1 Appendix – WP1A Characterisation – Advanced Data Analytics for Detection of Marine Mammals and Seabirds from Large Datasets

1.1 Overview and Specification

1.1.1 Problem Definition

Passive acoustic monitoring (PAM) is a common method for tracking marine mammal populations, particularly species like dolphins, porpoises, and whales that rely on sound for navigation, feeding, and communication. Work has focussed on processing raw data, comprised of signals recorded by hydrophones. Using Machine Learning (ML) and deep learning models, the main objective is to identify the species of interest based on the hydrophone locations, to determine presence and abundance of target species.

In the pre-processing phase (click detection), PAMGuard, a software package for audio signal processing to analyse recorded sounds and extract clicks, audio records closely resembling clicks. These clicks may originate from the target species, other species, ship-related noise, shrimp activity, and more.

After the process of detecting and extracting clicks, the next phase involves their classification or labelling. While manual labelling entails expert examination of raw click waveforms and their time and frequency attributes for accurate classification, automated classification employs ML models, often termed intelligent methods, effectively reducing both time and cost associated with the task. The vast datasets produced by PAM can benefit from the application of ML models, specifically deep learning models, to facilitate rapid automated detection and classification of marine mammal vocalizations.

1.1.2 What Has Been Accomplished

The application of a deep learning approach for the supervised classification of dolphins' echolocation clicks has been investigated. Deep learning automatically extracts effective features from the raw waveform (or its spectrogram) without the need for a separate pre-processing step for feature extraction.

The approach required the use of two popular deep neural networks, namely Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). Both models were fed with raw waveform data and spectrograms. The data under examination were recorded in the vicinity of Anglesey's west coast, an environment characterized by high background noise. These data were labelled for three species of interest: broadband species, Risso's dolphins, and Harbour porpoises.

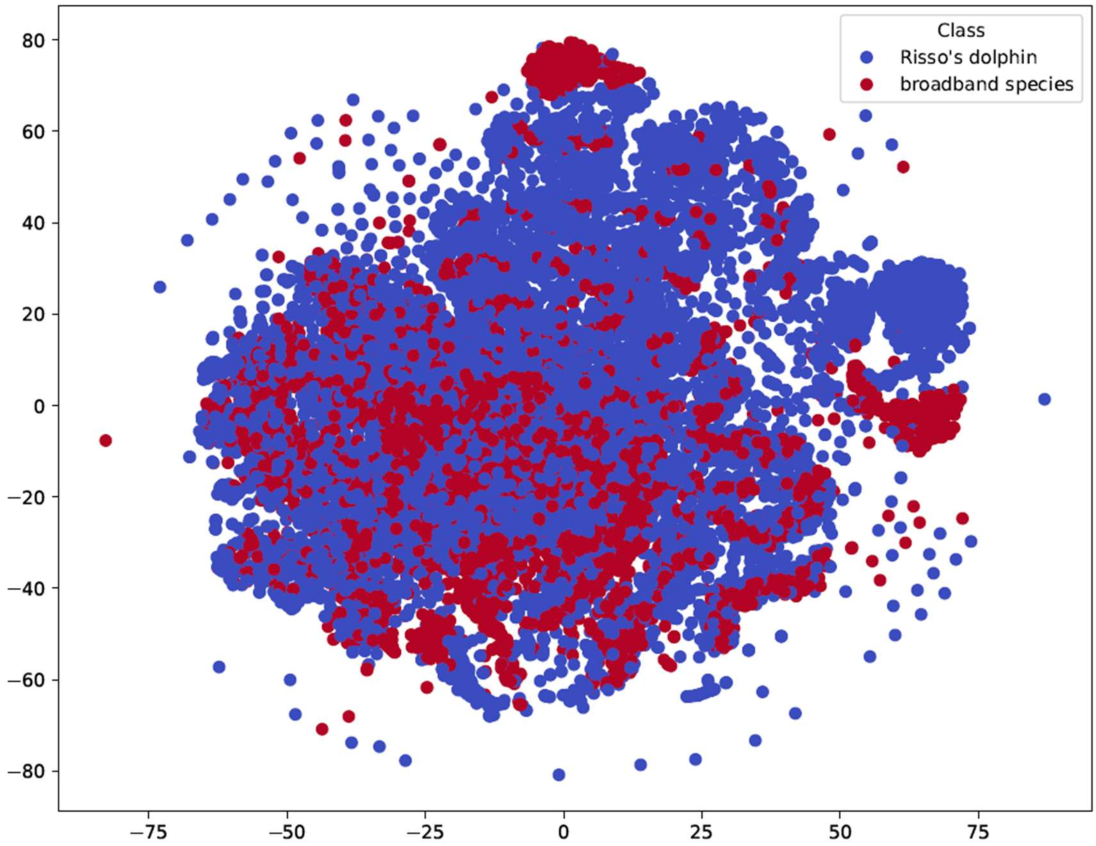
The Figure 1-1 illustrates a t-SNE plot, providing a visual depiction of the classification task's intricacy when distinguishing between broadband species and Risso's dolphins. Remarkably, even after reducing the original high-dimensional signals to a two-dimensional space, the plot reveals the complex interplay among data points. This complexity underscores the challenges associated with accurately segregating these distinct classes.

Of note, the highest performance levels were achieved by the CNN and RNN models, allowing clicks to be classified based on their raw waveforms with accuracy, precision, and recall rates all exceeding 90% (resulting in F1 scores of 95% and 94%, respectively). It's worth mentioning that CNN and RNN models using spectrograms yielded slightly lower results, with CNN's performance dipping by 4% and RNN's by 6%.

While all four proposed structures performed well according to the standard criteria for classification tasks, it has been found that the RNN, when supplied with raw data, offered advantages in terms of reduced processing time.

Of utmost significance is that the model has been developed, with this high level of accuracy, can classify marine species based solely on their individual click signals Concerning the accuracy level of the model in classifying individual clicks, the suggestion is to employ majority voting for identifying the species associated with a "click train". To achieve this objective, there's no need to create a new machine learning model. This can be accomplished solely through the establishment of a confidence threshold and majority voting .

FIGURE 1-1: T-SNE PLOT ILLUSTRATING THE COMPLEXITY OF THE CLASSIFICATION TASK AFTER REDUCING THE ORIGINAL HIGH-DIMENSIONAL SIGNALS (VECTOR LENGTH 512) TO A TWO-DIMENSIONAL SPACE.



1.1.3 Ongoing Work for Data Analytics

Initially, it is necessary to establish two fundamental definitions:

- **Target species clicks:** These clicks specifically pertain to the specimens aimed to be identified.
- **Noise clicks:** These are other clicks extracted from the original signal by the click detector.

This work has demonstrated the capabilities of machine learning and deep learning models in identifying and classifying clicks from species of interest (target species). These models effectively apply what they have learned. The training of the models involved just clicks from the target species., enabling them to distinguish these species. It's essential to note that the models have not been trained on noise clicks, as there is no logical rationale for doing so. In general, it is not feasible to provide these models with a mixture of noise clicks and target species clicks and expect them to accurately identify the target species.

Solution 1:
One potential approach is to submit all detected clicks to the deep learning model for labelling. Since the model has not been trained on noise clicks, it may assign low-confidence labels from the categories it was trained on. Subsequently, by averaging the confidence scores within a click train, it would be possible to assess whether the species of interest is likely present or not. This approach utilizes the model's confidence scores as an indicator of the likelihood of the species of interest being present within the detected clicks.

Solution 2:
If the initial solution proves ineffective, an alternative approach involves implementing a deep learning-based one-class classifier. This method entails training a classifier exclusively on data associated with a single species against whatever. For each specific target category, a one-class classifier is designed. Once again, confidence levels can be used to perform reliable labelling based on this approach.

1.1.4 Data Analytics Conclusion

The proposed mechanism offers substantial advantages in two key aspects:

- **Automated Species Identification:** The entire process of species identification can be carried out automatically, resulting in significant cost and time savings.
- **Deterrent System Integration:** This system holds the potential for future integration into active deterrent systems. These systems could recognize in real-time the presence of the target species if it approaches a danger zone, triggering the activation of the deterrent system.

1.2 Introduction

The growth of tidal-stream energy is picking up speed, with an increasing number of underwater turbines being deployed in areas with high tidal flow. It is crucial to comprehend how marine mammals respond to these operational devices to mitigate uncertainties related to the risk of underwater collisions, posing a key challenge in the consenting process. Therefore, there is an imperative need to comprehend and quantify the risks associated with tidal-stream devices, specifically investigating the movement of cetaceans in proximity to operating infrastructure.

In 2019, a PAM system was devised to monitor cetacean movement by localising vocalisations, specifically echolocation clicks, around the DG500—a full-scale tidal kite developed by Minesto Ltd—in the Holyhead Deep (Veneruso et al., 2020). While there is a lack of localization data or studies on cetacean response during kite operation to date, the project uncovered regular encounters with dolphins and harbour porpoises, demonstrating that cetaceans consistently inhabited the vicinity of the DG500 infrastructure (Veneruso et al., 2020).

The volume of data amassed through the PAM system is substantial and currently necessitates manual examination by an expert investigator to detect and classify cetaceans at the species level. This classification process is labour-intensive and time-consuming. Various techniques with mixed outcomes have been developed for the automatic detection and classification of signals from different cetacean species (reviewed by Usman et al., 2020). However, no single technique can effectively identify the vocalizations of the > 80 known cetacean species, each characterized by distinct physical properties and a range of sounds.

Echolocation clicks, which are impulsive short pulses with substantial strength lasting for microseconds, pose a challenge for identification due to their diverse physical properties. Dolphins, for example, produce echolocation clicks with similar spectral and temporal features, complicating the task of distinguishing between species.

The primary objective of this project is to develop a tool capable of automatically classifying delphinid click detections to the species level where possible.

1.3 Methodology

1.3.1 Description of PAM System and Deployment

As the data for the Morlais Demonstration Zone (MDZ) was not initially accessible at the project's outset, the investigations have been conducted using the closest available similar data. The deployment of the PAM system in Holyhead Deep, Wales, UK (coordinates 53° 17.796 N, 4° 47.948 W) at a depth of 85 meters during the period of August to September 2019 marked a crucial phase in this study (see Figure 1-2). The site, enriched with Marine Renewable Energy (MRE) infrastructure, introduced a unique challenge due to the anthropogenic noise produced, particularly from the dynamic movements of metallic components such as mooring gear and chains. The acoustic recording setup was meticulously designed, featuring a Sonar Point recorder sourced from Desert Star Systems, USA, connected to four HTI 99-UHF hydrophones (HTI, High Tech Inc, USA), capturing data at an impressive rate of 312,500 samples per second. It is noteworthy that the focus was on a single hydrophone channel, chosen strategically for detailed analyses. The investigation into background noise fluctuations involved the calculation of the RMS (root-mean-squared) sound level for each hour of recorded data, providing valuable insights into the environmental acoustics.

FIGURE 1-2: AVAILABLE PAM DATA CLOSE TO THE MDZ

1.3.2 Click Detection and Data Labelling

The utilization of PAMGuard, a widely recognized software tool (Gillespie et al., 2009; www.pamguard.org) added a robust layer to the data processing capabilities. The click detection mechanism, configured to trigger on transient signals with peak frequencies surpassing 20 kHz and rising 10 dB above the continuous measure of background noise, ensured a comprehensive identification of relevant acoustic events. The subsequent manual classification by an experienced analyst (GV) further refined the dataset, resulting in the meticulous labelling of 16,520 clicks. These were categorized into Risso's dolphin clicks (10,367) and broadband click species events (6,153), relying on the nuanced analysis of click frequency characteristics and waveforms, as outlined by (Palmer et al., 2017) and (Soldevilla et al., 2008).

1.3.3 Feature Extraction

The study employed two distinct approaches to prepare the data for the subsequent deep learning models. The first approach harnessed the innate capability of neural networks to decipher inherent rules directly from the raw dataset, automating the extraction of features crucial for classification. The second approach involved the extraction of click frequencies utilizing a short-time Fourier transform (STFT), presenting the model with spectrograms. The hyperparameter tuning process explored various values to ascertain near-optimal configurations for the STFT window size and the overlap size between adjacent windows.

1.3.4 Machine Learning Classifier

To propel the study into the realm of advanced machine learning, two prominent classification models within the deep learning domain: CNN and RNN were tested. CNNs, with their ability to automatically extract features through convolution filters based on local correlations, were juxtaposed against RNNs, which specialize in processing sequential data by incorporating outputs from past or future inputs. Recognizing the limitations of standard RNNs in handling long-term dependencies, the exploration extended to Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks. These additions allowed delving into the intricate nuances of acoustic signal classification, especially in environments where long-term dependencies play a crucial role.

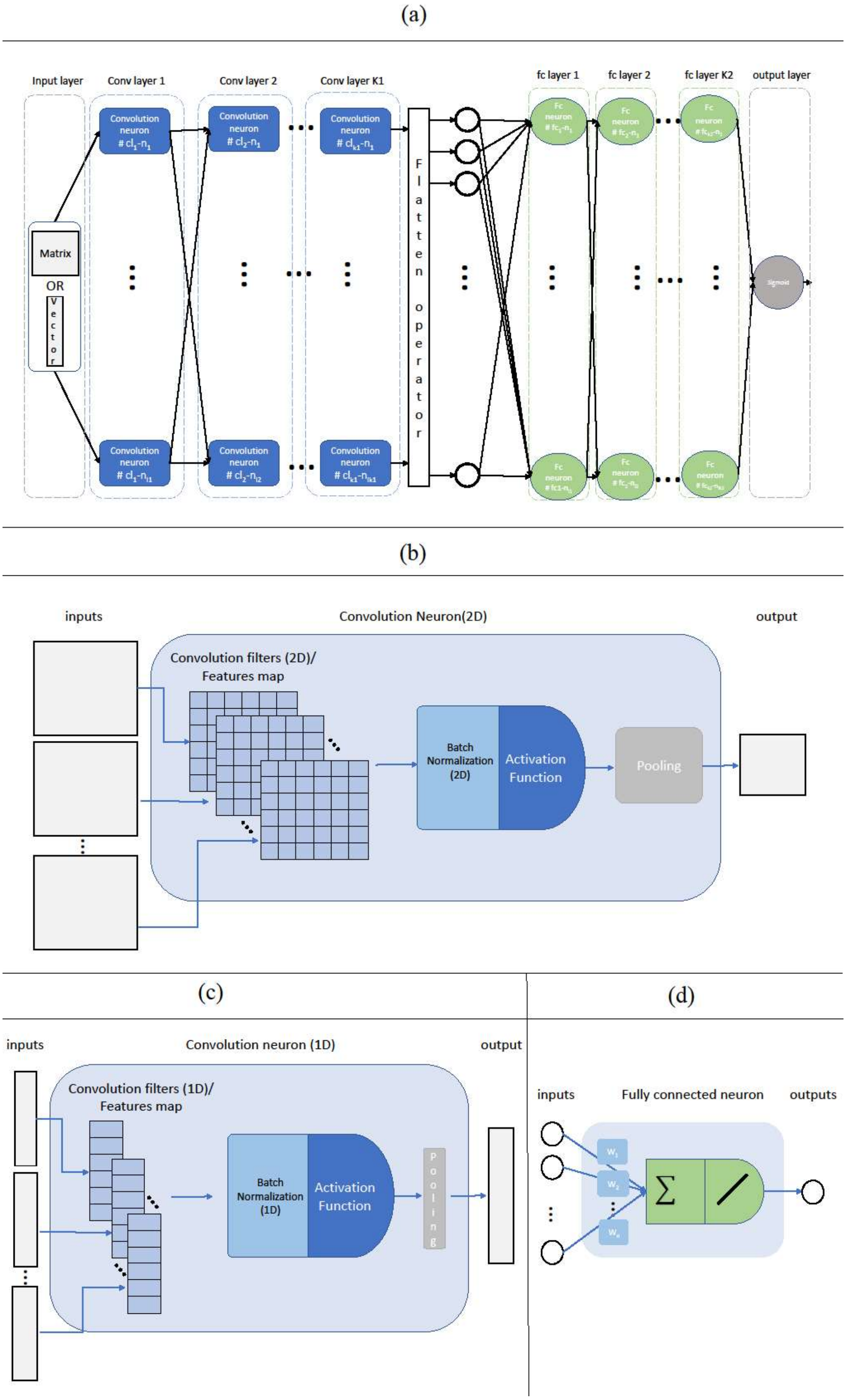
The supervised learning process, as adopted in this study, traversed three critical phases: the training phase, where the model's weights or parameters were honed on labelled data; the validation phase, focusing on hyperparameter tuning; and the test phase, where the model's effectiveness as a classifier was evaluated using unseen labelled data. The meticulous execution of these phases elevated the robustness of the model, ensuring its applicability in real-world scenarios.

1.3.5 Convolution Neural Networks Architecture

The incorporation of CNN into this study introduced a multi-faceted approach to the architecture, depending on the input type. For raw data, one-dimensional convolution filters were applied, while for spectrogram datasets, two-dimensional convolution filters were employed. The architecture, as illustrated in Figure 1-3, comprised several crucial blocks, each contributing to the overall effectiveness of the CNN model:

- The Input layer served as the gateway for the raw signal waveform in one-dimensional convolutional networks or the spectrogram in two-dimensional convolutional networks. The dimensions of the spectrograms were intricately determined by the size of the STFT window and the distance between adjacent windows.
- The Convolution layer, a pivotal component, housed numerous convolution filters (1D or 2D), representing weights or model parameters. The dimensions and stride of these filters, considered hyperparameters, were optimized during the validation process. These convolution filters played a central role in the feature extraction process.
- The Batch Normalization layer contributed to maintaining the mean output close to 0 and the output standard deviation close to 1, ensuring stability in the learning process.
- The Activation function, a nonlinear entity, took the output of the batch normalization layer as input, capturing the inherent nonlinearity of the samples. ReLU and Leaky-ReLU, two widely used activation functions, were implemented in the CNN models.
- The Pooling layer, another crucial element, focused on extracting the most significant features from the convolution layer's output. This layer facilitated a down-sampling mechanism through averaging or maximizing, with the size of the filter and stride as hyperparameters.
- The Fully Connected layers, strategically placed at the end of the network, played a pivotal role in mapping the extracted features to the desired output class probabilities or regression values. These layers performed a linear mapping of the features, culminating in a final non-linear activation function in the last layer to produce the desired output.
- The SoftMax or Logistic layer, the concluding component of the CNN architecture, appended to the end of the fully connected layer. The choice between Logistic and SoftMax classes depended on the nature of the classification task – binary or multi-class. Given the binary classification task, the Logistic class was utilized in the proposed CNN model.

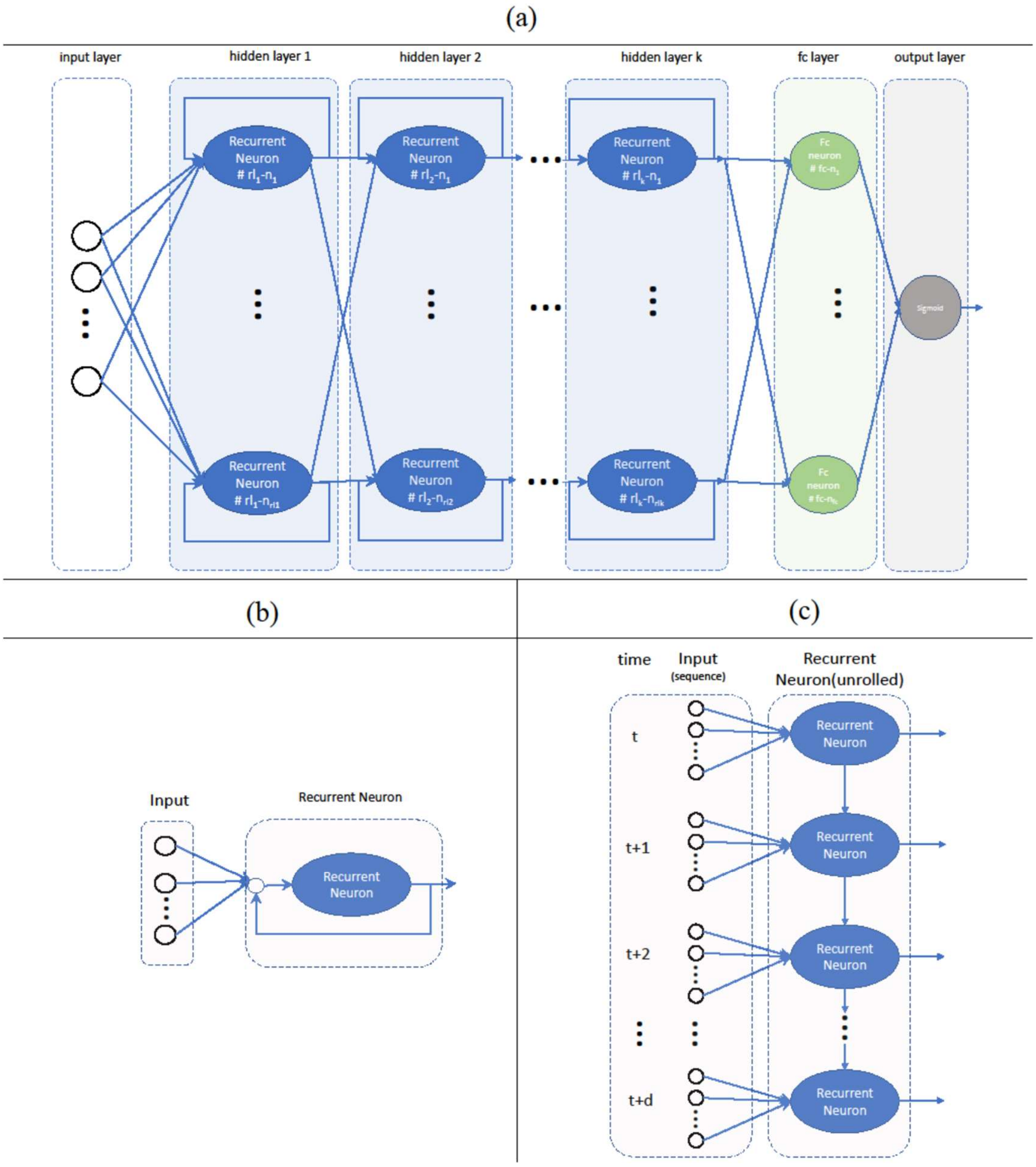
FIGURE 1-3: ARCHITECTURE OF THE CONVOLUTIONAL NEURAL NETWORK: A) GENERAL VIEW OF THE CNN MODEL B) ENLARGED VIEW OF A ONE-DIMENSIONAL CONVOLUTION NEURON C) ENLARGED VIEW OF A TWO-DIMENSIONAL CONVOLUTION NEURON D) ENLARGED VIEW OF A FULLY CONNECTED NEURON.



1.3.6 Recurrent Neural Networks Architecture

The RNN architecture, as depicted in Figure 1-4, introduced a distinctive perspective on processing sequential data. A sample was dissected into a sequence of consecutive states, with these states presented to the first layer of the model. This layer, in turn, encompassed several RNN/GRU/LSTM gates (neurons), arranged sequentially. The standard RNN gate, a fundamental neuron fed with inputs directly from the input and from the output of its preceding state, was enriched by the inclusion of the GRU gate and the more intricate LSTM gate. This diversity added flexibility to the model, allowing for the accommodation of varying historical information.

FIGURE 1-4: ARCHITECTURE OF THE RECURRENT NEURAL NETWORK: A) GENERAL VIEW OF THE RNN B) ENLARGED VIEW OF A RECURRENT NEURON C) ENLARGED VIEW OF AN UNROLLED RECURRENT NEURON.



To augment the flexibility of the model, several recurrent layers could be stacked, each contributing to the extraction of features from the input sequence. At the conclusion of the recurrent layers, a Fully Connected layer and a Logistic layer facilitated the transformation of the extracted features into the target label. This architectural design allowed for the consideration of the history of the input sequence, a critical aspect in extracting meaningful features from sequential data.

In this structured framework, hyperparameters such as the number of gates in each layer, the number of recurrent layers, and the type of gates were carefully considered. The choice of waveform or spectrogram inputs led to distinct methods of transforming a sample into a sequence of states. For waveform inputs, consecutive samples were considered as a state, with the dimension of input (n) determining the length of the sequence. Spectrogram inputs, on the other hand, considered each column of the matrix, representing the magnitude of frequency coefficients in a window of the STFT, as a state. The length of the input vector and the sequence depended on both the size of the window and the distance between two adjacent windows in the STFT.

These intricacies, determined through hyperparameter tuning, played a pivotal role in optimizing the performance of the RNN architecture. The details of these optimizations, as outlined in Table 1-1, showcase the meticulous approach adopted in tailoring the architecture to the specific characteristics of the data.

TABLE 1-1: DESCRIPTION OF THE FOUR MODELS TESTED AND THEIR HYPERPARAMETERS

Model	Category	Parameters	Values
1. CNN, raw signals	Convolution Blocks	Number of CNN blocks	5
		Number of channels	[32,128, 16, 48, 64]
		Convolution filter size	[15, 5, 7, 5, 9]
		Convolution filter stride	[3, 1, 3, 1, 3]
		Activation function	ReLU
		Max-pool filter size*	[1, 2, 2, 1, 2]
		Max-pool stride	[1, 1, 1, 1, 1]
	Linear Blocks	Number of fully connected layers	3
		Number of neurons	[64, 512, 128]
	Others	Drop out probability	0.3
		Learning rate	1e-4
2. CNN, spectrograms	Spectrogram	n-fft	16
		Hop size	12
	Convolution Blocks	Number of CNN blocks	3
		Number of channels	[48, 16, 128]
		Convolution filter size	[3, 3, 3]
		Convolution filter stride	[1, 1, 1]
		Activation function	ReLU
		Max-pool filter size*	[1, 1, 1]
		Max-pool stride	[1, 1, 1]
	Linear Blocks	number of layers	2
		Number of neurons	[512, 64]
	Others	Drop out probability	0.05
		Learning rate	1e-3
3. RNN, raw signals	Neural network	input size	4
		Number of layers	2
		Number of neurons	[33, 27]
		Type of gate	GRU
		Learning rate	0.00098
4. RNN, spectrograms	Spectrogram	n-fft	128
		Hop size	32
	Neural network	Number of layers	3
		Number of neurons	[40, 40, 40]
		Type of gate	LSTM
		Learning rate	0.00078

* for max-pool filter size, number 1 means not using max-pool filter in that block.

In essence, this study extended beyond the conventional boundaries of deep learning, delving into the nuances of CNN and RNN architectures to address the unique challenges posed by the classification of acoustic signals, especially in the context of marine environments.

Overall, the expansion and refinement of the PAM system deployment, click detection, data labelling, feature extraction, and machine learning classifier methodologies in this report provide a comprehensive foundation for future research endeavours in the field of marine mammal monitoring. The intricacies of the CNN and RNN architectures, combined with the detailed exploration of hyperparameter tuning, position this study at the forefront of advancing the understanding of cetacean behaviour in real-world marine environments.

1.4 Results

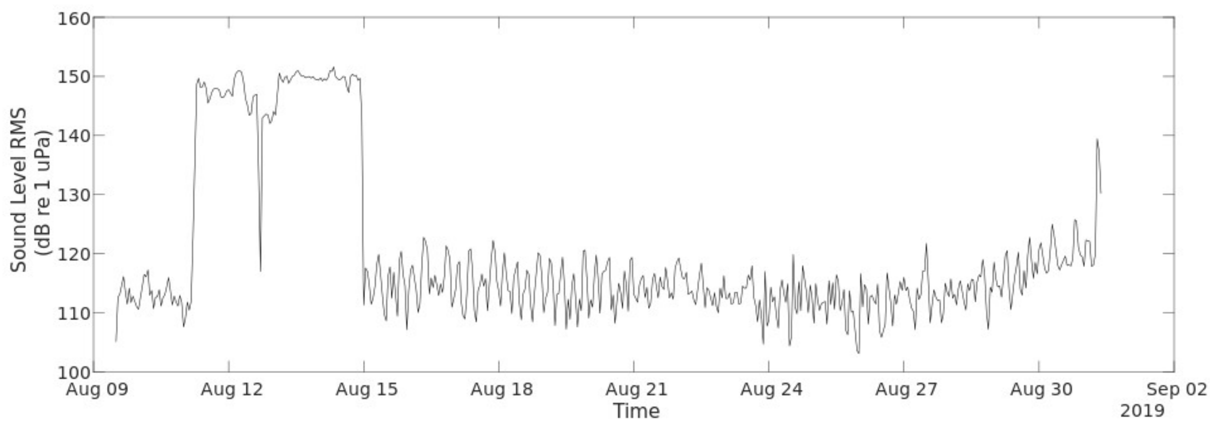
1.4.1 Data processing and performance metrics

As previously detailed, this dataset comprises 16,520 samples, each containing the waveform of a click and its corresponding label. Among these, 10,367 samples are attributed to Risso's dolphins, while 6,153 samples belong to the broadband species. To ensure consistency and facilitate analysis, all waveforms underwent normalization and zero-padding, resulting in a standardized length of 512 points.

The culmination of the rigorous training, validation, and evaluation processes reinforces the reliability and robustness of the deep learning models in addressing the complexities of cetacean click classification. The careful consideration of performance metrics provides a nuanced understanding of the models' strengths and emphasizes their potential applicability in broader marine mammal monitoring initiatives.

Figure 1-5 delves into the background noise levels at the recording site throughout the month of PAM recordings. The plot reveals a continuously changing RMS level, with a notable spike between August 11th and August 15th. This spike corresponds to heightened anthropogenic activities, including shipping and infrastructure movement, emphasizing the impact of human-related noise on the acoustic environment.

FIGURE 1-5: VARYING BROADBAND BACKGROUND NOISE LEVEL (SPLRMS) AT RECORDING SITE DURING THE MONTH OF STUDY



To ensure a robust evaluation of these models, the dataset was partitioned into six folds. One of these folds was set aside for hyperparameter tuning, a critical step in optimizing the model's performance.

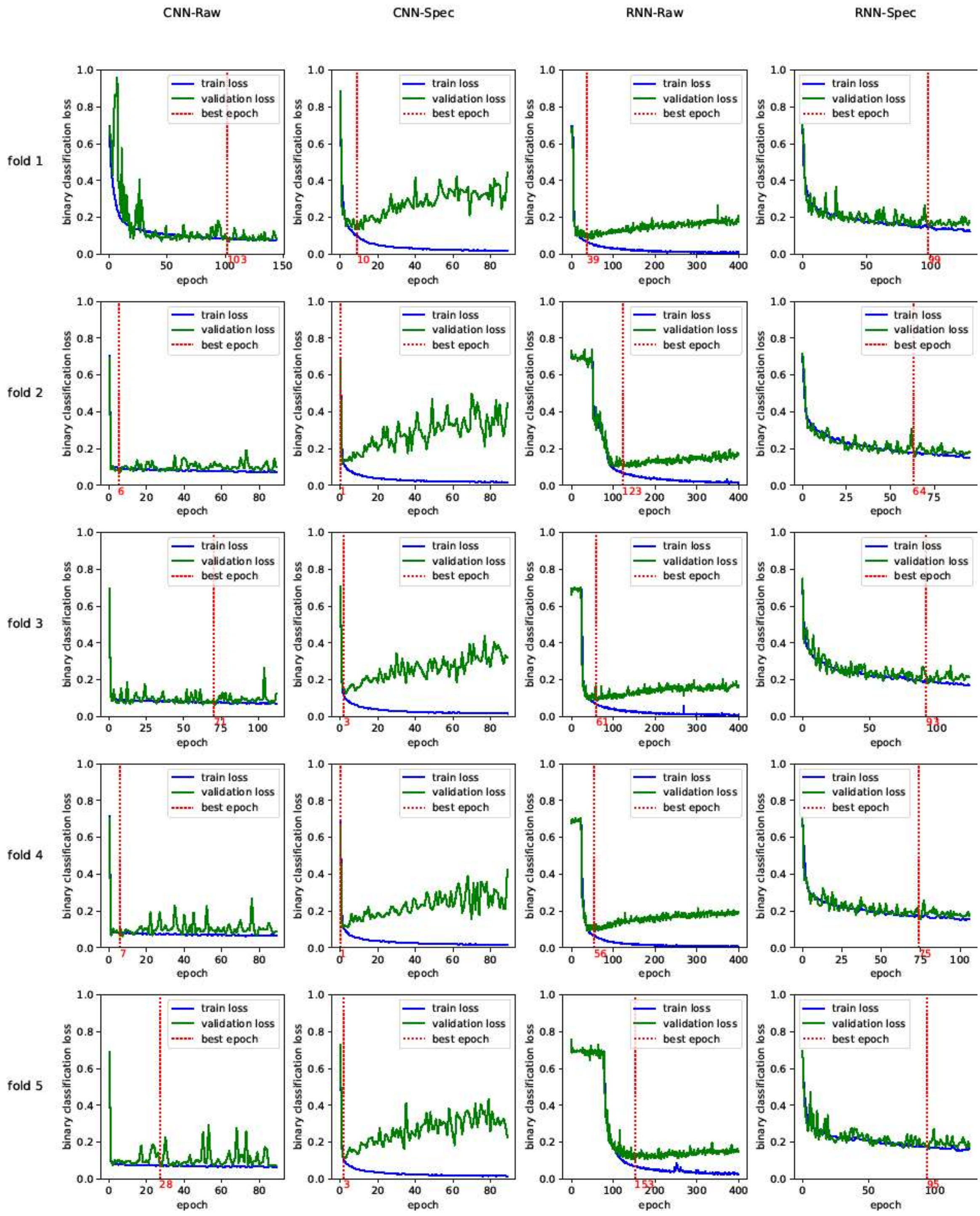
Table 1-1 summarizes the hyperparameters employed in the models, obtained through a Bayesian optimization process and kernel-fitting sampling method. The remaining five folds were utilized in a five-fold-cross-validation structure, allowing for comprehensive training and evaluation of the models. The models underwent independent training five times, utilizing approximately 11,000 samples during each training iteration and evaluation on about 2,750 distinct samples. This meticulous approach ensures a thorough exploration of the dataset and robust assessment of the models' performance under diverse conditions.

1.4.2 Training and Validation Phases

In the pursuit of developing robust and effective deep learning models for cetacean click classification, careful strategies were implemented during the training and validation phases. To mitigate the risk of overfitting, an early stopping condition was incorporated into the training loops, allocating 15% of the training data as a validation set. The training process was designed to halt if the validation loss exhibited continuous increases over 30 epochs. The model with the lowest validation loss was then selected as the best-performing model. Balancing the dataset was a key consideration, achieved by oversampling the broadband species. The binary classification loss function and the Adam optimization algorithm were employed, and the training dataset was divided into batches, each containing 16 samples.

Figure 1-6 provides a comprehensive overview of individual training and validation loss curves for each model across different folds. Notably, the proposed models demonstrated low variance within the five folds, indicative of their robustness. The 'best epoch' points on the graphs, where training could be halted to prevent overfitting, underscore the success of the hyperparameter tuning method in optimizing both the model structure and training algorithm.

FIGURE 1-6: TRAINING AND VALIDATION LOSS CURVE FOR FIVE INDEPENDENT FOLDS FOR ALL MODELS. THE RED DOTTED LINES SHOW THE EPOCH FOR THE BEST MODEL STATE, ACCORDING TO THE VALIDATION DATA LOSS.



While CNN models converged quickly due to the substantial number of parameters, the stability of the models for validation data varied. Specifically, CNN with spectrogram and RNN with raw data exhibited signs of overfitting, suggesting a high degree of freedom in the model. On the other hand, CNN with raw data and RNN with spectrogram demonstrated minimal overtraining and overfitting, indicating the appropriateness of their model structures.

1.4.3 Performance Metrics and Test Phase

The evaluation of model performance involved key metrics that provide insights into their effectiveness. Accuracy, a metric gauging overall model performance across all classes, was calculated as the ratio of correct Risso's dolphin predictions to the total number of click predictions. Precision, a metric assessing the accuracy of classifying a click as Risso's, was determined by the ratio of correctly classified Risso's clicks to the total number of clicks classified as Risso's. Recall, measuring the model's ability to detect Risso's dolphin clicks, was calculated as the ratio of correctly classified Risso's clicks to the total number of Risso's clicks. The F1-score, a harmonic mean between precision and recall, provided a balanced evaluation of the model's performance.

The results, summarized in Table 1-2, showcase the high performance of all models, with metrics consistently exceeding 90%. Notably, both CNN and RNN models fed with raw data achieved precision, accuracy, recall, and F1-scores of approximately 95%. This indicates a high degree of correct positive classifications while minimizing false positives. Despite the class imbalance (prevalence 63%), the high F1 scores (96% and 93% for CNN and RNN, respectively) indicate a commendable balance between precision and recall. These findings underscore the effectiveness of the developed models in accurately classifying Risso's dolphin clicks in real-world marine environments.

The culmination of the rigorous training, validation, and evaluation processes reinforces the reliability and robustness of the deep learning models in addressing the complexities of cetacean click classification. The careful consideration of performance metrics provides a nuanced understanding of the models' strengths and emphasizes their potential applicability in broader marine mammal monitoring initiatives.

TABLE 1-2: PERFORMANCE METRICS (ACCURACY, PRECISION, RECALL, F1 SCORE) STATISTICS (MEAN, MINIMUM, MAXIMUM AND STANDARD DEVIATION) AND NUMBER OF PARAMETERS FOR EACH MODEL

Metric		1) CNN-raw	2) CNN-spectrogram	3) RNN-raw	4) RNN-spectrogram
Accuracy	max	0.98	0.96	0.97	0.94
	mean	0.974	0.954	0.966	0.932
	min	0.97	0.95	0.96	0.93
	std	0.005	0.005	0.005	0.004
Precision	max	0.97	0.90	0.95	0.89
	mean	0.964	0.898	0.944	0.886
	min	0.96	0.89	0.93	0.88
	std	0.005	0.004	0.009	0.005
Recall	max	0.94	0.95	0.93	0.87
	mean	0.936	0.924	0.914	0.858
	min	0.93	0.89	0.90	0.85
	std	0.005	0.022	0.013	0.008
F1-Score	max	0.95	0.92	0.94	0.88
	mean	0.948	0.910	0.930	0.872
	min	0.94	0.89	0.92	0.87
	std	0.004	0.012	0.007	0.004
Number of parameters		215,937	7,334,257	8,939	43,442

1.5 Discussion and future work

The significance of the developed model lies in its remarkable accuracy in classifying marine species based on individual click signals. This achievement opens the door to a ground-breaking application: the identification of species within a sequence of clicks, commonly known as a "click train." With the demonstrated high accuracy in classifying individual clicks, it is suggested that reliable identification of the species corresponding to a click train can be achieved without the need for a new machine learning model. This can be accomplished through the ingenious use of majority voting and the establishment of a confidence threshold.

To operationalize this concept, the first step involves establishing two key definitions for clarity:

Target Species Clicks: These clicks specifically pertain to the specimens aimed to be identified.

Noise Clicks: These are clicks extracted from the original signal by the click detector but do not belong to the target species.

The approach leverages the capabilities of machine learning and deep learning-based models in effectively identifying and classifying clicks from the target species. These models excel in applying their learned knowledge, having been trained exclusively on clicks from the species of interest, thereby enabling them to distinguish these species accurately. It is essential to highlight that these models have not been trained on noise clicks, as there is no logical rationale for doing so. Expecting the models to accurately identify the target species when presented with a mix of noise clicks and target species clicks is generally impractical.

Solution 1: Leveraging Confidence Scores for Click Train Assessment

An initial approach involves providing all detected clicks to the deep learning model for labelling. Since the model lacks training on noise clicks, it may assign low-confidence labels from the categories it was specifically trained on. averaging the confidence scores within a "click train", allows the assessment of whether the species of interest is likely present or not. This method utilizes the model's confidence scores as an indicator of the likelihood of the species of interest being present within the detected clicks.

Solution 2: One-Class Classification for Enhanced Reliability

If the initial solution proves insufficient, an alternative approach is to implement a deep learning-based one-class classifier. This method involves training a classifier exclusively on data associated with just one species against everything else. For each specific target species, a dedicated one-classifier is designed. Confidence levels can once again be utilized to perform reliable labelling based on this approach, providing an added layer of sophistication and reliability in species identification within click trains.

These innovative solutions represent a strategic shift from conventional classification approaches, allowing for more nuanced and accurate assessments of species presence within acoustic data. As the exploration and refinement of these methods continue, the potential applications for real-world marine mammal monitoring become increasingly apparent, opening avenues for advancements in marine biology and environmental conservation.

The Morales dataset, which is expert-labelled, would be incorporated into the research efforts. To integrate this valuable dataset into these simulations and to align with its characteristics, necessary adjustments would be applied. The Morales dataset stands out as a well-constructed and meticulously labelled resource, providing a robust foundation for ongoing work.

Furthermore, it is essential to note that examining the performance of these simulated models on the Morales dataset is a key aspect of the future plans. This strategic move allows validating the efficacy and generalizability of these models on real-world data, providing a crucial step towards real-world applications.

As the project continues to refine and optimize the models based on expert-labelled data, it is anticipated to uncover valuable insights that will not only enhance the robustness of proposed algorithms but also contribute to the advancement of the broader field.

In summary, the ongoing evaluation of proposed models on the Morales dataset, coupled with continuous refinement and collaboration with domain experts, positions the project at the forefront of developing cutting-edge solutions with real-world impact.

1.6 References

Gillespie, D., Mellinger, D. K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., & Trinder, P. (2009). PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *J Acoust Soc Am*, *125*, 2547. https://doi.org/10.1121/1.4808713

Palmer, K. J., Brookes, K., & Rendell, L. (2017). Categorizing click trains to increase taxonomic precision in echolocation click loggers. *J Acoust Soc Am*, *142*, 863–877. https://doi.org/10.1121/1.4996000

Soldevilla, M. S., Henderson, E. E., Campbell, G. S., Wiggins, S. M., Hildebrand, J. A., & Roch, M. A. (2008). Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. *J Acoust Soc Am*, *124*, 609–624. https://doi.org/10.1121/1.2932059

Usman, A. M., Ogundile, O. O., & Versfeld, D. J. J. (2020). Review of Automatic Detection and Classification Techniques for Cetacean Vocalization. *Ieee Access*, *8*, 105181–105206. https://doi.org/10.1109/access.2020.3000477

Veneruso, G., Torrens-Spence, Green, A., & Gillepsie, D. (2020). Investigating cetacean response around an operating Deep Green Unit. *(No. SC2-RD-B11), Bangor University.*