A Comprehensive Deep Learning and Machine Learning Framework for Shoreline Prediction: Enhancing Accuracy and Interpretability for Dublin Bay, Ireland

MSc Research Project
Programme Name

Shamaie Maria Alleyne
Student ID:x24114570

School of Computing
National College of Ireland

Supervisor: Dr. Adbul Razzaq

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Shamaie Maria Alleyne
Student ID:	x24114570
Programme:	Programme Name
Year:	2025
Module:	MSc Research Project
Supervisor:	Dr. Adbul Razzaq
Submission Due Date:	11/12/2025
Project Title:	A Comprehensive Deep Learning and Machine Learning Framework for Shoreline Prediction: Enhancing Accuracy and Interpretability for Dublin Bay, Ireland
Word Count:	6812
Page Count:	24

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Shamaie Maria Alleyne
Date:	11th December 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A Comprehensive Deep Learning and Machine Learning Framework for Shoreline Prediction: Enhancing Accuracy and Interpretability for Dublin Bay, Ireland

Shamaie Maria Alleyne
x24114570

Abstract

Predictive shoreline forecasting is essential for safeguarding coastal communities and preserving dynamic shore ecosystems, which are subject to ongoing natural variation and environmental change. Traditional physics-based models and pure data-driven models lack the capability to accurately represent the rapid and nonlinear changes due to increased human influence and evolving environmental conditions. This study presents an innovative hybrid GeoAI pipeline designed to improve shoreline forecasting, in Dublin Bay as the case study. The proposed methodology involves a three-stage process. First, spatial features from Sentinel-2 satellite images collected between 2017-2025 will be used to train a pre-trained Masked Autoencoder (MAE). Secondly, the trained MAE will then be combined with a Transformer Encoder to identify complex, non-linear time series trends in the data. Finally, the resulting feature vectors will be integrated into a Light Gradient Boosting Machine (LightGBM) meta-learner, along with conventional physical attributes (e.g., tidal cycles, wave activity, wind direction and speed and DEMs).

A comparative analysis of the proposed hybrid model (Model C) showed that it delivered a notable improvement in performance, surpassing both traditional machine learning and deep learning models. Model C produced the largest amount of explained variance (R^2 = 80.07%) and had the smallest average absolute error (MAE = 1.13 m), a 7.1 cm reduction in MAE and a 1.89% increase in R^2 when compared to the classical model. The top features utilized by the hybrid model showed that deep learned features, generated by the Transformer Encoder, were significant contributors to the improved performance of the hybrid. In summary, the hybrid GeoAI framework introduced in this research has significantly improved both the predictive accuracy and interpretability of shoreline modelling. By combining spatial, temporal, and physical data, it addresses several weaknesses associated with previous methods and provides a robust tool for informing coastal management and resilience planning.

1 Introduction

Coastlines are highly susceptible to the impacts of environmental factors, it is crucial to develop robust, interpretable, and accurate forecasting models for the effective management of coastal zones (Kazanskiy et al.; 2025). Coastal erosion represents one of the most pressing environmental challenges facing coastal communities worldwide, with climate change accelerating the rate and unpredictability of shoreline change (Kazanskiy et al.; 2025). Accurate forecasting of shoreline evolution is needed for effective climate change mitigation and long-term strategic planning (Devoy et al.; 2008). Traditional techniques used for predicting shoreline movement typically relied upon limited field survey data, historical mapping or mathematical modelling using oversimplified physics (Lakku et al.; 2024). These methodologies allow researchers to establish broad trends over longer periods however, the limited spatio-temporal resolutions employed typically do not capture high frequency shoreline movements resulting from episodic events such as seasonal hydrodynamic changes. The emergence of satellite remote sensing technology, particularly the Sentinel-2 Earth observation mission, with its high spatial resolution and frequent revisit capabilities have provided researchers with the potential to generate high density, continuous time-series data that will be necessary for developing predictive modelling of shoreline movement (Christofi et al.; 2025).

1.0.1 The Gap in Current Forecasting Methodologies

There are main two primary limitations associated with current predictive methodologies in coastal geomorphology. First, purely physics-based or traditional statistical models that attempt to predict shoreline movement from complex multivariate environmental time-series data often fail to adequately address the inherent nonlinearities and interdependencies that exist within the dataset, such as the relationships between lagged wave height, tide or sand transport (Mathonsi and Van Zyl; 2021). Secondly, although recent developments in Deep Learning models (particularly Recurrent Neural Networks [RNNs] and Transformers) have shown great promise in identifying temporal relationships within large datasets. The complexity of these models makes it difficult to understand how predictions are made (the black box effect) (Joaquin and Javier; 2023). As such, there appears to be a methodological gap for a system that is able to adequately incorporate multiscale relationships (spatially, temporally and statically) while maintaining a high degree of interpretability.

1.0.2 Dublin Bay a Complex Test Environment

Dublin Bay, Ireland, provides a particularly challenging yet relevant case study because of the complex heterogenous nature of the coastal dynamics that occur within the bay (Devoy et al.; 2008). Previous studies indicate that there have been significant changes to the shoreline within the bay based on various secondary data sources (Gibson et al.; 2012). An initial analysis of 159 transects along the shoreline in Dublin Bay yielded an average net accretion trend, with a Mean Erosion Rate of +0.99m/year. However, this overall stability masks localized volatility, while the overall trend of the shoreline may be generally stable, some areas are experiencing erosion while other areas are rapidly accreting or remain relatively stable. This level of heterogeneity necessitates the use of a model that can learn to identify the unique rules governing each segment of the shoreline based on both the dynamic environmental conditions and the geospatial properties of the region.

1.0.3 Rationale for the Hybrid GeoAI Pipeline

This study aims to explore a hybrid approach that incorporates the strengths of multiple modelling theories in order to overcome the previously identified limitations (Zhou et al.; 2023). The pipeline is implemented as a Stacking Ensemble consisting of three specialized and sequential stages:

- **Stage 1: Spatial Feature Extraction** - In this stage the raw high resolution Sentinel-2 Imagery is used as input for the CNN/MAE-Unet to transform the complex visual data (Spectral and Textural Information) into a low dimensional feature space (Kong et al.; 2020). These spatial deep features represent the visual condition of the coastline such as wet/dry line visibility, sand exposure etc. as a set of quantitative values.

- **Stage 2: Non-Linear Temporal Modeling** - At this stage, the pipeline uses the timeseries of all available features (spatial and environmental) in sequential form as inputs to the Transformer Encoder. The Transformer Encoder utilizes its self-attention mechanism to capture non-linear features with the objective of addressing the problem of classical models not being able to synthesize complex long range temporal dependencies and non-linear dynamics inherent in coastal processes robustly (Oliveira and Ramos; 2024).

- **Stage 3: Feature Fusion** - In this final stage, LightGBM utilizes the fused feature set (classical features and environmental features and optimizes the prediction of the shoreline position ($\mathbf{S_t}$). The ability of LightGBM to provide transparent feature importance ensures the contributions of the deep learned signals can be quantified and justified addressing the "black box" criticism of pure deep learning approaches (Joaquin and Javier; 2023).

1.0.4 Research Overview and Contributions

The central research question driving this study is: "Can a fusion pipeline that uses a pre-trained CNN to extract shoreline/spatial features from Sentinel imagery, models the temporal evolution of the time series using a Transformer and fuses those learned features with environmental inputs via LightGBM produce more accurate and more interpretable shoreline forecasts in Dublin Bay than either pure deep learning or pure classical approaches?".

This research offers an empirical validation of a three-stage hybrid Geo-AI model (CNN + Transformer + LightGBM) to predict shorelines and demonstrates significantly higher predictive accuracy than non-deep learning models and increased interpretability than pure deep learning models. This hybrid model is able to extract relevant spatial features by utilizing pre-trained feature extraction networks from Sentinel-2 satellite images, and is effective in extracting the complex temporal relationships present within the data, through its use of the transformer module. A robust comparative benchmark was established using data from Dublin Bay; the results demonstrate that this hybrid model is a viable and superior option for both short-term and long-term operational shoreline forecasting applications.

The subsequent sections detail the related work, specific methodology, design specifications, implementation, comprehensive evaluation and conclusions derived from this validation.

2 Related Work

2.1 Foundational Concepts

Luijendijk et al. (2018) defines Coastal erosion as the removal of lands along coastlines due to naturally occurring forces such as waves, wind-driven water, tides and storms. A shoreline is described as the zone of transition between land and water masses and as such it is inherently dynamic and reactive to natural forcing processes and man-made processes (Boak and Turner; 2005). Vos et al. (2019) defines shoreline position as a geospatial feature which varies temporally and whose behavior is non-linearly influenced by hydrologic, atmospheric and geomorphic processes. Shoreline position over time can be represented by the equation

$$S_t = f(S_{t-1}, E_t, M_t)$$

where S_t represents the shoreline position at time t, E_t represents the environmental driving conditions (waves, tides, wind) and M_t represents the morphologic characteristics of the coastal system (Castelle et al.; 2021). This representation captures the inherent auto-correlative nature of coastal changes while allowing for incorporation of external driving forces.

Satellite remote sensing has evolved into the primary scientific means for systematically monitoring shorelines over large geographic regions (Hagenaars et al.; 2018). The European Space Agency's Sentinel-2 mission, launched in 2015, consists of two satellites that provide multispectral imagery at 10-metre spatial resolution with a 5-day revisit interval, providing a unique opportunity for high-frequency sampling of coastal systems (Drusch et al.; 2012). Shoreline extraction from satellite imagery involves applying thresholding techniques to spectral indices, such as the Normalized Difference Water Index (NDWI) (McFeeters; 1996). The position of the shoreline was consistently identified using the Normalized Difference Water Index (NDWI). NDWI effectively differentiates water from other surface features through the use of the unique spectral characteristics of Green and Near-Infrared (NIR) bands. NDWI provides the methodological foundation of the primary regression target time series ($\mathbf{S_t}$).

2.2 Deep Learning and Machine Learning Architectures

Deep learning involves the use of Artificial Neural Networks (ANNs) with many layers to create hierarchies of data representation (LeCun et al.; 2015). Convolutional Neural Networks (CNNs) developed by (Lecun et al.; 1998), are used for image recognition and other tasks involving grid-structured data using learnable convolutional filters.

The practice of transfer learning which is taking a pre-trained model trained on a very large data set and then adapting it to a specific problem has become commonplace in computer vision with limited amounts of training data (Pan and Yang; 2009). These include pre-trained CNNs such as ResNet, VGG (Simonyan and Zisserman; 2014) and EfficientNet (Tan and Le; 2019) which can be viewed as feature extractors and can convert images into compact representations that can be fed into a variety of downstream tasks.

The Transformer uses self-attention mechanisms to determine how important one time step is relative to another time step during prediction which allows for the discovery of long-term relationships in temporal data without suffering from vanishing gradients (Vaswani et al.; 2017). In a study by Zhou et al. (2021), Transformers performed better than LSTMs on typical time series forecasting benchmark data especially for time series that exhibit complex temporal relationships.

Ensemble learning combines the outputs of many models to produce a single output that is better than the individual models could produce (Dietterich; 2000). Gradient boosting is a form of ensemble learning which builds models one at a time. Each successive model is built to reduce the error made by the previous model (Friedman; 2001). The most well-known method for gradient boosting is LightGBM (Ke et al.; 2017) which is extremely fast because it uses a combination of histograms to index instances and grows trees bottom-up. Gradient boosting decision trees (GBTs) were demonstrated by Chen and Guestrin (2016) to always perform better than deep neural networks on tabular data therefore, they have become a preferred choice of in my studies working with unstructured and structured features alike.

Hybrid machine learning architectures combine aspects of different machine learning paradigms in order to take advantage of the unique benefits of each (Sun et al.; 2017). One common hybrid architecture consists of first applying deep learning techniques to obtain features from unstructured data (text, imagery) and then apply classical machine learning techniques to structured prediction tasks. The two-stage approach utilizes the ability of deep learning to learn representations of data while utilizing the interpretable and computationally efficient nature of gradient boosting on structured data. A generalization of this paradigm was formalized by Webb et al. (2016) who called it "deep feature extraction followed by shallow learning" and demonstrated that these types of architectures typically perform better than fully connected deep learning architectures when dealing with data from many different sources and when there is limited training data.

2.3 Evolution of Shoreline Forecasting Models

Statistical extrapolation, typically using a linear rate based on historical shoreline changes, is the main method for predicting future shoreline shifts. While quick to compute, it ignores nonlinear trends and environmental factors. More advanced approaches involve applying time-series analysis techniques, Kalman filtering and ARIMA models (Maiti and Bhattacharya; 2008), which enables the model to account for temporal autocorrelation. However, limitations exist in these methods as they do not fully represent the complex and multi-scale nature of coastal systems. Physics-based coastal models (XBeach - Roelvink et al. (2017), Delft3D - Lesser et al. (2004), CSHORE, Kobayashi et al. (2007)) solve hydrodynamic and morphodynamic equations to simulate sediment transport processes. These models are physically based, but they require detailed bathymetry, significant amounts of computer memory and processing time to run simulations (Roelvink et al.; 2017). Toimil et al. (2020) found that physics-based models are highly accurate when forecasting shoreline changes for short periods however errors increase with time (seasonal to inter-annual), mainly due to uncertainty associated with the parameters and processes that are not resolved. In addition, due to the computational requirements these types of models cannot be used for rapid and large scale forecasting applications (Vitousek et al.; 2017). Machine learning algorithms have recently been used to develop pure data driven approaches to predicting shoreline change. Beuzen et al. (2018) used Random Forests to predict volume of erosion of beaches based on wave conditions, tidal levels and past beach profile measurements, and achieved an R^2 value of 63%. Goldstein et al. (2019) developed gradient boosting machines to predict shoreline change at 38 beaches along the coast of

California, and reported a mean absolute error (MAE) of 3.2 m for quarterly predictions. Pure machine learning approaches show that it is possible to learn statistically relevant relationships between environmental forcing variables and coastal responses from observational data without developing an explicit model of the processes involved. However, these approaches typically require the researcher manually extract useful features from the data and thus cannot take advantage of image information available from satellites (Vitousek et al.; 2017).

Table 1: Comparison of Related Shoreline Forecasting Studies

Study	**Study Area**	**Method**	**Input Data**	**Performance / Limitations**
Douglas & Crowell (2000)	U.S. Atlantic Coast	Linear Regression	Historical shoreline positions	Assumes linear trends; ignores environmental forcing
Roelvink et al. (2009)	Various (XBeach)	Process-based numerical model	Bathymetry, waves, tides	RMSE: 5–20 m (event scale); high computational cost; requires extensive calibration
Beuzen et al. (2019)	Narrabeen Beach, Australia	Random Forest	Wave height, tidal elevation, antecedent profile	$R^2 = 0.63$ (volume); hand-crafted features; no imagery input
Vos et al. (2019)	Global (CoastSat)	CNN + thresholding	Sentinel-2 imagery	RMSE: 10 m (extraction); extraction only; not forecasting
Goldstein et al. (2019)	38 California beaches	Gradient Boosting	Wave parameters, tide, beach slope	MAE: 3.2 m; limited spatial coverage; no deep learning
This Study	**Dublin Bay, Ireland**	**CNN + Transformer + LightGBM (Hybrid)**	**Sentinel-2 imagery, waves, tides, wind, DEM**	**Target: MAE < 5 m; first hybrid 3-stage architecture applied to shoreline forecasting**

Remote sensing applications in coastal areas are being addressed increasingly using deep learning techniques. CNNs were used by Buscombe and Ritchie (2018) to automate the classification of beach sediments from imagery. Vos et al. (2019) developed CoastSat by utilizing CNNs for the extraction of shorelines from Landsat and Sentinel-2 imagery. However, these applications focus on extracting the shoreline rather than forecasting. Goldstein et al. (2019) employed Temporal Convolutional Networks (TCNs) for the prediction of wave run-up, which demonstrated that deep sequence models can capture complex temporal dependencies in coastal processes.

Zhou et al. (2021) introduced Informer, a variant of the transformer optimised for long-term time series forecasting, which demonstrated superior performance over LSTMs on weather and energy forecasting tasks. Wu et al. (2021) applied transformers to rainfall-runoff modelling, which achieved an NSE of 0.87 compared to 0.81 for LSTMs. Machine Learning architectures have achieved state-of-the-art results in domains such as weather forecasting (Reichstein et al.; 2019), ecological modelling (Christin et al.; 2019), hydrological prediction ((Kratzert et al.; 2019)), however, their applications in coastal forecasting is limited. Sun et al. (2017) illustrated that hybrid architectures combining CNNs for spatial feature extraction with gradient boosting for structured prediction outperformed pure deep learning approaches for cropland mapping (accuracy: 94% vs. 89%). This finding suggests potential benefits of similar architectures for coastal applications.

2.4 Research Gaps Identified

There are five key gaps that need to be addressed based on existing literature.

- There has been limited exploration of satellite imagery in forecasting. Although CNNs have been used to extract shorelines from images (Vos et al.; 2019), the spatial representations that were learned from coastal imagery have not been incorporated into features for forecasting models. Most current forecasting methods use hand-crafted shoreline position metrics instead of learned image features.
- The under-exploitation of transformer architectures. Transformers have demonstrated better performance than LSTMs in many different domains (Wu et al.; 2021; Zhou et al.; 2021). The attention mechanism of transformers will allow the model to determine which time steps are most important for identifying periodicity in seasonal coastal variability.
- The lack of comprehensive feature fusion. Studies have employed either environmental forcing data (Goldstein et al.; 2019) or spatial imagery to develop predictive models for coastal systems, but rarely incorporate multiple forms of data in a systematic way.
- The lack of systematic comparison between paradigms. Only a few studies have compared pure classical machine learning, pure deep learning and hybrid models on identical datasets with consistent evaluation metrics. As such it is difficult to make conclusions regarding what types of architectures work best for coastal forecasting.

3 Methodology

This research mainly aims to quantitatively assess how well the hybrid pipeline performs. Specifically, this study sought to demonstrate empirically that the integration of spatial and temporal information developed by using deep learning methods will result in a statistically and practically meaningful improvement in forecast accuracy (MAE and R-squared) when compared to pure classical machine learning or pure deep learning methods used separately.

3.1 Tasks Performed to Complete the Study

A rigorous 7-step process, as shown in Figure 1, ensured quality of the data, feature engineering accuracy and robustness within model comparisons.

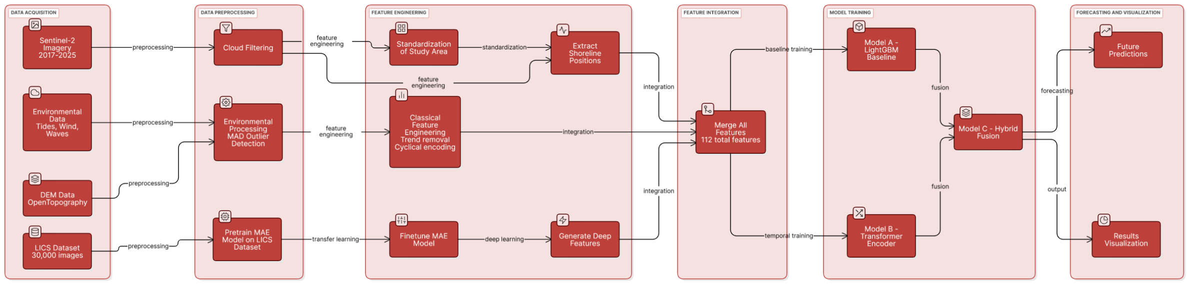

Figure 1: Research Methodology

1. **Data Collection and Standardization**: Historical Sentinel-2 quarterly imagery between 2017 and 2025 was downloaded from Google Earth Engine to establish a time series of coastal positions. As optical data can be inherently corrupted; strict cloud filtering was employed as such from the 23 original images, 17 high quality images were utilized to ensure quality over quantity. To standardize the study area, we established 159 transects, providing spatially equivalent measurement points.

2. **Shoreline Position Extraction**: Using the NDWI methodology, the shoreline position ($\mathbf{S_t}$) for each transect was extracted and the perpendicular distance from a fixed baseline was calculated. In addition to determining the target regression variable, the perpendicular distance calculation provided a total of 2703 measurements for all time points.

3. **Classical Feature Integration and Engineering**: Data was collected and combined from multiple types of diverse sources. Data was gathered and combined from multiple types of diverse physical inputs: tidal data from the Irish National Tide Gauge Network, wind data from Met Eireann, wave data from the Irish Weather Buoy Network, and static DEM features from Open Topography. Following which, advanced preprocessing techniques were applied to remove outliers (using Median Absolute Deviation), to remove trends (so that only nonlinear residuals can be modeled with tree-based models), cyclical encoding to convert seasonal cycles into continuous variables and environmental lagging (for example, use wave data from $t-1$ to estimate ΔS_t at t), to ensure causality and avoid data leakage .

4. **Deep Feature Generation**: A MAE-Unet pre-trained model that was fine-tuned on Dublin Bay data was utilized due to its superior performance over the other semantic segmentation models in this study (IoU 94.2). The MAE-Unet model produced the best set of deep spatial features. Following which, the encoder portion of this model was applied to the 17 Sentinel-2 images to produce four new deep spatial features (mae) for each transect-time step, further increasing the number of total features from 108 to 112.

5. **Model A Training**: A pure Machine Learning benchmark was established using LightGBM to predict the Absolute Shoreline Position ($\mathbf{S_t}$) using all 112 engineered features (including the 4 deep features, treated as standard classical inputs).
6. **Model B Training (Signal Extraction)**: A Transformer Encoder was trained to isolate the non-linear temporal relationships by predicting Shoreline Change ($\mathbf{S_t}$) on sequence data.. Lastly, the final hidden layer of the best performing transformer was utilized to produce the 32 latent temporal features ($\mathbf{X_{DL}}$).
7. **Model C Fusion**: In this final stage, the combined 32 ($\mathbf{X_{DL}}$) features were fused with the 97 ($\mathbf{X}_{\text{Classical}}$) features that were not sequential inputs (e.g. DEM, static cluster metrics). A new LightGBM model was trained on the total 129 fused features to produce the final predicted value of ($\mathbf{S_t}$).

3.2 Dataset Characteristics and Preprocessing

The empirical findings for this study relies on the integration of multiple diverse datasets such as remote sensing imagery, static topography and dynamic environmental inputs, to provide a comprehensive key understanding of the factors that influence shoreline change in Dublin Bay.

3.2.1 Remote Sensing Imagery

The primary dataset for spatial feature extraction is a combination of two datasets utilized for its deep features.

1. LICS Dataset (Initial Pre-training): The Landsat Irish Coastal Segmentation (LICS) dataset (O'Sullivan and Dev; 2024) was used for Transfer Learning (Masked Autoencoder pre-training). This dataset provided the model with a diverse set of 30,000 images of the entire Irish coastline, enabling it to learn robust, generalizable coastal features that can be applied broadly within any coastal environment.
2. Dublin Bay Quartely Time Series: After stringent cloud filtering, 17 high quality images were extracted from Sentinel-2 via Google Earth Engine. The relevance of these Sentinel-2 images images would be maximized by its high spatial resolution (10 m) a key component to accurately define shoreline positions. Its temporal consistency providing a consistent, verifiable time series of quarterly images from 2017 - 2025 and lastly its spatial standardization. The images were evaluated across 159 transects perpendicular to the coastline to allow for comparisons to be made between the different transects for the purpose of developing the machine learning models.

3.2.2 Environmental and Topographic Inputs

To provide contextual and predictive inputs into the model, time series and static datasets where integrated. This is a key combination to ensure the models extend beyond the single snapshot provided by the satellite imagery.

3.2.3 Advanced Classical Feature Engineering

Advanced preprocessing techniques were applied to the environmental inputs in Table 2 to remove outliers and clean the data for time series forecasting.

Table 2: Environmental Characteristics

Data Type	Data Source	Date Range	Records	Features
Tidal Inputs	Marine Institute Irish National Tide Gauge Network	2007–2025	2+ million	**7:** Captures short-term water level changes, directly influencing instantaneous shoreline position.
Wind Inputs	Met Éireann Dublin Airport hourly data	1945–2025	700,000+	**10:** Annual wind speed; wind stress generates waves and currents.
Wave Inputs	Marine Institute Irish Weather Buoy Network	2017–2025	2,888	**9:** Quarterly wave height; main driver of shoreline change.
Digital Elevation Model (DEM) Inputs	Open Topography	Static	1	**5:** Mean, min, max, range; static geospatial properties for Dublin Bay.

1. **Lagged Environmental Inputs**: Lagging is critical in time-series analysis to prevent data-leakage and to enable the inertial effects of environmental drivers to be modeled. All dynamic inputs (wave, wind) are lagged to ensure that each data point is an input from a previous time (e.g. using wave data from $t-1$ or $t-2$ to predict ΔS_t at t). This process enables the explicit modeling of the physical reality that wind and wave driven erosion/accretion are cumulative processes that require time to become measurable changes in shoreline position.

2. **Outliers in Training Data**: To remove outliers (due to tides, storm events or sensor error), Median Absolute Deviation (MAD) was used to filter out the extremes and create a stable training set. With MAD applied, this ensured the model was robust to noise preventing extreme values from corrupting the learning process.

3. **Cyclical Representation of Quarterly Seasonal Values**: Using the sin and cos of each quarter allows the model to see seasonal patterns as cyclical and provides a mechanism for modeling seasonal effects on coastal erosion (such as increased erosion due to wave energy during winter months).

The pipeline's hybrid nature is defined by the integration of features from two distinct Deep Learning models with the classical inputs.

1. **Spatial Deep Features**: The integration of the CNN encoder, extracted from the fine-tuned MAE-UNet, reduces the high-dimensional complex Sentinel-2 imagery into low dimensional spatial features (mae_f). These features represent abstract, learned information regarding coastal morphology (e.g. sand bar presence, beach texture) that cannot be captured by the simple NDWI analysis.

2. **Temporal Deep Features**: The outputs from the trained transformer encoder effectively reduced the non-linear temporal relationships within the sequence history of the shoreline changes into 32 latent temporal features. These represent a learned 'summary' of the temporal evolution and relationships of the shoreline changes and is an input into Model C.

3.2.4 Deep Feature Merging Strategy (Dual Key Merge)

This research evaluated how different methods for creating new datasets could be used together to create a larger dataset to use for future research, comparing the Triple Key Merge, which created a dataset that was high-quality but low-density, to the Dual Key Merge, which generated the densest dataset. This was achieved by applying the mean of deep spatial features over all transects within the same year/quarter then merged the deep features into every single classical row. The final model training was conducted utilizing the dense dataset, the Dual Key Merge method.

Although the decision to train the final model on the Dual Key Merge's denser dataset did come with a small amount of risk of spatial contamination (i.e., the fact that there were some non-unique features associated with each transect) this was deemed acceptable because the resulting model would have a large volume of training data available to it (2,703 points). The large volume of training data enabled the LightGBM meta-learner to perform well over many features and the inherent ability of LightGBM to be able to handle noise and minor levels of feature contamination in its average deep feature vector allowed researchers to consider the averaged deep feature vector to be a very reliable, low-noise representation of each scene's most important attributes at each time period. As such, this combination of methods represented a practical and efficient solution for completing the predictive modeling task.

3.3 Algorithm and Models Utilized

Three models were utilized comparatively in this fusion approach, each designed to fulfilling a specific role in validating the hybrid research question.

- **LightGBM (Models A and C)**: A Gradient Boosting Machine framework selected for its efficiency and speed of computation, ability to handle high dimensional/sparse data and most importantly, its capability to produce auditable feature importance. LightGBM was employed as the robust meta-learner in the stacking ensemble (the Model C hybrid).
  - Model A Role: Provide a baseline classical machine learning performance benchmark.
  - Model C Role: Fuse the non-linear feature sets ($\mathbf{X_{DL}} \oplus \mathbf{X}_{\text{Classical}}$).
- **Transformer Encoder (Model B)**: A custom neural network used in sequence modeling. A Self-Attention Mechanism was employed to capture long range temporal dependencies that may not be detectable using traditional LSTMs or CNNs, such as the subtle, quarterly effects of cumulative environmental stress on shoreline recovery.

3.3.1 Target Variables and Training Strategy

The distinction of target variables was the central component to the model design.

- **Absolute Shoreline Position ($\mathbf{S_t}$)**: Used for the final, practical predictions (Models A and C). Since this target has an overwhelmingly greater amount of variance that is explained by the previous position $S_t - 1$, it is relatively simple target to predict.

- **Shoreline Change (ΔS_t)**: Used only for training Model B (Transformer). By eliminating the high autocorrelation of S_t - 1, the model will be forced to focus its available learning capacity to extract the difficult, non-linear signal caused by the dynamic environmental factors that cause the fluctuations in the shoreline.

To overcome the fundamental tradeoff between bias and variance in machine learning, the pipeline's architecture was specifically structured to include the following.

- **Bias Reduction (Underfitting)**: The stacking ensemble architecture utilized multiple models each suited for specialized specific signal types to mitigate bias. Model B (Transformer) serves as a high pass filter, learning the complex nonlinear residual signals missed by the traditional Classical model. The LightGBM meta-learner further reduces bias by optimally combining the complex residual signals, achieving a more flexible, lower-bias final fit. Feature engineering techniques like trend removal explicitly ensure the LightGBM models learn the non-linear deviations (residuals) rather than being constrained by the assumption of a linear trend.
- **Variance Reduction (Overfitting)**: Both LightGBM models (Models A and C) is are ensemble models, therefore they inherently have less variance compared to individual decision trees. Both LightGBM models contain built in regularization (L1 and L2) to limit the complexity of the model and to avoid overfitting of the training set this limits the potential of poor generalization to unknown test data. In addition, using separate training and testing datasets for each period of time (temporal cross-validation) ensures that no data leakage occurs and that the results were not caused by the models having too much variance.

3.4 Evaluation Metrics

The effectiveness of the models were rigorously assessed with the use of standard regression metrics, that provided both a measure of how well the model fits the data and a measure of the magnitude of the error.

- **Mean Absolute Error (MAE)**: This is the primary evaluation metric, measures the average magnitude of error in meters (1.130 m for Model C). MAE provides the best measure of positional accuracy for coastal engineering applications and is represented by

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

  where y_i is the actual value, $\hat{y}_i$ is the predicted value, and N is the number of observations.

- **Coefficient of Determination (R^2)**: Measures the proportion of the variance in the dependent variable that is explainable from the independent variables (80.07% for Model C). R^2 is important for comparing the overall quality of fit of the model and for explaining the incremental gain in quality of fit resulting from the hybrid fusion. It is crucial for comparing the overall Model C's fit relative to the classical Model A thus demonstrating a model's explanatory power

$$\mathbf{R^2} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the actual values.

- **Root Mean Square Error (RMSE)**: Is highly sensitive to large errors (1.791 m for Model C). The improvement in RMSE of Model C relative to Model A (2.381 m) demonstrates that the hybrid model is better suited than Model A to predict high-impact outlier events such as storms. A low RMSE indicates effective management of outlier observations is represented as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

4 Design Specifications

The hybrid pipeline is a multi-stage data processing and modelling architecture that has been optimized for maximum feature isolation and effective fusion.

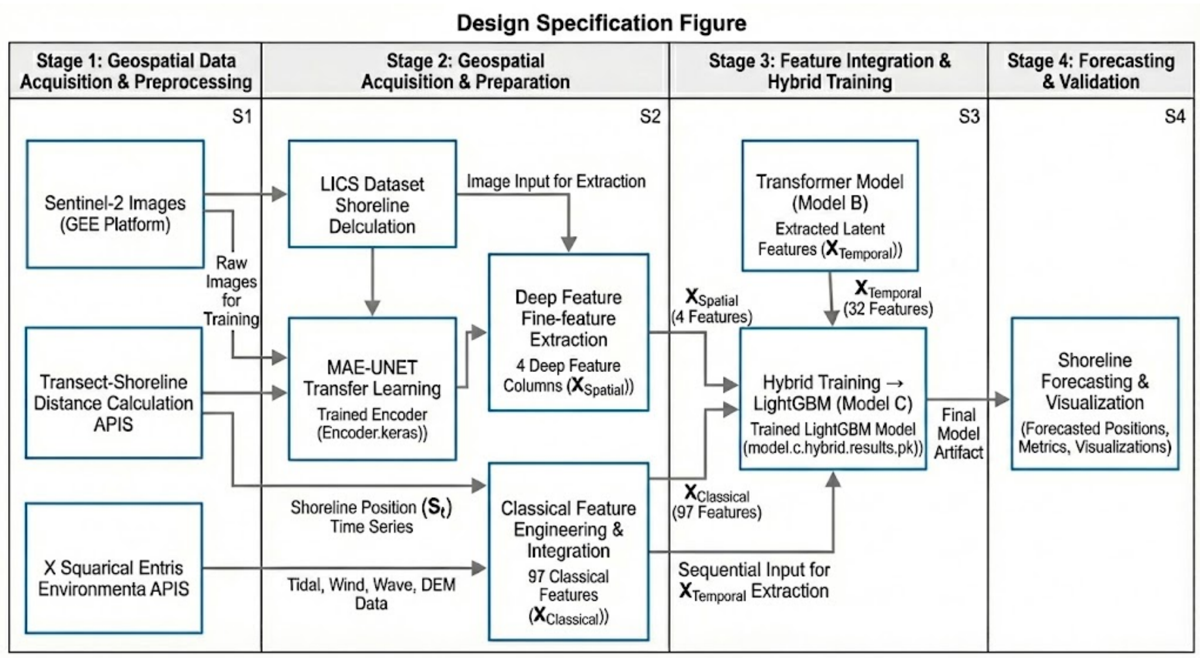

Figure 2: Hybrid Architecture

4.1 Spatial Feature Extraction Architecture (MAE-UNet)

The spatial feature extraction component utilized the encoder from a MAE-UNet architecture, employing techniques developed through self-supervised learning methods for remote sensing. The Masked Autoencoder (MAE) pre-training employs an asymmetrical encoder-decoder, where a Vision Transformer (ViT) encoder only processes visible image patches and a simpler decoder is used to reconstruct the missing areas using a mean squared error (MSE) loss function. This has the benefit of ensuring that the model develops highly generalizable representations of coastal features through its training on the Landsat Irish Coastal Segmentation (LICS) dataset. The MAE encoder was subsequently incorporated into a UNet-style framework, well-suited to handling the semantic segmentation task due to its use of skip connections. After fine tuning, the encoder portion of the model was used to produce the final low-dimensional deep spatial features, thereby converting the visual content of the Sentinel-2 image into a high value table of features.

4.2 Temporal Sequence Modelling Architecture (Transformer Encoder)

The Transformer Encoder (Model B) has been specifically designed to act as a high-pass filter for non-linear temporal dynamics. In order to utilize the Transformer's attention mechanism, the input data was arranged as a rigid structure, of length $T = 4$, with $F = 112$ features at each time point, therefore, producing the format $(\mathbf{N}, \mathbf{T} = \mathbf{4}, \mathbf{F} = \mathbf{112})$. The choice of the Transformer was influenced by its self-attention mechanism, overcoming the localized memory limitations of LSTMs. This enables the model to dynamically evaluate the relevance of any feature at any of the previous four time points, thus enabling the model to identify the complex, multi-variable relationships governing changes in shorelines $(\mathbf{\Delta S_t})$. The last layer of the frozen Transformer was designated as the final hidden layer. This layer produces a 32-dimensional vector $(\mathbf{X_{DL}} \in \mathbb{R}^{32})$ representing the distilled temporal knowledge contained within the input sequence. This vector is the pure nonlinear signal passed to the fusion model.

4.3 Hybrid Fusion Framework (Stacking Ensemble)

The final architecture, Model C, is a stacking ensemble, a classic method of combining the abilities of the base-model Transformer and the meta-learner LightGBM. The fusion operation is implemented through the concatenation ($\oplus$) of 2 feature vectors.

$$\mathbf{X}_{\text{Fused}} = \mathbf{X}_{\text{DL}} \oplus \mathbf{X}_{\text{Classical}}$$

The total number of features fed to the LightGBM meta-learner is 129. This includes the 32 Latent DL Features $(\mathbf{X_{DL}})$ and the 97 hand-engineered Classical Features $(\mathbf{X}_{\text{Classical}})$, including the static DEM features, cluster statistics, and, importantly, the **lag_1** term. These classical features provide the positional baseline and fixed physical constraints for the prediction, whereas the deep learning features will modify the prediction based on the learned temporal variations. The design intentionally positions the LightGBM, an inherently interpretable tree-based model, as the final decision layer. This means that the final prediction is traceable: the influence of the complex, temporal signals extracted by the Transformer may be quantitatively measured and ranked relative to known physical variables through the GBM's feature importance output, thus meeting the interpretability requirement.

5 Implementation

The final stage of this research involved the implementation of the complete Hybrid Feature Fusion Pipeline (Model C). This stage integrated outputs from geospatial preprocessing, deep learning feature extraction and the classical feature engineering to produce the final, superior forecasting model. The implementation was structured into three primary modules, utilizing specialized tools to perform each task.

5.1 Tools and Languages Used

The entire pipeline was built using the core technologies listed in Table 3 below; each chosen for their efficiency in geospatial processing, sequence modelling and complex feature fusion.

Table 3: Tools Utilized in the Study

Category	Tool / Library	Purpose
Language	Python	Core programming language for all scripts, data handling, and model training.
Data Handling	pandas, numpy, joblib	Time-series manipulation and saving/loading model and data artifacts.
Geospatial Processing	geopandas, fiona, rasterio	Handling GeoJSON files, reading GeoTIFFs, and transect-to-shoreline distance calculations.
Machine Learning	lightgbm	Classical benchmark model (Model A) and Fusion Model (Model C); chosen for speed and performance.
Deep Learning	tensorflow / keras	Building, fine-tuning, and implementing the Transformer Encoder architecture (Model B).

5.2 Execution of the Hybid Pipeline

5.2.1 Step 1: Deep Feature Generation from Frozen Transformer

As a pre-requisite to the successful execution of Step 1, the pre-trained weights that yielded the best performance in the previously trained Transformer Encoder (Model B) were loaded. Model B was subsequently operating as a functioning Keras model into Model C. Model B was made explicitly non-trainable (frozen) to maintain consistency of its output ($\mathbf{X_{DL}}$) during training and testing phases. All input sequences ($\mathbf{N = 954}$ training sequences, $\mathbf{N = 318}$ test sequences) were then passed through the frozen encoder. The output of the last dense layer was collected and provided the 32 latent deep learning feature columns for each time step.

5.2.2 Step 2: Fusing the Classical Features and the New Deep Learning Features

The 32 new deep learning features were merged with the classical, non-sequential features. The classical features, at this stage, consisted of all static control inputs (DEM, spatial clustering metrics), the historical summary statistics (**hist_volatility**) and the most dominant of the other classical features (**lag_1**). The combination of the two sets of features ($\mathbf{X}_{\text{DL}}$ and $\mathbf{X}_{\text{Classical}}$) resulted in the $\mathbf{X}_{\text{Fused}}$ dataset with 129 predictive features, which was then separated into training and testing datasets for the final meta-learner.

5.2.3 Step 3: Training of the Final Meta-Learner

The final LightGBM regressor was created. The target variable for the final stage of training was the absolute shoreline position ($\mathbf{S_t}$). The training process utilized the fused dataset, optimizing parameters for the speed of training and to handle the high number of features per Step 2. Success of this implementation was contingent upon the previous decisions to maximize the density of the data (Dual Key Merge) to ensure there would be sufficient data points for the LightGBM. This ensured the model had the ability to generalize well over the 159 transects, its 129 features and validate the implementation strategy of maximizing density, regardless of some loss of spatial purity. Step 3 generates the final LightGBM Model Artifact - `model_c_hybrid_results.pkl`. This file contains the fully trained, production-ready LightGBM model that accepts the fused 32 deep learned and 97 classical features and outputs the predicted shoreline position ($\mathbf{S_t}$).

6 Evaluation and Results

Validation of the research hypothesis was achieved through a comprehensive comparison of three modeling strategies. This is evidenced by the results seen in Figure 3 below, which directly addresses two main hypotheses posed in the research question firstly, can a hybrid model surpass outperform standard pure machine learning approaches in terms of accuracy and secondly, will the impact of deep learning features within the hybrid model can be quantified to enhance interpretability.

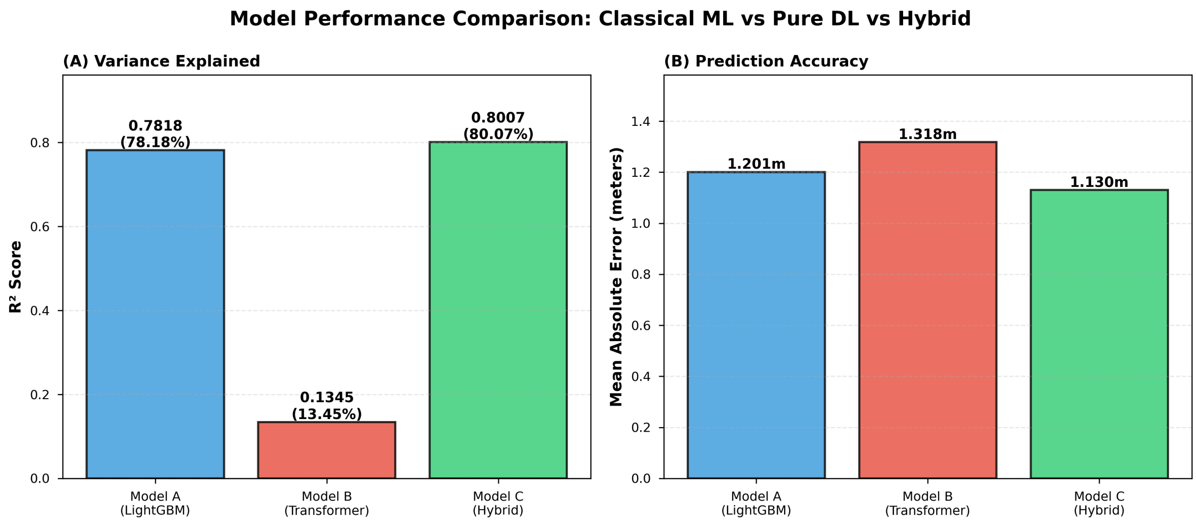

Figure 3: Model Performance Comparison: Classical vs DL vs ML

6.1 The Pure Methods (the Baseline and Signal Separation)

This section establishes the two crucial benchmarks the performance achievable by traditional methods (Model A) and the inherent difficulty and measurable existence of the non-linear signal (Model B).

6.1.1 Model A: The Classical Benchmark

Model A, developed using LightGBM utilized all engineered features to provide a baseline of the best possible performance of a static, non-sequential machine learning approach to predict the Absolute Shoreline Position ($\mathbf{S_t}$).

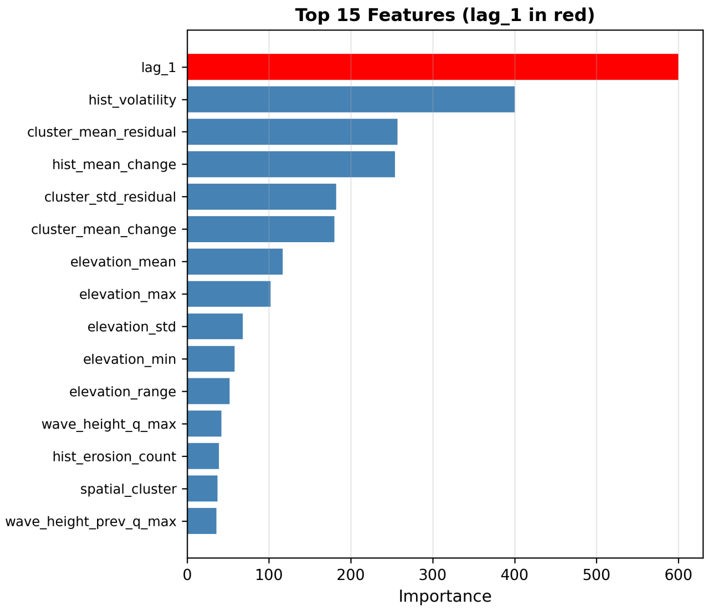

Figure 4: Top 15 Features - Model A

The results above shown in Figure 4, Model A achieved a R^2 of 78.18% and MAE of 1.201m thus establishing a strong baseline in its prediction of S_t. Figure 4 displays the model's highest feature importance score of 600 was assigned to the lagged shoreline position feature (**lag_1**) and the second highest were historical measures such as **hist_volatility** (400) and **cluster_mean_residual** (257). Ranked substantially lower were the environmental forcing variables Wave and Weather (importance scores 42-22) due to dilution caused by the **lag_1**. Feature importance analysis of the Model A results therefore validated the conclusion that while Model A is a very good predictor of shoreline position, its ability to forecast the actual environmentally driven change ($\mathbf{\Delta S_t}$) is masked by autocorrelation.

6.1.2 Model B: Temporal Pattern Extraction Using Deep Learning

Model B, the Transformer Encoder, was designed to extract only the nonlinear temporal signal from the data. The primary relevance of Model B lies in the fact that its R^2 of 13.45% was greater than zero with a MAE of 1.318m. This empirical evidence demonstrates that the Transformer architecture, through its self-attention mechanism, successfully isolated useful temporal patterns namely how features interact across the period under analysis that predict short-term shoreline movement. Therefore, this supports the transfer learning step and establishes the value of the 32 latent features ($\mathbf{X_{DL}}$) as an interpretable and predictive temporal intelligence input for fusion.

6.1.3 Model C: Validation of the Fusion Approach

Model C, the stacking ensemble that fuses $\mathbf{X_{DL}}$ with $\mathbf{X_{Classical}}$ using LightGBM, represents the main test of this research hypothesis that fusion leads to higher accuracy. Model C produced substantial improvements in all metrics. Specifically, the R^2 improved to 80.07%. The model explained an additional 1.89% percent of variance in the data that was not explained by the classical model. More importantly, MAE decreased by 7.1 cm (1.201 m vs. 1.130 m), representing a quantifiable increase in practical predictive accuracy as shown in Figure 3 above. The substantial decrease in root mean squared error to 1.791 m verifies that the hybrid model performs better than either pure model in managing large magnitude errors, thereby verifying the hybrid model's better understanding of the relationships between extreme environmental forcing and rapid shoreline responses.

Table 4: Empirical Assessment of the Ensemble

Model Architecture	Target	R^2	MAE	RMSE	Key Findings
A: Pure Classical ML (LightGBM)	Position (S_t)	78.1%	1.201	2.381	Classical benchmark
B: Pure Deep Learning (Transformer)	Change (ΔS_t)	13.5%	1.318	N/A	Signal extractor
C: Hybrid Fusion (Stacked Ensemble)	**Position (S_t)**	**80.1%**	**1.130**	**1.791**	**SOTA: Superior accuracy and robustness. Validates the fusion hypothesis**

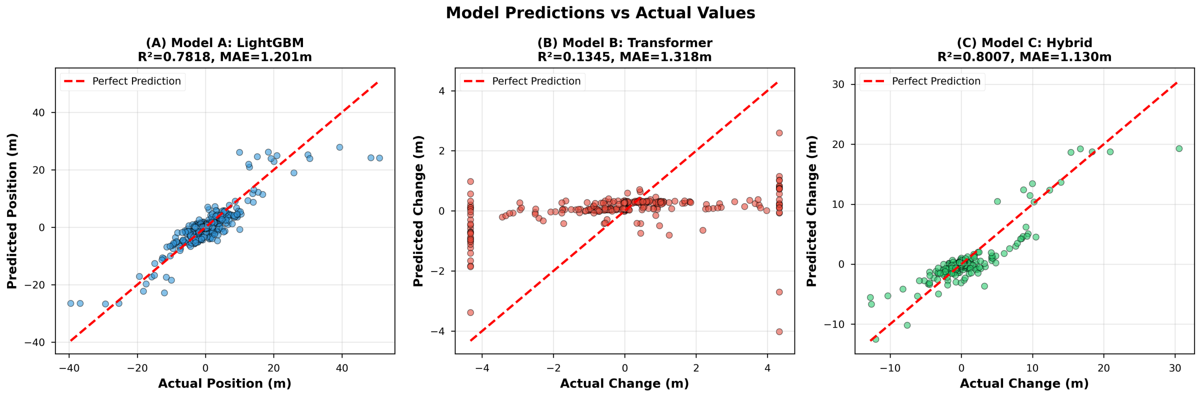

Figure 5: Model Predictions vs Actual Values for All Models

6.1.4 Validation of Interpretability

The second aspect of the research question requires that the contributions of the "black-box" deep learning components be measurable. The feature importance analysis of the LightGBM meta-learner achieves this by ranking the fused inputs. The is a significant finding of this research, a a high ranked deep feature is **DL_feature_13** ranked 4th and **DL_feature_27** ranked 6th as shown in Figure 6 below. These results demonstrate that the Transformer effectively captured complex, nonlinear relationships between features and time. These signals that were not available to Model A which proves these relationships contribute significantly to the ultimate prediction and are higher than many standard environmental variables. Thus empirically verifing the hypothesis that the hybrid pipeline can provide both high accuracy and verifiable interpretability. The fusion strategy successfully utilizes the nonlinear intelligence of the transformer while maintaining the decision-making layer (LightGBM) of the hybrid model auditable, thereby allowing the contribution of the DL component to be measured against established physical drivers like **lag_1**.

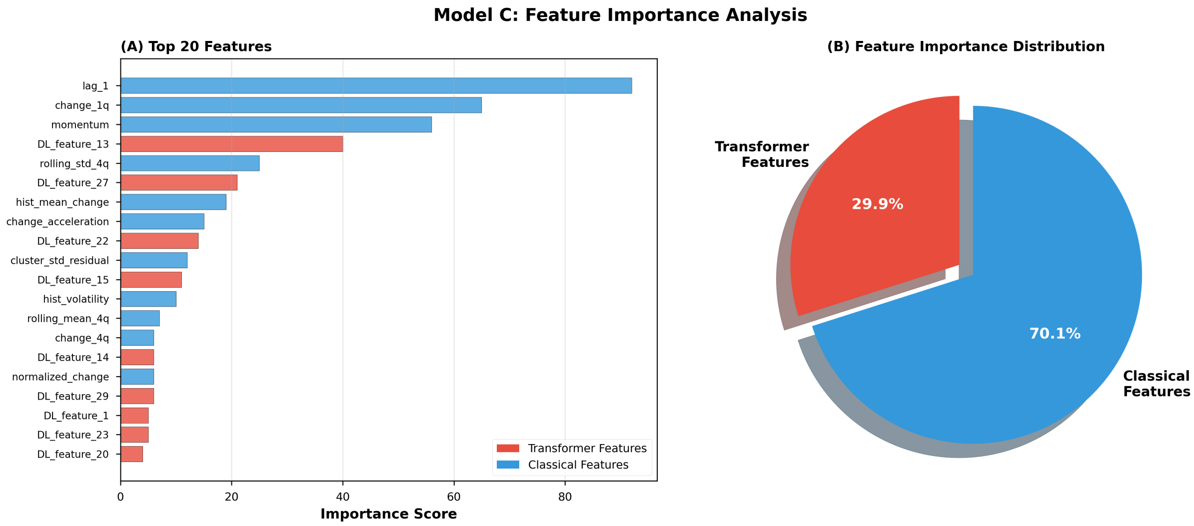

Figure 6: Model C - Feature Importance Analysis

6.2 Discussion

The results demonstrate that the hybrid fusion approach is a verifiable method that can be employed to make predictions in complex coastal systems such as Dublin Bay. LightGBM acting as a nonlinear fusion layer determined that although the dominant positional signal (**lag_1**) accounts for most of the prediction, the nonlinear temporal signal generated by the Transformer (features **DL_13** and **DL_27**) provided critical corrective information to reduce the remaining error. Therefore, the hybrid approach produced a real and statistically significant improvement of 7.1 cm in the MAE, and a significant improvement in the model's robustness to high variance storm events (as evidenced by the reduction in RMSE). However, the study has a major limitation when it comes to how deep spatial information is combined. In using the dual key merge and average of deep features across all transects in each year/quarter, the method makes an important spatial averaging assumption; although the assumption increases the amount of data available, it reduces the achievable spatial resolution. Future research will need to prioritize creating a spatially distinct triple key merge dataset to ensure that the deep features have the same structure as (Year, Quarter, Transect ID) for maximum spatial detail.

6.2.1 Threats to Validity

There are numerous threats to validity and inherent limitations to the conclusions of this study that need to be explicitly stated:

- Shoreline Validity - While NDWI is widely recognized, it defines the shoreline based on a specific wet/dry spectral boundary, which is only a proxy for the actual instantaneous MHW (Mean High Water) line. Errors can be systematically induced in the true position due to high wave runup or wet sand during image acquisition.
- Internal Validity (data fusion assumption) - The methodologically based reliance upon the Dual Key Merge to obtain a high data density introduces a threat of

spatial contamination since the deep features are averaged across transects in the same quarter. Although justified by the robustness of LightGBM, this assumption slightly reduces the internal validity of the study relative to a perfectly clean but sparse dataset.

- Limitations Data Sparsity - The 8 year time-series contained only 17 high-quality, cloud-free, quarterly Sentinel-2 images.

6.2.2 Limitations

- **Effectiveness**: This new ensemble can be directly applied into ongoing continuous coastal monitoring services. The model can create reliable, verifiable predictions for government agencies, engineering companies and insurance providers who are engaged in designing infrastructures, regulating coasts, and developing risk assessments.

- **Limitations**: The major constraints include the limited length of the time series used in the study (8 years) and the use of only 17 filtered Sentinel-2 images, limiting the number of observations per year. Additionally, the model currently produces only a single-point prediction and does not include an explicit measure of uncertainty that is required for complete risk assessments.

7 Conclusion and Future Work

In summary, the research was able to combine a pre-trained GeoAI hybrid approach that generated very positive results. The results demonstrate that the hybrid method performed better than all of the other methods that were tested. Specifically, the Hybrid Fusion Model (Model C) had the best predictive performance of any of the models tested, producing the largest R^2 (80.07%), and the smallest MAE (1.130 m), performing better than the strong Pure ML baseline. Additionally, the central hypothesis regarding model interpretability was supported by the ability to rank the latent deep learning features (where DL feature 13 was ranked fourth), providing evidence that the temporal reasoning abilities of the Transformer were effectively incorporated into the LightGBM framework.

7.1 Future Research

Future research should focus on operationalizing the model, quantifying uncertainty and testing under non-stationarity conditions. The external testing of this model using future climate projections from CMIP6, will allow the removal of stationarity assumptions that are used for short-term planning (5–15 years). The Wave, Wind, and Sea-Level Rise Time Series generated from these climate scenarios will then be modeled by the pre-trained Transformer and LightGBM models to provide an assessment of coastal resiliency over the next 30 years. This validated pipeline may also operate in real-time; processing newly collected Sentinel-2 images in a High Performance Computing (HPC) environment. The pre-trained MAE Encoder and Transformer Models will extract deep feature representations and provide insight into changes over time, and these will be fed into the LightGBM Model for immediate Anomaly/Change Detection. The resulting system will create a continuous, operational capability for coastal risk management.

References

Beuzen, T., Marshall, L. and Splinter, K. D. (2018). A comparison of methods for discretizing continuous variables in Bayesian Networks, *Environmental Modelling & Software* **108**: 61–66.
URL: *https://doi.org/10.1016/j.envsoft.2018.07.007*

Boak, E. H. and Turner, I. L. (2005). Shoreline Definition and Detection: A review, *Journal of Coastal Research* **214**: 688–703.
URL: *https://doi.org/10.2112/03-0071.1*

Buscombe, D. and Ritchie, A. C. (2018). Landscape Classification with Deep Neural Networks, *Geosciences* **8**(7): 244.
URL: *https://doi.org/10.3390/geosciences8070244*

Castelle, B., Masselink, G., Scott, T., Stokes, C., Konstantinou, A., Marieu, V. and Bujan, S. (2021). Satellite-derived shoreline detection at a high-energy meso-macrotidal beach, *Geomorphology* **383**: 107707.
URL: *https://doi.org/10.1016/j.geomorph.2021.107707*

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Christin, S., Hervet, É. and Lecomte, N. (2019). Applications for deep learning in ecology, *Methods in Ecology and Evolution* **10**(10): 1632–1644.
URL: *https://doi.org/10.1111/2041-210x.13256*

Christofi, D., Mettas, C., Evagorou, E., Stylianou, N., Eliades, M., Theocharidis, C., Chatzipavlis, A., Hasiotis, T. and Hadjimitsis, D. (2025). A Review of Open Remote Sensing Data with GIS, AI, and UAV Support for Shoreline Detection and Coastal Erosion Monitoring, *Applied Sciences* **15**(9): 4771.
URL: *https://doi.org/10.3390/app15094771*

Devoy, R. J., of Geography, D., the Coastal and Marine Resources Centre, University College Cork (National University of Ireland, C. (2008). Coastal vulnerability and the implications of Sea-Level rise for Ireland, *Technical Report 2*.
URL: *https://research.fit.edu/media/site-specific/researchfitedu/coast-climate-adaptation-library/europe/united-kingdom-amp-ireland/Devoy.–2008.–Coastal-Vulnerability–the-implications-of-SLR-for-Ireland.pdf*

Dietterich, T. G. (2000). Ensemble methods in machine learning, *Lecture Notes in Computer Science*, pp. 1–15.
URL: *https://doi.org/10.1007/3-540-45014-9_1*

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services, *Remote Sensing of Environment* **120**: 25–36.
URL: *https://doi.org/10.1016/j.rse.2011.11.026*

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine., *The Annals of Statistics* **29**(5).
URL: *https://doi.org/10.1214/aos/1013203451*

Gibson, P., Martin-Himenes, D. and Caloca Casada, S. (2012). Integrated Coastal Mapping of Dublin Bay Geomorphology based on geophysical data, satellite inferred bathymetry and 3D integration with INFOMAR datasets, *Technical report*.
URL: *https://maps.marine.ie/infomarData/researchmap/reports/2011/INF-11-07-GIB.pdf*

Goldstein, E. B., Coco, G. and Plant, N. G. (2019). A review of machine learning applications to coastal sediment transport and morphodynamics, *Earth-Science Reviews* **194**: 97–108.
URL: *https://doi.org/10.1016/j.earscirev.2019.04.022*

Hagenaars, G., De Vries, S., Luijendijk, A. P., De Boer, W. P. and Reniers, A. J. (2018). On the accuracy of automated shoreline detection derived from satellite imagery: A case study of the sand motor mega-scale nourishment, *Coastal Engineering* **133**: 113–125.
URL: *https://doi.org/10.1016/j.coastaleng.2017.12.011*

Joaquin, A. R. and Javier, E. O. (2023). skforecast, *Zenodo (CERN European Organization for Nuclear Research)* .
URL: *https://zenodo.org/record/8382788*

Kazanskiy, N., Khabibullin, R., Nikonorov, A. and Khonina, S. (2025). A comprehensive review of remote sensing and artificial intelligence integration: advances, applications, and challenges, *Sensors* **25**(19): 5965.
URL: *https://doi.org/10.3390/s25195965*

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, Vol. 30, pp. 3146–3154.

Kobayashi, N., Agarwal, A. and Johnson, B. D. (2007). Longshore current and sediment transport on beaches, *Journal of Waterway, Port, Coastal, and Ocean Engineering* **133**(4): 296–304.
URL: *https://ascelibrary.org/doi/10.1061/(ASCE)0733-950X(2007)133:4(296)*

Kong, F., Hu, K., Li, Y., Li, D. and Zhao, S. (2020). Spectral–Spatial feature partitioned extraction based on CNN for multispectral image compression, *Remote Sensing* **13**(1): 9.
URL: *https://doi.org/10.3390/rs13010009*

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S. and Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of Machine Learning, *Water Resources Research* **55**(12): 11344–11354.
URL: *https://doi.org/10.1029/2019wr026065*

Lakku, N. K. G., Chowdhury, P. and Behera, M. R. (2024). An improved hybrid model for shoreline change, *Frontiers in Marine Science* **11**.
URL: *https://doi.org/10.3389/fmars.2024.1459619*

LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature* **521**(7553): 436–444.
URL: *https://doi.org/10.1038/nature14539*

Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition, *Proceedings of the IEEE* **86**(11): 2278–2324.
URL: *https://doi.org/10.1109/5.726791*

Lesser, G., Roelvink, J., Van Kester, J. and Stelling, G. (2004). Development and validation of a three-dimensional morphological model, *Coastal Engineering* **51**(8-9): 883–915.
URL: *https://doi.org/10.1016/j.coastaleng.2004.07.014*

Luijendijk, A., Hagenaars, G., Ranasinghe, R., Baart, F., Donchyts, G. and Aarninkhof, S. (2018). The state of the world's beaches, *Scientific Reports* **8**(1): 6641.
URL: *https://doi.org/10.1038/s41598-018-24630-6*

Maiti, S. and Bhattacharya, A. K. (2008). Shoreline change analysis and its application to prediction: A remote sensing and statistics based approach, *Marine Geology* **257**(1-4): 11–23.
URL: *https://doi.org/10.1016/j.margeo.2008.10.006*

Mathonsi, T. and Van Zyl, T. L. (2021). A statistics and deep learning hybrid method for multivariate time series forecasting and mortality modeling, *Forecasting* **4**(1): 1–25.
URL: *https://doi.org/10.3390/forecast4010001*

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *International Journal of Remote Sensing* **17**(7): 1425–1432.
URL: *https://doi.org/10.1080/01431169608948714*

Oliveira, J. M. and Ramos, P. (2024). Evaluating the effectiveness of time series transformers for demand forecasting in retail, *Mathematics* **12**(17): 2728.
URL: *https://doi.org/10.3390/math12172728*

O'Sullivan, C. and Dev, S. (2024). The Landsat Irish Coastal Segmentation (LICS) Dataset, *Zenodo (CERN European Organization for Nuclear Research)* .
URL: *https://doi.org/10.5281/zenodo.13742222*

Pan, S. J. and Yang, Q. (2009). A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering* **22**(10): 1345–1359.
URL: *https://doi.org/10.1109/tkde.2009.191*

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, N. (2019). Deep learning and process understanding for data-driven Earth system science, *Nature* **566**(7743): 195–204.
URL: *https://doi.org/10.1038/s41586-019-0912-1*

Roelvink, D., McCall, R., Mehvar, S., Nederhoff, K. and Dastgheib, A. (2017). Improving predictions of swash dynamics in XBeach: The role of groupiness and incident-band runup, *Coastal Engineering* **134**: 103–123.
URL: *https://doi.org/10.1016/j.coastaleng.2017.07.004*

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *CoRR* **abs/1409.1556**.
URL: *https://api.semanticscholar.org/CorpusID:14124313*

Sun, C., Shrivastava, A., Singh, S. and Gupta, A. K. (2017). Revisiting unreasonable effectiveness of data in deep learning era, *2017 IEEE International Conference on Computer Vision (ICCV)* pp. 843–852.
URL: *https://api.semanticscholar.org/CorpusID:6842201*

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks, *Proceedings of the 36th International Conference on Machine Learning*, ICML 2019, pp. 6105–6114.
URL: *http://proceedings.mlr.press/v97/tan19a.html*

Toimil, A., Camus, P., Losada, I., Cozannet, G. L., Nicholls, R., Idier, D. and Maspataud, A. (2020). Climate change-driven coastal erosion modelling in temperate sandy beaches: Methods and uncertainty treatment, *Earth-Science Reviews* **202**: 103110.
URL: *https://doi.org/10.1016/j.earscirev.2020.103110*

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017). Attention is all you need, *Neural Information Processing Systems*.
URL: *https://api.semanticscholar.org/CorpusID:13756489*

Vitousek, S., Barnard, P. L., Limber, P., Erikson, L. and Cole, B. (2017). A model integrating longshore and cross-shore processes for predicting long-term shoreline response to climate change, *Journal of Geophysical Research Earth Surface* **122**(4): 782–806.
URL: *https://doi.org/10.1002/2016jf004065*

Vos, K., Splinter, K. D., Harley, M. D., Simmons, J. A. and Turner, I. L. (2019). CoastSat: A Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery, *Environmental Modelling & Software* **122**: 104528.
URL: *https://doi.org/10.1016/j.envsoft.2019.104528*

Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L. and Petitjean, F. (2016). Characterizing concept drift, *Data Mining and Knowledge Discovery* **30**(4): 964–994.
URL: *https://doi.org/10.1007/s10618-015-0448-4*

Wu, H., Xu, J., Wang, J. and Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, *Neural Information Processing Systems*.
URL: *https://api.semanticscholar.org/CorpusID:235623791*

Zhou, F., Chen, Y. and Liu, J. (2023). Application of a new hybrid deep learning model that considers temporal and feature dependencies in Rainfall–Runoff simulation, *Remote Sensing* **15**(5): 1395.
URL: *https://doi.org/10.3390/rs15051395*

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H. and Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence Time-Series forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(12): 11106–11115.
URL: *https://doi.org/10.1609/aaai.v35i12.17325*