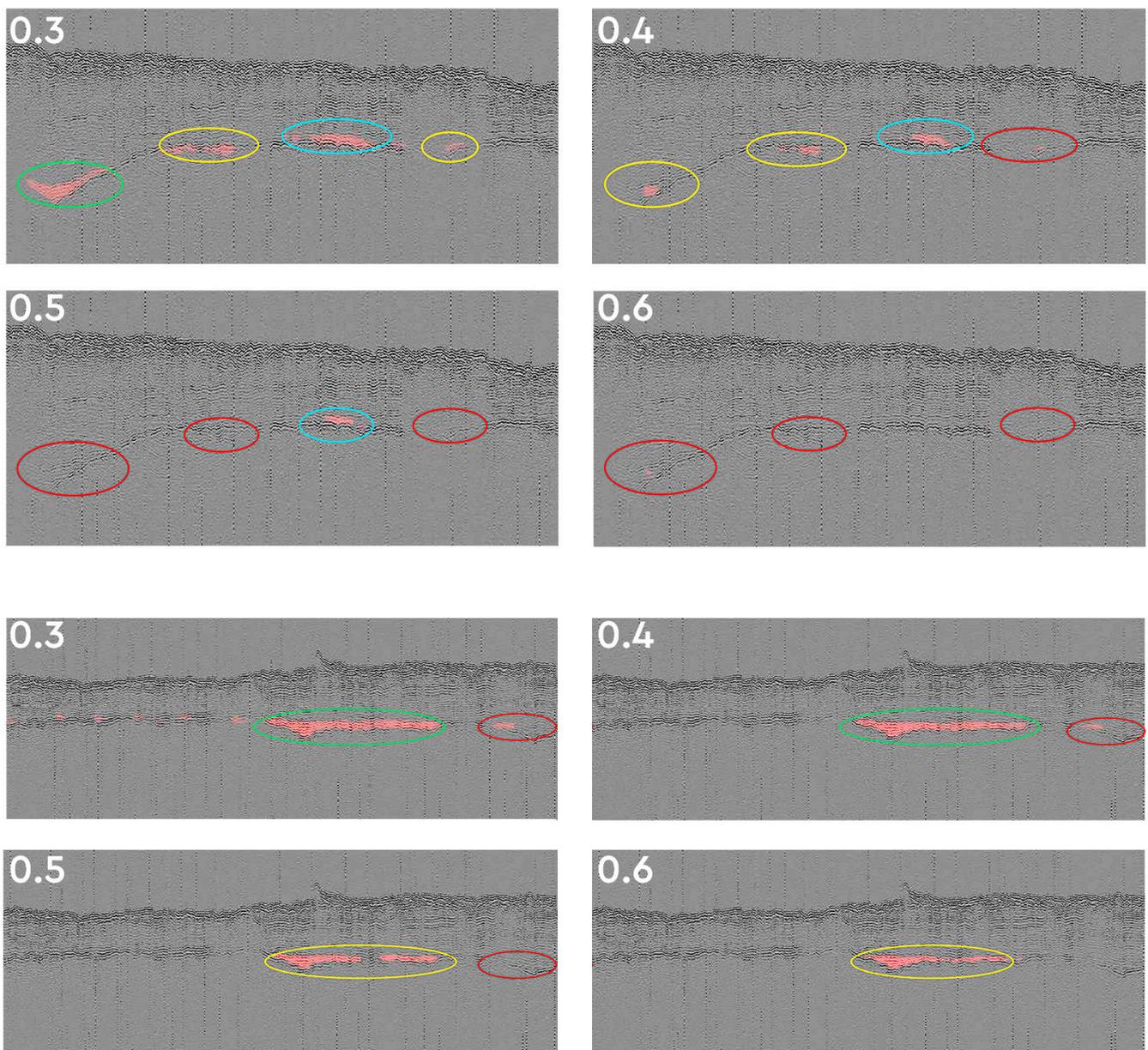
**Figure 5.** Predictive classification results using different confidence thresholds (0.3–0.6). Correctly classified (green), incorrectly or partially classified in yellow, false positives (blue) and features missed by the CNN (red).

When the predictive model was repeated at the 0.4 confidence threshold, the results demonstrated a suitably high level of accuracy. The model was able to reliably classify features in the sub-seafloor seismic dataset at reasonably accessible depths and containing sediment deposits.

The results of running the CNN at a confidence threshold of 0.4 can be seen in Figure 6. The top set (A) is of survey line 19_1EW_pt1, shotpoints 20,000–24,000; the set below (B) is of survey line BB7_WE_pt1, shotpoints 64,001–68,000. The upper images in both 6 A and B illustrate the simplified manual interpretation of a continuous horizon identified as a Holocene terrestrial surface. The green fill shows the extent of a sub-seafloor topographical feature, potentially a palaeochannel, and the dashed line shows potentially important sediment deposition. The lower image in each set is the output of the CNN that was run on the same seismic data but without the manual markup. The manual interpretations of these profiles were not provided to the CNN during training or the test runs.

While the CNN has not detected the full extent of the shallow features in either image, it has classified the general area and identified the continuous land surface horizon. The model has also attempted to identify possible sediment deposits within reach of a vibrocorer for sampling. It has missed the potential organic sediment outside of the filled feature and the deposition within the second image (B2) may be too deep to recover the highlighted sediment, at least in places, but the small feature to the left was within 3 m of the seabed and easily accessible by a 6 m vibrocore.

Figure 7 shows a manually interpreted seismic profile image (A) and the output of the CNN (B). Both images are from the same point of survey line Dredge2b_LF_16,000–20,000. Again, the CNN was run at a confidence threshold of 0.4. The merged output (B) clearly shows that the CNN was able to classify sediment deposition within the central feature at a depth that would be accessible to coring. The output also shows that the CNN correctly ignored high-amplitude signals that lay outside of the feature, as well as identifying a further, small feature to the right-hand side of the image. Significantly, this feature was missed or considered too small to be significant by the manual interpreter (A). However, the high-negative-amplitude signals suggest that this as a potential continuation of the peat layer outside the identified feature. This makes the result more significant in the fact that river banks were preferred to river valleys as occupation sites. It demonstrates the AI tool's ability to address our intention to identify such signals outside of, but in the same horizon context as, the features.
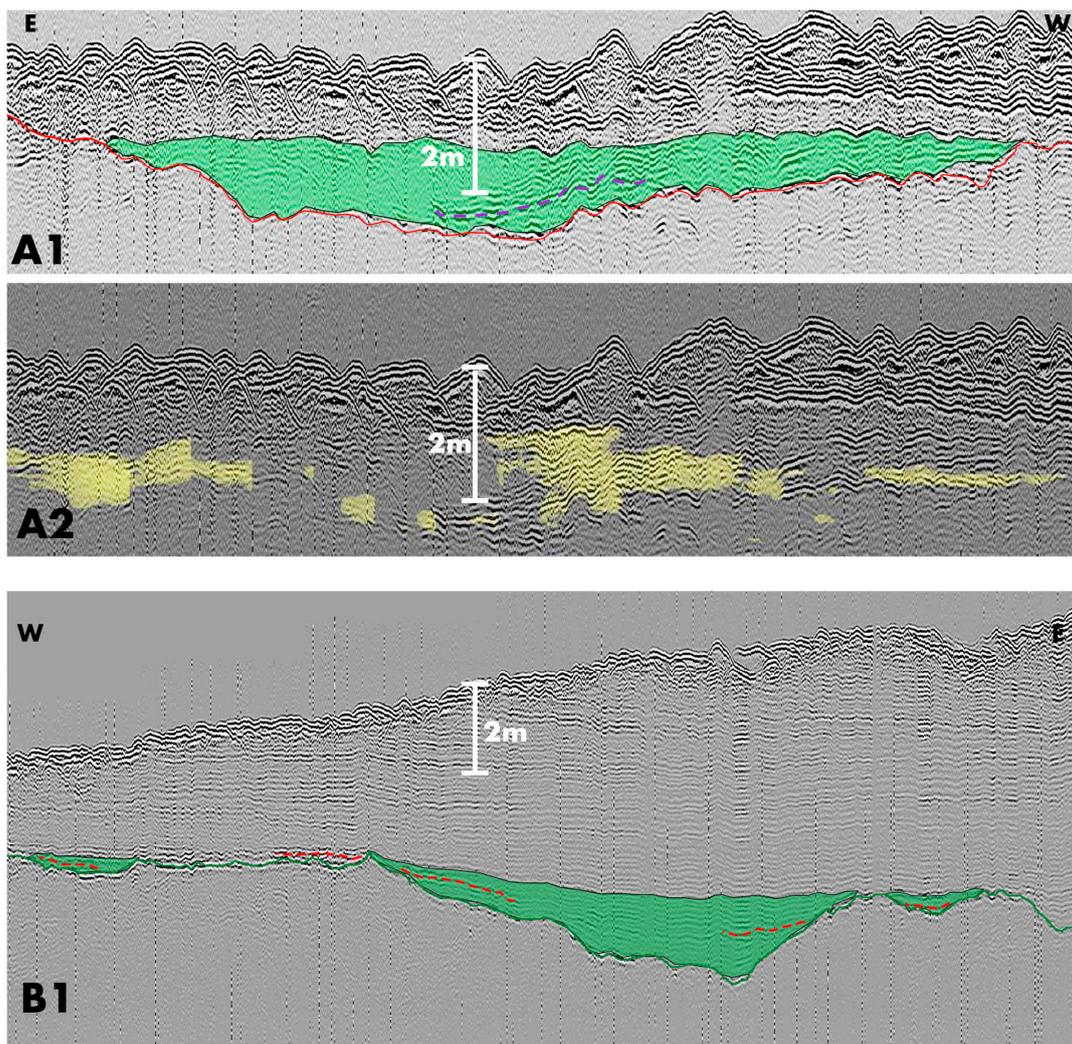


**Figure 6.** *Cont.*
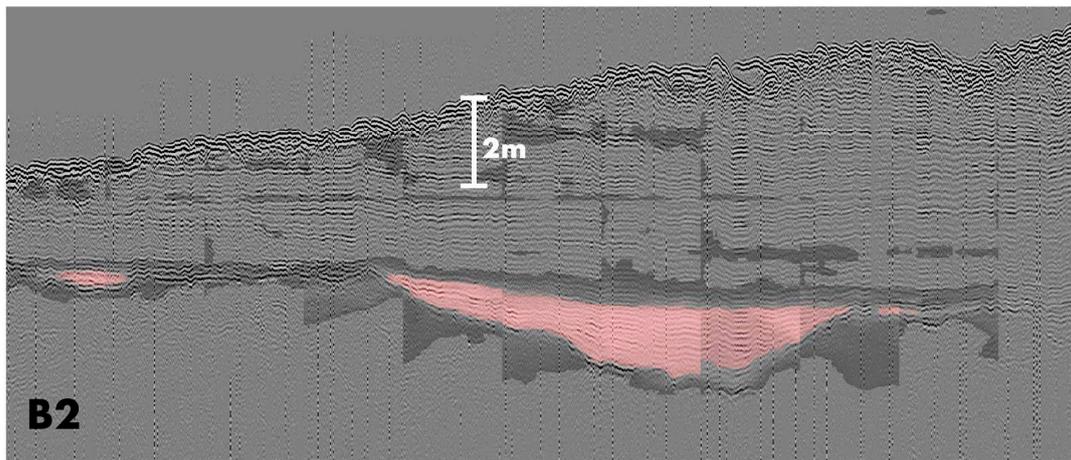
**Figure 6.** Two sets of images (**A**,**B**) showing a feature with depositional fill. Each set comprises a simplified manual interpretation (**A1**,**B1**) with features identified in green and potential organic deposition marked by a dashed line in purple (**A1**) or red (**B1**). The AI test results for the same images (**A2**,**B2**) pick the same features. The AI interpretation colours (yellow and pink) represent test runs using the different versions of the CNN.
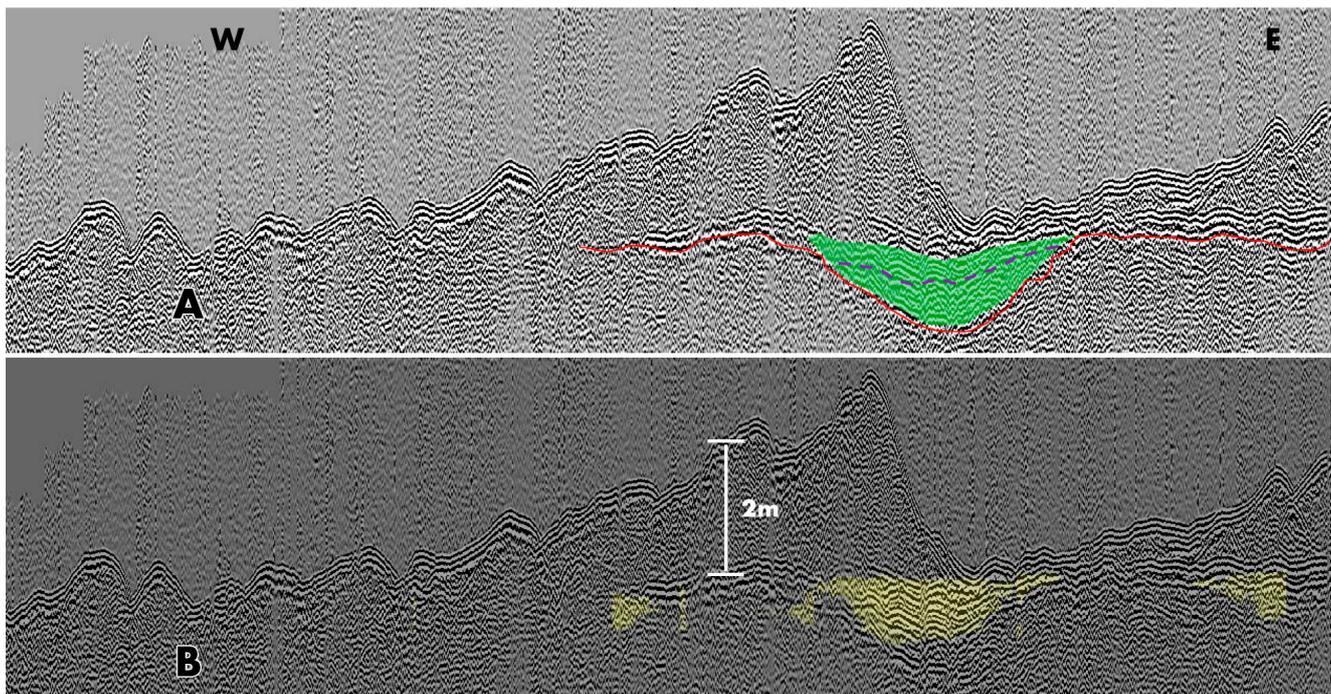


**Figure 7.** Two images of the same seismic profile (**A**,**B**). The upper shows a simplified manual interpretation with the continuous horizon picked out in red, a channel feature in green and significant deposition lines in dashed purple. The lower shows the AI interpretation in yellow. It is worth noting the deposition identified by the AI outside the channel feature as well as inside.

## 4. Discussion and Conclusions

The output from the AI model demonstrates that shallow seismic interpretation with a focus on archaeological potential is possible using these technologies. Although still at an early development phase, the method has proven adept at identifying features containing potential organic deposits at depths less than 6 m below the sea floor, which are accessible to sampling using a 6 m vibrocore. The identification of potential organic deposits in the same context but outside of such features proved more unreliable, and testing was unable to rate the success as highly positive. However, it was shown that the AI was

identifying the high negative amplitude signals and not merely following the features blindly, as shown in Figure 7. As a proof-of-concept model, the project demonstrates that the deep learning solution has considerable potential for the exploration of seismic data for archaeological purposes.

The model achieved these results using a number of state-of-the-art AI methods to identify and classify continuous facies and sediment deposition. Positive feature identification rates were high (up to 88% positive classifications vs. non-detection at a threshold of 0.4), with false positives being reduced to an acceptable level. The employment of semantic segmentation demonstrated that the model can be run using images of any size as inputs.

There are, of course, improvements that could be implemented in future work. The results suggest that, while the AI model was effective at feature identification where a strong negative-amplitude signal was present, the model still performed inconsistently in accurately detecting such sediments outside these features. Added to this, there is currently no way for the model to differentiate between organics in the form of materials such as wood, shallow gas pockets, or peat, as they all possess many of the same amplitude characteristics. In fact, images that demonstrated free gas blanking were removed to keep the model as focused as possible. Data in which image quality was considered an issue were not used. Early training of the model demonstrated that this may pose an issue, and further work is required to ascertain what level of image degradation would be acceptable. Any future work should aim to add complexity to the model to enable it to interpret a wider range of images. A quality control method could be implemented to bring in the entire image prior to segmentation to enhance precision and validate training. Before significant developments are implemented, there remains a real need to ground truth the current results of the model in the field.

Manual interpretation of the seismic profiles has provided consistently high success rates in discovering peat at an accessible depth, as evidenced by the coring missions performed by VLIZ in 2018 and 2019 [7]. If the CNN can be shown to have the capacity to replicate, or even improve on, the work of manual interpretation, the outputs of AI analysis will greatly assist researchers to locate areas of high archaeological potential and ensure that coring success rates remain high.

To this end, future iterations of the CNN should focus on running the model using 'live' survey data, as is acquired during a marine survey. If that were achieved, this would be a true breakthrough in shallow marine geophysics. A successful application would prove immensely useful when processing the vast swathes of data now becoming available following marine development. Such automation would allow for efficient processing of larger datasets, with an enhanced expectation of positive archaeological results.

# References

1. Münster, S.; Maiwald, F.; di Lenardo, I.; Henriksson, J.; Isaac, A.; Graf, M.M.; Beck, C.; Oomen, J. Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe. *Heritage* **2024**, *7*, 794–816. [CrossRef]
2. Character, L.; Ortiz, A., Jr.; Beach, T.; Luzzadder-Beach, S. Archaeologic Machine Learning for Shipwreck Detection Using Lidar and Sonar. *Remote Sens.* **2021**, *13*, 1759. [CrossRef]
3. Fitch, S.; Gaffney, V.; Harding, R.; Fraser, A.; Walker, J. A Description of Palaeolandscape Features in the Southern North Sea. In *Chapter 3, Europe's Lost Frontiers—Volume 1: Context and Methodologies*; Gaffney, V., Fitch, S., Eds.; Archaeopress: Oxford, UK, 2022; pp. 54–59. [CrossRef]
4. Astrup, P.M. Chapter 2 The Mesolithic in southern Scandinavia. In *Sea-Level Change in Mesolithic Southern Scandinavia: Long- and Short-Term Effects on Society and the Environment*; Jutland Archaeological Society; Aarhus Universitetsforlag: Aarhus, The Netherlands, 2018; Volume 106, pp. 20–28.
5. Peeters, J.H.M.; Amkreutz, L.W.S.W.; Cohen, K.M.; Hijma, M.P. *North Sea Prehistory Research and Management Framework (NSPRMF) 2019: Retuning the Research and Management Agenda for Prehistoric Landscapes and Archaeology in the Dutch Sector of the Continental Shelf*; Rijksdienst voor het Cultureel Erfgoed: Groningen, The Netherlands, 2019; Volume 63.
6. Amkreutz, L.; van der Vaart-Verschoof, S. (Eds.) *Doggerland. Lost World under the North Sea*; Sidestone Press: Leiden, The Netherlands, 2022; pp. 97–106, 209, ISBN 9789464261134.
7. Missiaen, T.; Fitch, S.; Harding, R.; Muru, M.; Fraser, A.; De Clercq, M.; Moreno, D.G.; Versteeg, W.; Busschers, F.S.; van Heteren, S.; et al. Targeting the Mesolithic: Interdisciplinary approaches to archaeological prospection in the Brown Bank area, southern North Sea. *Quat. Int.* **2021**, *584*, 141–151. [CrossRef]
8. Gaffney, V.; Allaby, R.; Bates, R.; Bates, M.; Ch'ng, E.; Fitch, S.; Garwood, P.; Momber, G.; Murgatroyd, P.; Pallen, M.; et al. Doggerland and the lost frontiers project (2015–2020). In *Under the Sea: Archaeology and Palaeolandscapes of the Continental Shelf*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 305–319.
9. Gaffney, V.; Fitch, S.; Bates, M.; Ware, R.L.; Kinnaird, T.; Gearey, B.; Allaby, R.G. Multi-proxy characterisation of the Storegga tsunami and its impact on the early Holocene landscapes of the southern North Sea. *Geosciences* **2020**, *10*, 270. [CrossRef]
10. Walker, J.; Gaffney, V.; Harding, R.; Fraser, A.; Boothby, V. Winds of Change: Urgent Challenges and Emerging Opportunities in Submerged Prehistory, a perspective from the North Sea. *Heritage* **2024**, 10–12, *preprint*. [CrossRef]
11. Louwe Kooijmans, L.P.; van der Sluijs, G.K. *Mesolithic Bone and Antler Implements from the North Sea and from The Netherlands*; Leiden University: Leiden, The Netherlands, 1971; ROB.
12. Glimmerveen, J.; Mol, D.; van der Plicht, H. The Pleistocene reindeer of the North Sea—Initial palaeontological data and archaeological remarks. *Quat. Int.* **2006**, *142*, 242–246. [CrossRef]
13. Cohen, K.M.; Westley, K.; Hijma, M.P.; Weerts, J.T. Chapter 7: The North Sea. In *Submerged Landscapes of the European Continental Shelf: Quaternary Paleoenvironments*; Flemming, N.C., Harff, J., Moura, D., Burgess, A., Bailey, G.N., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2017; Volume 1, pp. 152–166.
14. Phillips, E.; Hodgson, D.M.; Emery, A.R. The Quaternary geology of the North Sea basin. *J. Quat. Sci.* **2017**, *32*, 117–339. [CrossRef]
15. Törnqvist, T.E.; Hijma, M.P. Links between early Holocene ice-sheet decay, sea-level rise and abrupt climate change. *Nat. Geosci.* **2012**, *5*, 601–606. [CrossRef]
16. Splash-COS Database. Available online: http://splashcos.maris2.nl/ (accessed on 24 May 2020).
17. Vashist, P.C.; Pandey, A.; Tripathi, A. A Comparative Study of Handwriting Recognition Techniques. In Proceedings of the International Conference on Computation, Automation and Knowledge Management (ICCAKM), Dubai, United Arab Emirates, 9–10 January 2020; pp. 456–461. [CrossRef]
18. Mohan, M.; Jyothi, R.L. Handwritten character recognition: A comprehensive review on geometrical analysis. *IOSR J. Comput. Eng.* **2015**, *17*, 83–88.
19. Bahlmann, C.; Haasdonk, B.; Burkhardt, H. Online handwriting recognition with support vector machines—A kernel approach. In Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, Niagara on the Lake, ON, Canada, 6–8 August 2002; pp. 49–54. [CrossRef]
20. Berkson, J.M. Measurements of coherence of sound reflected from ocean sediments. *J. Acoust. Soc. Am.* **1980**, *68*, 1436–1441. [CrossRef]
21. Kreuzburg, M.; Ibenthal, M.; Janssen, M.; Rehder, G.; Voss, M.; Naumann, M.; Feldens, P. Sub-marine continuation of peat deposits from a coastal peatland in the southern baltic sea and its holocene development. *Front. Earth Sci.* **2018**, *6*, 103. [CrossRef]
22. Ball, J.E.; Anderson, D.T.; Chan, C.S. Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *J. Appl. Remote Sens.* **2017**, *11*, 042609. [CrossRef]
23. Verschoof-van der Vaart, W.B.; Landauer, J. Using CarcassonNet to automatically detect and trace hollow roads in LiDAR data from the Netherlands. *J. Cult. Herit.* **2021**, *47*, 143–154. [CrossRef]
24. Wrona, T.; Pan, I.; Gawthorpe, R.L.; Fossen, H. Seismic facies analysis using machine learning. *Geophys. J.* **2018**, *83*, O83–O95. [CrossRef]
25. Plets, R.; Dix, J.; Bates, R. *Marine Geophysics Data Acquisition, Processing and Interpretation: Guidance Notes*; English Heritage: London, UK, 2013. Available online: https://www.thecrownestate.co.uk/media/3917/guide-to-archaeological-requirements-for-offshore-wind.pdf (accessed on 8 April 2024).

26. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Lecture Notes in Computer Science. Springer: Cham, Switzerland, 2018; Volume 11211. [CrossRef]
27. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13001–13008. [CrossRef]
28. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988. [CrossRef]