



## Article

# Image-Based Classification of Double-Barred Beach States Using a Convolutional Neural Network and Transfer Learning

Stan C. M. Oerlemans <sup>1</sup>, Wiebe Nijland <sup>1</sup> , Ashley N. Ellenson <sup>2</sup> and Timothy D. Price <sup>1,\*</sup>

<sup>1</sup> Department of Physical Geography, Faculty of Geosciences, Utrecht University, P.O. Box 80.115, 3508 TC Utrecht, The Netherlands

<sup>2</sup> College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, OR 97330, USA

\* Correspondence: t.d.price@uu.nl

**Abstract:** Nearshore sandbars characterize many sandy coasts, and unravelling their dynamics is crucial to understanding nearshore sediment pathways. Sandbar morphologies exhibit complex patterns that can be classified into distinct states. The tremendous progress in data-driven learning in image recognition has recently led to the first automated classification of single-barred beach states from Argus imagery using a Convolutional Neural Network (CNN). Herein, we extend this method for the classification of beach states in a double-barred system. We used transfer learning to fine-tune the pre-trained network of ResNet50. Our data consisted of labelled single-bar time-averaged images from the beaches of Narrabeen (Australia) and Duck (US), complemented by 9+ years of daily averaged low-tide images of the double-barred beach of the Gold Coast (Australia). We assessed seven different CNNs, of which each model was tested on the test data from the location where its training data came from, the *self-tests*, and on the test data of alternate, unseen locations, the *transfer-tests*. When the model trained on the single-barred data of both Duck and Narrabeen was tested on unseen data of the double-barred Gold Coast, we achieved relatively low performances as measured by *F1* scores. In contrast, models trained with only the double-barred beach data showed comparable skill in the self-tests with that of the single-barred models. We incrementally added data with labels from the inner or outer bar of the Gold Coast to the training data from both single-barred beaches, and trained models with both single- and double-barred data. The tests with these models showed that which bar the labels used for training the model mattered. The training with the outer bar labels led to overall higher performances, except at the inner bar. Furthermore, only 10% of additional data with the outer bar labels was needed for reasonable transferability, compared to the 20% of additional data needed with the inner bar labels. Additionally, when trained with data from multiple locations, more data from a new location did not always positively affect the model's performance on other locations. However, the larger diversity of images coming from more locations allowed the transferability of the model to the locations from where new training data were added.

**Keywords:** machine learning; Argus; ResNet50; transfer learning; CNN; deep learning; beach state; nearshore morphology



**Citation:** Oerlemans, S.C.M.; Nijland, W.; Ellenson, A.N.; Price, T.D. Image-Based Classification of Double-Barred Beach States Using a Convolutional Neural Network and Transfer Learning. *Remote Sens.* **2022**, *14*, 4686. <https://doi.org/10.3390/rs14194686>

Academic Editor: Javier Marcello

Received: 28 July 2022

Accepted: 15 September 2022

Published: 20 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

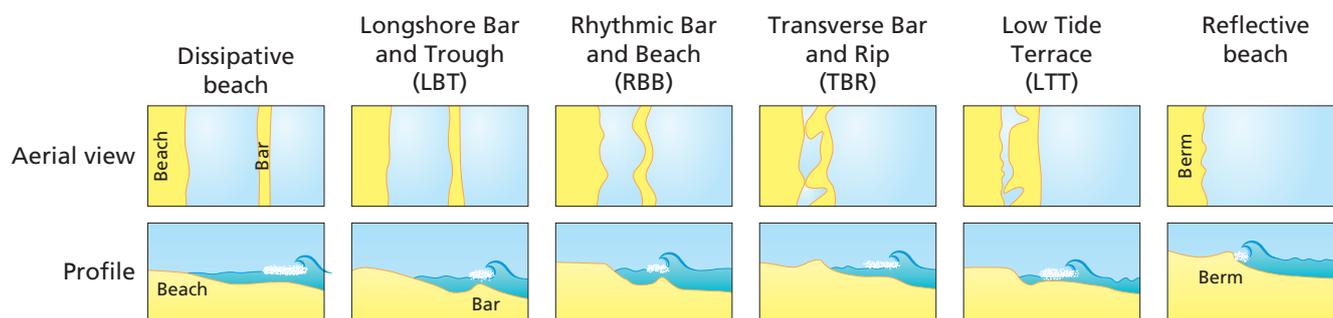


**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The nearshore zones of sandy coasts are highly dynamic areas, where wave breaking and wave-driven currents constantly rearrange nearshore sediment into complex, consistently occurring morphological patterns. Sandbars, in particular subtidal sandbars, are responsible for significant spatiotemporal variations in nearshore bathymetric profiles [1,2]. They are fundamental to various nearshore processes: they can reduce beach-dune erosion through wave dissipation [3–6]; exchange sand with the subaerial beach [7]; control advection, mixing and distribution of nutrients and pollutants [8]; and determine recreational safety [9]. Hence, monitoring and understanding the spatiotemporal variability of nearshore morphology is essential to both scientists and coastal managers.

Sandbar morphology at a given beach may range from shore-parallel ridges to an alongshore alternation of shore-attached bars and rip channels [10,11]. Ref. [12] created the most widely used beach state classification scheme for single-barred coasts (Figure 1). In this scheme, they identified three basic beach types: reflective, intermediate and dissipative, consisting of six beach states in total with distinct sandbar configurations. The two end members, Reflective (R) and Dissipative (D), relate to, respectively, low and high-energetic conditions. The intermediate states, corresponding to high to low-energetic conditions, were identified as Longshore Bar and Trough (LBT), Rhythmic Bar and Trough (RBB), Transverse Bar and Rip (TBR) and Low Tide Terrace (LTT) (Figure 1) [12]. In general, during low-energetic accretionary conditions, sandbars advance sequentially in downstate direction, whereas sandbar morphology may jump to a higher state during high-energetic erosional sequences [13].



**Figure 1.** Schematic representation of the beach states as defined in the classification scheme of [12]. (Adapted from: [14]).

Initially, this classification scheme was used for the classification of beach states of single-barred beaches. Eventually, the scheme of [12] was extended to be applicable on multi-barred beaches. A multi-bar state model was devised in which each bar can go through the same states as in the original single-bar model [10]. On a double-barred beach, the seaward-most outer bar evolves more slowly through the bar states than the landward-most inner bar. The outer bar often occurs as a quasi-inactive feature during low-energetic periods [11].

In situ measurements, where hydrodynamics, sediment transport and bed levels are measured directly in the field, have often been used to study sandbars and their dynamics. However, this method is spatially limited and expensive in terms of time and money. In addition, these field-based experiments are often restrained by the surf zone's harsh and potentially dangerous conditions [7,15–17]. Luckily, the nearshore zone exhibits many optical signatures that can be exploited. Wave breaking is very obvious to the eye as bright patches of foam on the water surface. Since waves tend to break in shallow water, the locations of concentrated foam and the spatial pattern this foam forms can be used to locate the position of submerged sandbars and to classify them into a beach state [16]. These optical signatures led to the extensive use of remote sensing applications such as the Argus video monitoring systems [11,18], permitting frequent, spatially extensive and high-resolution data monitoring [3,16,19].

The research on sandbar morphology and dynamics has thus far mainly been conducted using quantitative measures [11,20,21] or conventional machine learning (ML) algorithms. Ref. [22] used machine learning for the automated classification and mapping of the seabed. Ref. [23] used a Neural Network (NN) to study the predictability of nearshore sandbar migration. Additionally, ref. [24] used an Artificial Neural Network (ANN) to produce a model that estimates cross-shore bar location from raw Argus images of double-barred beach systems. However, the surge of deep learning (DL) in nearshore studies has enabled the development of innovative algorithms to study sandbars. Ref. [25] applied a Recurrent Neural Network (RNN) to study nearshore sandbar behaviour. Additionally,

ref. [26] applied a Convolutional Neural Network (CNN) to derive surf-zone bathymetry from video imagery.

In contrast with conventional machine learning (ML), DL algorithms do not need human-designed rules. Instead, they use large amounts of data to map the input to a specific label. DL can automate the learning of features and enables classification to be achieved in a single shot [27]. The tremendous progress in DL, particularly in image recognition and classification is primarily due to the application of CNNs [27]. CNNs are a specific DL method designed to learn spatial patterns, allowing image-specific features into the network's architecture and making the network more suited for image-focused tasks [28]. CNNs have been shown to produce state-of-the-art performances for various image recognition and classification tasks [29–33]. These state-of-the-art performances could partly be attributed to the use of transfer learning [34]. Instead of training a model from scratch, requiring a large amount of data, transfer learning utilizes the parameters of CNN architectures pre-trained on large public datasets, such as ImageNet [32], to be fine-tuned by new datasets in other image classification problems [35].

In a recent study, Ref. [36] trained a CNN from scratch for the automated classification of single-barred beach states from Argus imagery of the coasts of Duck (US) and Narrabeen (Australia). They implemented various data combinations to train and test the models and compared their results to the inter-labeller agreements resulting from the classification carried out by humans. They showed that CNNs trained and tested at the same site had comparable skills to the inter-labeller agreement, with a higher overall skill at Duck than at Narrabeen. For both sites, the highest accuracy was in classifying the low-energy R and LTT states, while the lowest skill of the CNN was classifying the rhythmic states of RBB (at Narrabeen) and TBR (at Duck). The performance decreased when the CNNs trained with the data of one location were tested on the other site's test data. However, they showed that the models trained on data from multiple locations achieved performances comparable to the CNNs trained with the data from a single location, with at least 25% of the training data coming from each location. Ref. [36] showed that single-barred beach systems could be successfully classified using CNNs. However, it is unknown how CNNs can be applied to double-barred systems. Hence, the aim of this study is to extend

## 2. Field Sites and Datasets

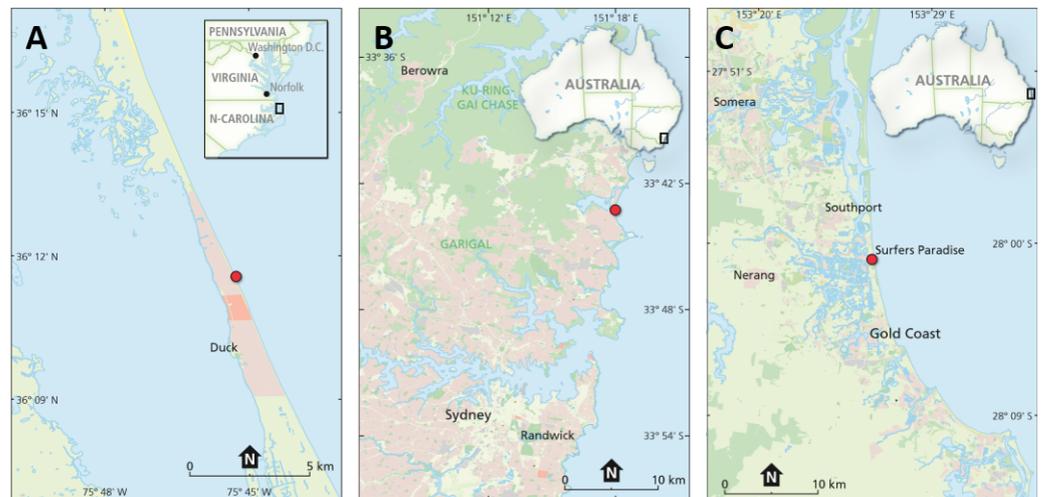
We used the data of three field sites, displayed in Figure 2. These datasets consist of grey-scale Argus imagery. Figure 3 shows examples of images from all three datasets, with each column containing images from a different site and each row containing a different beach state. In this section, we introduce the three field sites and the corresponding datasets.

### 2.1. Field Sites

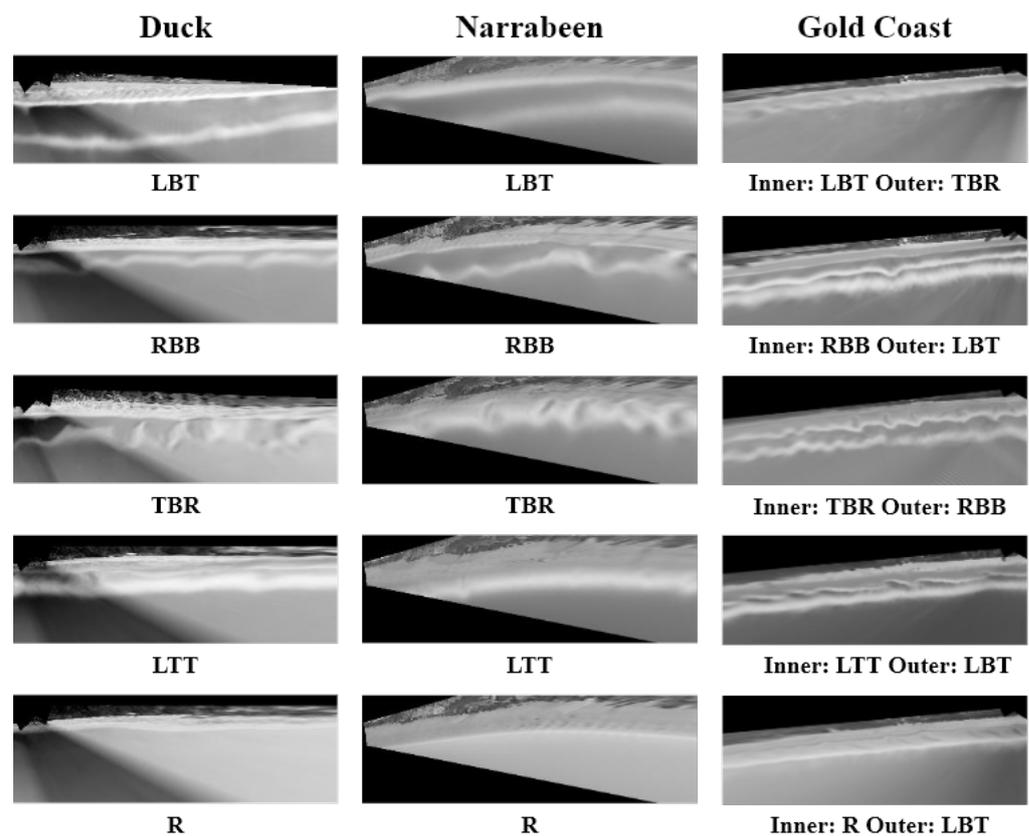
#### 2.1.1. Duck

The first site is the sandy barrier beach of Duck, situated at the U.S. Army Engineering Research and Development Center Field Research facility, North Carolina. As shown in Figure 2A, Duck is located between two water bodies, the Currituck Sound in the west and the North Atlantic Ocean in the east. The Argus installation at the Field Research Facility faces the North Atlantic Ocean. In general, the beach of Duck is classified as an intermediate beach with one and, occasionally, two sandbars present. The waves coming from the North Atlantic Ocean vary seasonally, with higher incident wave energy in winter and lower incident wave energy in summer [37]. The annual significant wave height is 1.1 m, with waves coming from the south during spring and summer and from the north during winter. In addition, winter storms consist of extra-tropical and tropical cyclones [38]. The mean spring tide range is micro-tidal at 1.2 m. The beach slope averages 0.108 at the foreshore and decreases further offshore to 0.006 at a depth of 8 m. The sediment consists of medium to fine-grain quartz with finer sands further offshore. The median grain size

between the bar and the shoreline is approximately 0.5 mm, with 20% carbonate material. Offshore of the bar, the median grain size becomes 0.2 mm [39].



**Figure 2.** The location of the three field sites, (A): Duck (US), (B): Narrabeen (Australia) and (C): Gold Coast (Australia).



**Figure 3.** Examples of Argus imagery showing the five different beach states. Left column: Duck, middle column: Narrabeen and right column: Gold Coast. Note that the inner and outer bars of the Gold Coast have individual labels, with the outer bar labels only exhibiting the TBR, LBT and RBB states.

### 2.1.2. Narrabeen-Collaroy

The second site is the 3.6 km-long embayment of Narrabeen-Collaroy, referred to as Narrabeen. Narrabeen is located within the Northern Beaches region of the metropolitan Sydney, Australia (Figure 2B). A lagoon backs the northern half of the barrier and is connected to the ocean via a shallow, narrow inlet, which opens and regularly closes at the embayment's northern end. Adjacent to the southern end of the beach is a prominent headland. This headland and the curvature of the embayment result in a distinct alongshore wave gradient [40–42]. The wave climate at Narrabeen is mildly seasonal, with high-energy cyclones and east-coast lows more prominent in Austral winter months and low-energy swells more prominent in Austral summer. At interannual time scales, the wave climate is influenced by the El Niño-Southern Oscillation (ENSO), resulting in periods of less energetic and a more southerly wave climate.

On the other hand, La Niña periods result typically in more energetic and easterly wave climates. The deep-water wave climate for the Sydney region is moderate to high wave energy with a mean wave height of 1.6 m. It is dominated by long continuous period swell waves coming from an SSE direction. These swell waves are generated from mid-latitude cyclones propagating in the southern Tasman Sea, south of Australia. Superimposed on these swell waves are storm events typically defined for this region by a significant wave height threshold of 3 m [43]. Tides are micro-tidal and semi-diurnal, with a mean spring tide of 1.3 m. The sediment is mainly uniform along the beach and consists of primarily fine to medium quartz sand with 30% carbonate materials [44].

### 2.1.3. Gold Coast

The third field site is from the popular tourist destination of Surfers Paradise, located at the northern end of the Gold Coast in South East Queensland, Australia (Figure 2C). The beach of the Gold Coast consists mainly of a double-barred system. In winter, persistent SSE swells and high-energy mid-latitude cyclones result in a net littoral drift to the north in the order of 500,000 m<sup>3</sup> per year [45]. The root-mean-square wave height at the Gold Coast is typically about 0.8 m. However, East Coast and tropical lows can increase the significant wave height to 1.5 m, with an estimated return interval of 2 years for a significant wave height of 3.5 m [46]. The tide is semi-diurnal with a spring tidal range from 1.5 m to 2 m [11]. Furthermore, the nearshore predominantly consists of quartz sand with a median grain size of 0.225 mm and exhibits an average slope of approximately 0.02.

Between 1999 and 2000, a 1.2 Mm<sup>3</sup> beach nourishment was undertaken to maintain and enhance the sub-aerial beach width [47]. The implementation of the nourishment near the study site of the Gold Coast commenced in early November 1999 and reached its southernmost extension in June 2000. The effect of the nourishment on the sandbars was most pronounced in March and April 2000. Our study site was restricted to the area south of the nourishment, and our sandbar data were not directly impacted. Furthermore, a hybrid coastal protection-surfing reef structure is located at Narrowneck, north of our study site [48]. Hence, only the imagery containing the southern area of our study site was used.

## 2.2. Datasets

We used the single-barred datasets of Duck and Narrabeen from [36]. These datasets consist of orthorectified time exposure (*timex*) grey-scale imagery collected hourly from Argus stations (for details on the Argus system see [36]). Argus images are the average of video frame observations of the surf zone collected over a period of time of fifteen minutes, which results in a picture of beach morphology over the surf zone [16]. The effects of lighting due to the angle of the sun or cloud cover, or a lack of wave-breaking signal due to low waves or high tide were reduced by using *daytimex* images. Daytimex images are the temporal mean of all Argus images collected hourly during a single day, thus averaging out tidal dependencies and natural modulations in wave-breaking, such as a lack of wave-breaking, and removing persistent optical signatures. Furthermore, the daytimex images

were orthorectified onto a domain of 900 m alongshore and 300 m cross-shore, with a ground resolution of 2.5 m  $\times$  2.5 m. The dataset was collected over the years 1987–2014 for Duck and 2004–2018 for Narrabeen. The images were combined from various camera views depending on the number of cameras installed and functional in the Argus system. This ranged from one to three for Duck and only one for Narrabeen.

The datasets of Duck and Narrabeen were manually labelled by [36] with small adjustments made by the authors of this paper. As can be seen in Figure 3, the datasets consist of images labelled with the beach states R, LTT, TBR, RBB and LBT, with 680 images of the site of Duck and 687 of Narrabeen. Table 1 shows the state distribution of these single-barred datasets and the number of adjustments to the original dataset made by the authors of this paper.

**Table 1.** Per state data distribution for Duck, Narrabeen and the Gold Coast’s inner and outer bar (\* Differences with [36] after adjusting labels).

State	Duck	Narrabeen (*)	Gold Coast Inner Bar	Gold Coast Outer Bar
R	125	125 (0)	377	0
LTT	126	164 (+7)	1779	0
TBR	140	157 (+8)	596	1118
RBB	126	106 (−24)	74	746
LBT	163	135 (+9)	197	1159
Total	680	687	3023	3023

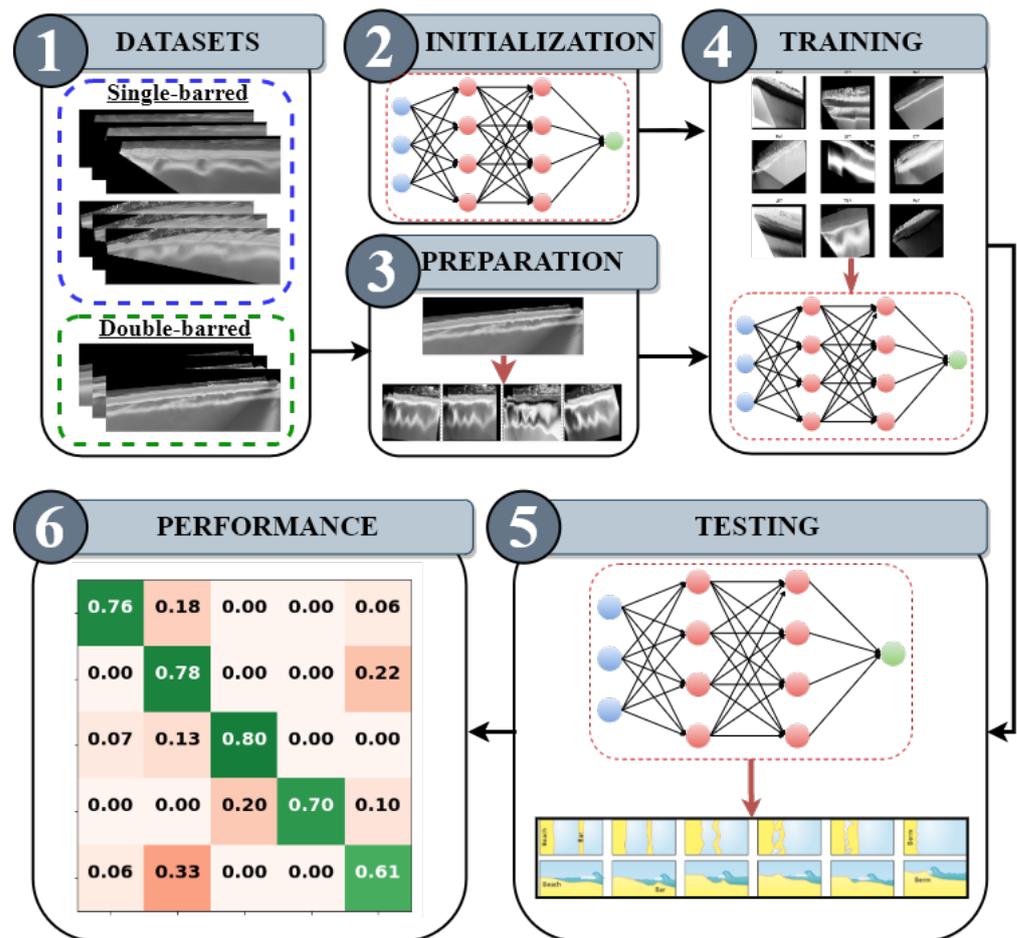
The Gold Coast dataset consists of labelled images collected over the years 1999–2008 [11]. However, instead of daytime-images as for Narrabeen and Duck, the dataset of the Gold Coast comprises 3023 daily low-tide time-exposure images (for details on the Argus system see [49]). These images were created by time-averaging 600 individual snapshots sampled at 1 Hz during low-tide. Each image spans 2500 and 900 m in the longshore and cross-shore direction, respectively. Ref. [11] used this dataset to individually classify the inner and outer bars. Ref. [11] identified two additional intermediate beach states, the erosive Transverse Bar and Rip (eTBR) and the rhythmic Low Tide Terrace (rLTT), related to the dominant oblique angle of wave incidence and the multiple bar setting, respectively. For this study, the labels of the images corresponding to these additional bar states have been adjusted to fit the best alternative beach state to correspond with the scheme of [12].

Table 1 shows the state distribution for the dataset of the Gold Coast over the entire collection. Every image has two labels, one belonging to the inner and another to the outer bar. The inner bar was labelled with all five beach states (R, LTT, TBR, RBB and LBT). In contrast, the outer bar was only observed to exhibit higher-energy states TBR, RBB and LBT (Figure 3).

### 3. Methodology

To explore and evaluate the application of a CNN in combination with transfer learning on the automated classification of beach states in a double-barred beach system, the following step-by-step process was used (Figure 4):

1. Collect the datasets;
2. Model initialization;
3. Prepare the data;
4. Train models with various datasets;
5. Test the models on various locations;
6. Report the performance for each test.



**Figure 4.** Schematic block diagram of the step-by-step methodology.

The datasets used in this study (step 1) are described in the previous section and comprise the Argus imagery from the single- and double-barred beach systems of Duck, Narrabeen and the Gold Coast.

### 3.1. Model Initialization

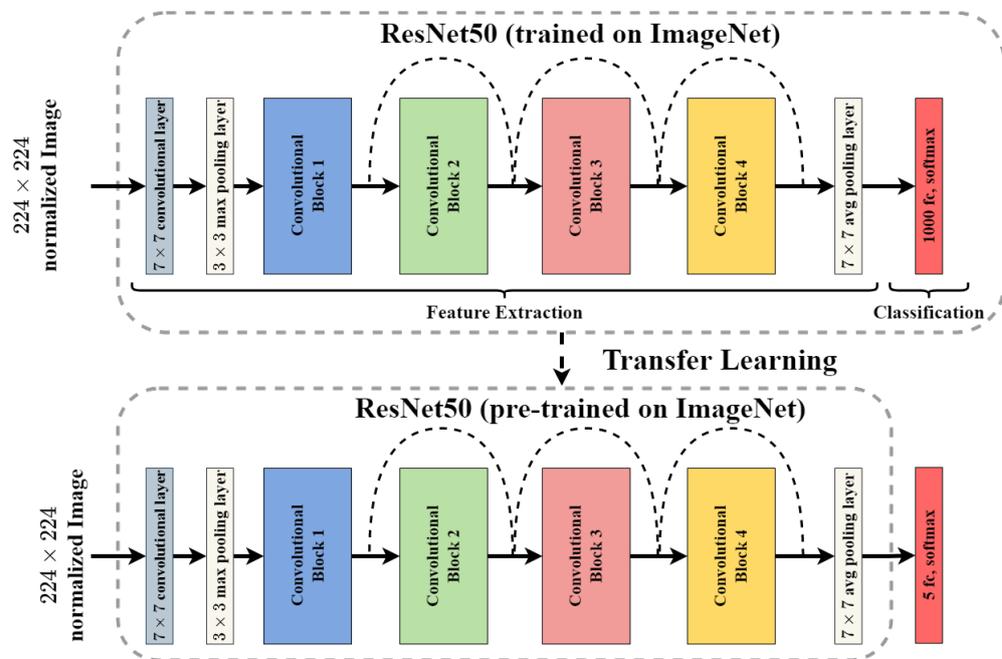
The second step is to prepare the model for the task at hand. In this application, a CNN takes the images of sandbar morphology as input and produces a beach state as an output. A CNN consists of various layers of artificial neurons. These artificial neurons, similar to neuron cells used by the human brain for passing input signals, are mathematical functions used to process the various inputs and to give a single output. The behaviour of each neuron contained in a CNN is defined by the value of its weights, the so-called trainable parameters [27].

The architecture of a common type of CNN consists of numerous layers, in general, a convolutional layer separating and identifying the distinct features of an image, a pooling layer responsible for the reduction in the spatial size of the convolved features and a fully connected (FC) layer that takes the output from the previous processes and predicts the image's class based on the retrieved features [27].

#### 3.1.1. ResNet50

Many CNN architectures have been created [50,51]. Modifications on these architectures have eventually resulted in state-of-the-art models [52–54]. One such architecture is the Residual Network (ResNet) [55]. ResNet became a family of models based on the number of layers, starting with 18 and going as deep as 1202 layers. In this study, the 50-layered architecture of ResNet50 was used following [55].

Figure 5 shows a schematic block diagram of a typical ResNet50 architecture. This architecture consists of two parts: the *feature extraction* part, consisting mainly of convolutional and two pooling layers, and the *classification* part existing of the FC layer. Note that the number of nodes in the FC layer corresponds to the number of unique classes in the training dataset.



**Figure 5.** Schematic block diagram of ResNet50 and transfer learning. Note that the classification layer is replaced with 5 as the number of unique classes in the datasets.

### 3.1.2. Training Protocols and Transfer Learning

Training a CNN is described as minimizing the differences between output predictions and the given labels on a training dataset. The training is carried out by feeding the CNN with a large dataset of images labelled with their corresponding class label. The images are added individually or in groups depending on the batch size. The CNN network processes the images and compares them with the class label of the input image. *Backpropagation* is a technique that is commonly used for training neural networks, where a *loss function* and an *optimization algorithm* play essential roles. The loss function measures the compatibility between output predictions and the given labels, and the optimization algorithm updates the trainable parameters according to the loss value. During training, the CNN goes through the entire dataset multiple times. Each time that all input images have had the opportunity to update their weights is called an *epoch* [56].

When the training of a CNN starts with random initialized weights, it is trained from scratch. A large amount of well-annotated data is required to train a model this way. This data are not always available as the acquisition and annotation of data can be very time-consuming. Additionally, training a model from scratch can take a long time. Rather than collecting and annotating large datasets, it is popular to use networks that have been pre-trained on vast datasets, such as *ImageNet* with over 14 M images [57] and to use a technique called transfer learning to train models [58]. There are two different ways in which transfer learning can be applied. The network can be used either for *feature-extraction*, in which only the classification part of the pre-trained network is updated for classification, or for *fine-tuning*, in which all trainable parameters of the pre-trained are updated using a new dataset, before using them for classification [59].

In this study, the ResNet50 model, pre-trained on the public dataset ImageNet, was fine-tuned to classify images into beach states. Initially, the classification part of the pre-trained network corresponds to the number of unique classes within ImageNet. Hence,

the classification part was reconfigured to correspond to the number of beach states within the unique datasets used in this study, as can be seen in Figure 5. In addition, several adjustments were made to the training settings as used by [36] to increase the computational speed and to optimize the model performance; these adjustments were made empirically. We used the OneCycleLR Policy following [60] to tune the learning process, and we implemented the Early Stopping method as described by [61] to stop training automatically when the performance is not increasing for a specific amount of training rounds defined by the *patience*. The optimization algorithm called stochastic gradient descent (SGD) was used [62]. The parameters and corresponding settings used in this study are shown in Table 2.

**Table 2.** The parameters and corresponding settings used in this study.

Type	Parameter	Setting
OneCycleLR	Max momentum	0.95
OneCycleLR	Min momentum	0.85
OneCycleLR	Max learning rate	0.01
SGD	Learning rate	0.002
Network	Batch size	32
Early stopping	Patience	20

### 3.2. Data Preparation

Step 3, before training the model, consists of data preparation. First, the data were distributed into three sets: one for training, one for validation to give an unbiased evaluation of the model performance while tuning the model parameters and one for testing the trained model on unseen data to evaluate the performance. The number of images used of each state is detailed in Table 1. As our data consist of time series from different periods, we sorted them chronologically per field site before splitting them. This sorting was carried out so that images from consecutive days and most likely having very similar optical signatures were not distributed in the different sets. Generally, the distribution was 80%, 10% and 10% for the training, validation and test datasets, respectively. An exception was made in the experiment in which the model's performance was evaluated as a function of training data composition, explained further in Section 4.2.

As the number of images belonging to each state were not equally distributed over the recorded periods, sorting the data before splitting resulted in datasets in which the state representation was highly unbalanced, as more images were available from some states than others. An unbalanced training dataset might adversely affect the training process leading to a bias during the classification phase of the model [63]. To counter this problem and to balance out the bias, we oversampled the data based on the states' occurrence ratios in the training dataset. Oversampling is the method in which images from a minority class in the training datasets are given a larger share in the training process. This is carried out by weighting, where a larger weight is applied to the images belonging to a minority class compared to the weights applied to a majority class.

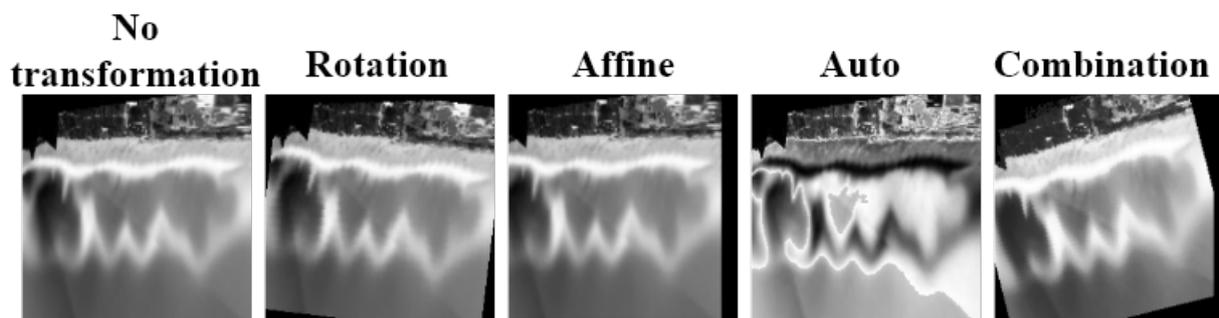
Furthermore, for optimal training performance, the images must be in the same input format as the images on which the network is pre-trained. Hence, each image was processed to meet the requirements of the pre-trained ResNet50 model. The resolution of the images was resized to meet  $224 \times 224$  pixels, and the values of these pixels were normalized to the range [0, 1].

The datasets used in this study were limited by the number of times each state occurred over the period spanned by the dataset and the number of images that distinctly exhibited one beach state [36]. We increased the number of training data by applying geometric transformations, or so-called augmentations, to the training images while simultaneously preserving the morphology and labels, similar to [36]. Table 3 shows the combination of

augmentations and their corresponding functions used for this study. Examples of the resulting images after transformation are shown in Figure 6.

**Table 3.** The augmentations used before training. In the left column, the transformations applied, and in the right column, the function of the corresponding augmentation.

Augmentation	Function
RandomRotation (15)	Rotate the image randomly between 0 and 15 degree angles
AutoAugment	Data augmentation method based on the proposed method of [64].
RandomAffine (0, translate = (0.15, 0.20))	Random affine transformation (translation with horizontal translates randomly between maximum fractions of 15 and 20) of the image, keeping the center invariant.



**Figure 6.** Examples of transformed images after applying the augmentations as described in Table 3: RandomRotation, RandomAffine, AutoAugment and the combination of the three.

#### 4. Experimental Setup

The writing of our algorithm, which is publicly available online at <https://github.com/StanOerlemans/DeepBeachStateV2>, accessed on 14 January 2022, and our two experiments were done using Google Colaboratory (Google Colab, <https://colab.research.google.com/>, accessed on 14 January 2022). Google Colab provides a RAM of 13 GB, an NVIDIA Tesla K89 12 GB GPU and an Intel(R) CPU @ 2.30 GHz processor. In the first experiment, only the single-barred beach datasets of Duck and Narrabeen were used to train the models. In the second experiment, we assessed the performance of models trained using additional data from the double-barred beach of the Gold Coast. The CNNs were trained using a dataset consisting of the data from a single location or on a combined dataset consisting of the data of multiple locations (step 4).

Additionally, each CNN was tested (step 5) on the test data from the same location as the training data, the *self-tests*, and on the test data of unseen locations, the *transfer-tests*. Ensembles of 10 CNNs were trained for each training setup, of which the highest performing CNN was used for further evaluation. The names given to the resulting models corresponding to these setups, the corresponding experiments and the data combinations on which they were trained and tested are shown in Tables 4 and 5, respectively.

**Table 4.** The two experiments, the corresponding models and the data combinations used to train the models.

Experiment	Model Name	Training Data
1.	DUCK-CNN	Duck
	NBN-CNN	Narrabeen
	CombinedSINGLE-CNN	Duck & Narrabeen
2.	INNER-CNN	Gold Coast's inner bar labels
	OUTER-CNN	Gold Coast's outer bar labels
	CombinedINNER-CNN	Duck, Narrabeen & Gold Coast's inner bar labels
	CombinedOUTER-CNN	Duck, Narrabeen & Gold Coast's outer bar labels

**Table 5.** The two experiments, the corresponding models and the data combinations on which the models were tested.

Experiment	Model Name	Test Data			
		Self-Tests		Transfer-Tests	
1.	DUCK-CNN	Duck		Narrabeen	
	NBN-CNN	Narrabeen		Duck	
	CombinedSINGLE-CNN	Duck & Narrabeen	Duck	Narrabeen	Gold Coast
2.	INNER-CNN	Gold Coast's inner bar labels			-
	OUTER-CNN	Gold Coast's outer bar labels			-
	CombinedINNER-CNN	Duck, Narrabeen & Gold Coast	Duck	Narrabeen	Gold Coast
	CombinedOUTER-CNN	Duck, Narrabeen & Gold Coast	Duck	Narrabeen	Gold Coast

#### 4.1. Experiment 1: Single-Bar Beach Models

In the first experiment (Table 4), only the single-barred beach datasets of Duck and Narrabeen were used to train the models. Three 10-member ensembles of CNNs were trained. The first two ensembles were single-location ensembles, meaning that the training data came from either Duck or Narrabeen. The data distribution at Duck resulted in 544 training images, and 68 validation and testing images. The CNNs trained at Duck are referred to as *DUCK-CNN*. The data distribution at Narrabeen resulted in 549 training images, and 69 validation and testing images. The CNNs trained at Narrabeen are referred to as *NBN-CNN*. The resulting models were self-tested and transfer-tested.

The third 10-member ensemble trained in Experiment 1 was trained with the training data of both Duck and Narrabeen combined and is referred to as *CombinedSINGLE-CNN*. The CombinedSINGLE-CNN was self-tested on the test data of both Duck and Narrabeen together and individually. In addition, to assess the skill of a model trained with only single-barred data on unseen double-barred beach data, the CombinedSINGLE-CNN was transfer tested on the data of the Gold Coast. When analysing the performance on the Gold Coast, the model was tested on the entire cross-shore of the Gold Coast and compared with both inner and outer bar labels separately. Hence, the performance on each bar can be assessed individually.

#### 4.2. Experiment 2: Double-Bar Beach Models

In the second experiment (Table 4), four models were trained to assess the performance of models trained using not only single-barred beach data, such as in the first experiment (Section 4.1), but also double-barred beach data. In this experiment, additional data from the Gold Coast were used for training.

Firstly, two models were trained with only the double-barred beach data, with either the inner or outer bar labels. These two models were trained by feeding it 2418 training

images from the Gold Coast with labels corresponding to either the inner bar or the outer bar. This resulted in the *INNER-CNN* and the *OUTER-CNN*.

Additionally, two models with data coming from Duck, Narrabeen and incrementally added data of the Gold Coast, with either the inner or outer bar labels, were trained (See Table 4). The training data of Duck and Narrabeen combined, as used for training the *CombinedSINGLE-CNN*, was supplemented with data of the Gold Coast. The Gold Coast's training data, with either the inner or outer bar labels, were added incrementally as a percentage of the total training data (1093) of Duck and Narrabeen combined. This created the *CombinedINNER-CNN* and the *CombinedOUTER-CNN*, respectively. In these cases, data with either the inner or outer bar labels were added in 10% of the Duck and Narrabeen training data until 100% of additional data were added and the total amount of training data were doubled. During this process, the test and validation data were kept constant, in which the test and validation data of the Gold Coast were added in an equal amount to the number of test and validation images from Duck and Narrabeen.

The performance of each run with increasing amounts of data from the Gold Coast was evaluated to determine the best-performing *CombinedINNER-CNN* and *CombinedOUTER-CNN*. Subsequently, these were used for further evaluation and testing. The testing included self-testing at all test sets individually and combined, following Table 5.

#### 4.3. Performance Measures

In step 6, several metrics were used to report the performances. The training and validation losses can be used to assess the model's training performance and optimization. The loss quantifies the error produced by the model. We used the training loss as a metric to assess how our model fitted the training data, and we used the validation loss as a metric to assess the performance of our model on the validation set.

We chose several performance metrics to evaluate and validate our models' effectiveness. Often chosen metrics are sensitivity, or so-called recall, specificity, and accuracy. The accuracy is interrelated with sensitivity and specificity. In addition, a confusion matrix in a classification task is a tabular representation of the per-state classifications of the model. Table 6 shows an example of such a confusion matrix with the evaluation factors described as follows:

- **True Positive (TP):** *TP* denotes the number of cases in which the CNN classified *Yes* correctly.
- **True Negative (TN):** *TN* signifies the number of cases the CNN classified *No* correctly.
- **False Positive (FP):** *FP* is the number of cases we classified *Yes* where the true class is *No*.
- **False Negative (FN):** *FN* denotes the number of cases the CNN classified *No* where the true class is *Yes*.

**Table 6.** Confusion matrix for the classification of actual and classified yes or no, with actual and classified, respectively, as True Positive (*TP*) for Yes–Yes, False Negative (*FN*) for Yes–No, False Positive (*FP*) for No–Yes and True Negative (*TN*) for No–No.

		Classified	
		Yes	No
Actual	Yes	<i>TP</i>	<i>FN</i>
	No	<i>FP</i>	<i>TN</i>

Mathematically specificity and accuracy are given as follows:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (1)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} * 100 \quad (2)$$

These metrics are especially effective in the case of a balanced dataset. However, they could give an unreliable representation of the classification performance in imbalanced data sets [65]. Hence, three different metrics were used to assess the overall and per-state performances.

The precision metric is the proportion of correct test results. A downside of precision is that it does not take the *FN* into consideration. It is mathematically represented as follows:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

The recall is the proportion of actual positives classified correctly. One downside of this metric is the fact that it does not take the *FP* into consideration. It is mathematically represented as follows:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The *F1* score is the harmonic mean between precision and recall therefore considering all four types of measures (*FP*, *TP*, *FN* and *TN*) in evaluating overall performance. In this metric, both recall and precision are evenly weighted, and therefore, it is often used when the class distribution in datasets is imbalanced. The *F1* score is mathematically given as follows:

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \quad (5)$$

When the *F1* is high, it signifies a high precision and recall, indicating a good balance between the two. A low *F1* score, on the other hand, only indicates that the model performance was poor over the entire test set and does not indicate whether we have a low precision or low recall. Hence, the overall performance was reported in *F1* scores, whereas the per-state performances were reported in normalized confusion matrices. Normalized means that each of the states is represented as having one samples. Thus, the sum of each row represents 100%. In the matrices, the diagonal represents the recall as calculated with Equation (4). The precision is calculated with Equation (3).

## 5. Results

### 5.1. Single-Bar Beach Models Performances

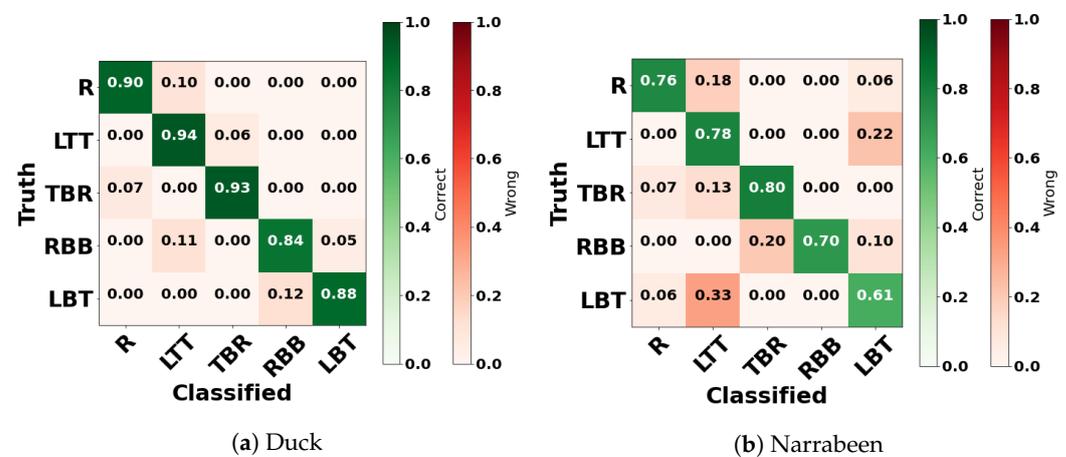
The results in terms of *F1* scores for the single-barred beach classifications are reported in Table 7. The training and validation losses corresponding to the models are reported in Appendix A. The highest performances are in the case of the self-tests, with *F1* values of 0.76 for NBN-CNN and 0.89 for DUCK-CNN. Transfer-testing resulted in lower performances for both single-location models: *F1* scores of 0.66 for NBN-CNN, a decrease of 13%, and *F1* scores of 0.73 for DUCK-CNN, a decrease of 18%. Testing the NBN-CNN at Duck yielded a smaller decrease in performance than vice versa, which is consistent with the results of [36]. This could indicate that the correlation of sandbars and beach states at Narrabeen was more informative than at Duck.

**Table 7.** *F1*-values for the single-barred beach models tested at Duck and Narrabeen.

Model	F1 Values	
	Duck	Narrabeen
NBN-CNN	0.66	0.76
DUCK-CNN	0.89	0.73
CombinedSINGLE-CNN	0.85	0.78

Testing the CombinedSINGLE-CNN gives comparable results to the single-location models. The  $F1$  score, when tested at Duck is 0.85 and when tested at Narrabeen is 0.78. Transfer-testing the CombinedSINGLE-CNN on the Gold Coast resulted in an  $F1$  of 0.32 for the inner bar and an  $F1$  score of 0.55 for the outer bar.

Figures 7 and 8 show the normalized confusion matrices for the tests of the DUCK-CNN and NBN-CNN, respectively. The self-tests of the DUCK-CNN (Figure 7a) resulted in per-state recall values of 0.84–0.94. In this case, 11% of the images with the true RBB state and 10% of the images with the true R state are classified as LTT. In comparison, the per-state performance of the self-test of the NBN-CNN (Figure 8a) has an overall bigger range, ranging between 0.59 and 0.94. The major confusions happen between the linear states, R, LTT and LBT, and between the rhythmic states, TBR and RBB. For the linear states, 29% of the R and 6% of the LBT images were classified as LTT and 11% of the LTT images were classified as LBT. In addition, 20% of the TBR images were classified as RBB.



**Figure 7.** Normalized confusion matrices for the DUCK-CNN tests: (a) The per-state performance of the self-test. (b) The per-state performance of the transfer-test on Narrabeen. The confusion matrices show the true states on the y-axis and the state as classified by the CNN on the x-axis. Note that the confusion matrices show the recall on the diagonal in green and that the precision and  $F1$  score can be calculated by Equations (3) and (5), respectively.

For the transfer-test of the DUCK-CNN (Figure 7b) at Narrabeen, the recall values decreased for all the states compared to the self-test; recall values range between 0.61 and 0.80. The main confusions occur in the adjacent states and between the linear states, in which 33% of the LBT images are classified as LTT and 22% of the LTT images as LBT.

Transfer-testing NBN-CNN at Duck (Figure 8b) resulted in recall values ranging between 0.00 and 1.00. The first significant observation is that the model achieved the lowest recall on the R state, similar to the self-test of NBN-CNN. However, in this case, the NBN-CNN failed to classify any images from the R state. Instead, the images of the R state are all classified as LTT. In contrast, LTT was often misclassified as TBR, RBB or LBT. In addition, the CNN resulted in misclassifications for TBR, of which 40% was classified as RBB, and for LBT, with 33% classified as RBB. The images exhibiting the RBB state, on the other hand, are all correct classified.

The confusion matrices corresponding to the CombinedSINGLE-CNN are shown in Figure 9. When this model was tested on the test data of Duck (Figure 9a), it resulted in recall values ranging between 0.81 and 0.90, with the confusion concentrated between adjacent states. In addition, the LTT and TBR states are most often confused with each other; 19% of LTT was classified as TBR and 13% of TBR images were classified as LTT.

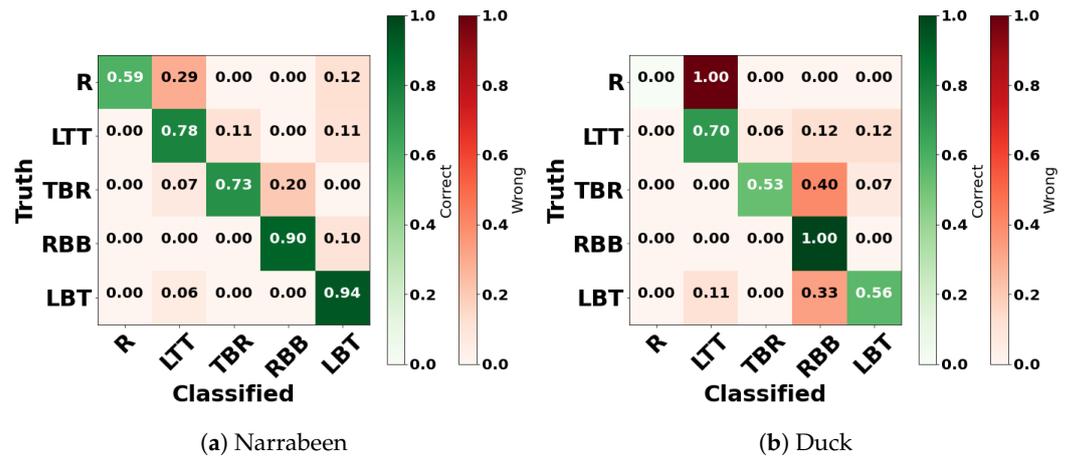


Figure 8. Normalized confusion matrices for the NBN-CNN tests: (a) the per-state performance of the self-test and (b) the per-state performance of the transfer-test on Duck.

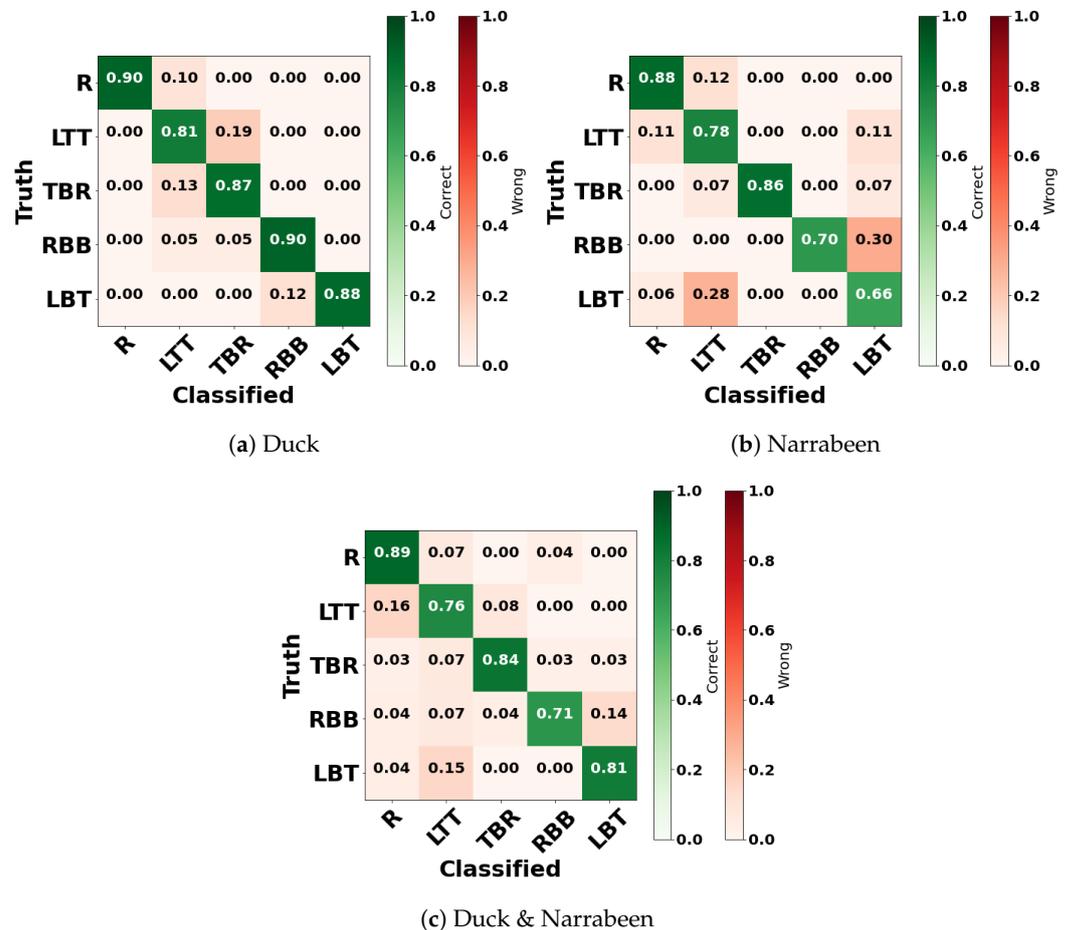
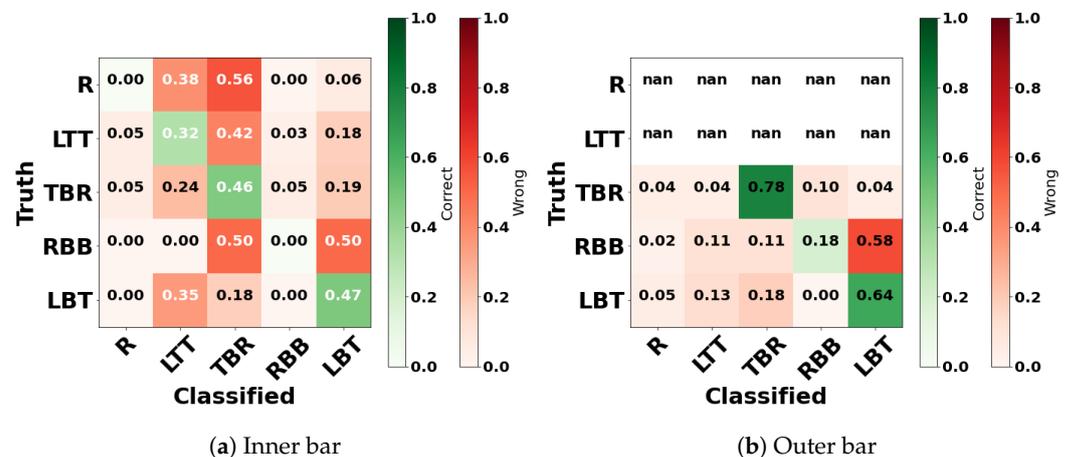


Figure 9. Normalized confusion matrices for the CombinedSINGLE-CNN tests at (a) Duck and (b) Narrabeen and (c) on the test data of Duck and Narrabeen combined.

When tested at Narrabeen (Figure 9b), the CombinedSINGLE-CNN achieved recall values ranging between 0.66 and 0.88. Note the high performance of the R state. At the transfer-test of the DUCK-CNN at Narrabeen, we noticed that the model failed to classify the R state, whereas the CombinedSINGLE-CNN resulted in a higher recall for the R state than the self-test of the NBN-CNN. In addition, we see high misclassifications for LBT, of which 28% is classified as LTT, and for RBB, of which 30% is classified as LBT.

When tested on the combined test datasets of Duck and Narrabeen (Figure 9c), the CombinedSINGLE-CNN achieved recall values ranging between 0.71 and 0.89. The most classification errors were consistent with the errors made when tested at Narrabeen, with 16% of the LTT images classified as R, 15% of the LBT images as LTT, and 14% of the RBB images classified as LBT.

To assess the transferability of single-barred models on unseen, double-barred data, the CombinedSINGLE-CNN was tested on the test data of the Gold Coast (Figure 10). For the inner bar (Figure 10a), recall values ranged between 0.00 and 0.47. The CombinedSINGLE-CNN especially failed to classify the R and RBB states. Rather, the images corresponding to the R state are most often classified as TBR and LTT states. On the other hand, images corresponding to the RBB state are classified as TBR, 50%, and LBT, 50%. Additionally, the LTT state is often classified as TBR, 42%, and as LBT, 18%. The TBR state is often classified as LTT, 0.24%, and as LBT, 19%. For the outer bar (Figure 10b), the recall values ranged between 0.18 and 0.78. The biggest confusion was made for the RBB state, of which 58% was classified as the LBT state. In addition, images were classified as R and LTT states, while the labels of the outer bar do not exhibit these states.



**Figure 10.** Normalized confusion matrices for the CombinedSINGLE-CNN tests at the test data of the Gold Coast. (a) The per-state performance on the inner bar. (b) The per-state performance on the outer bar.

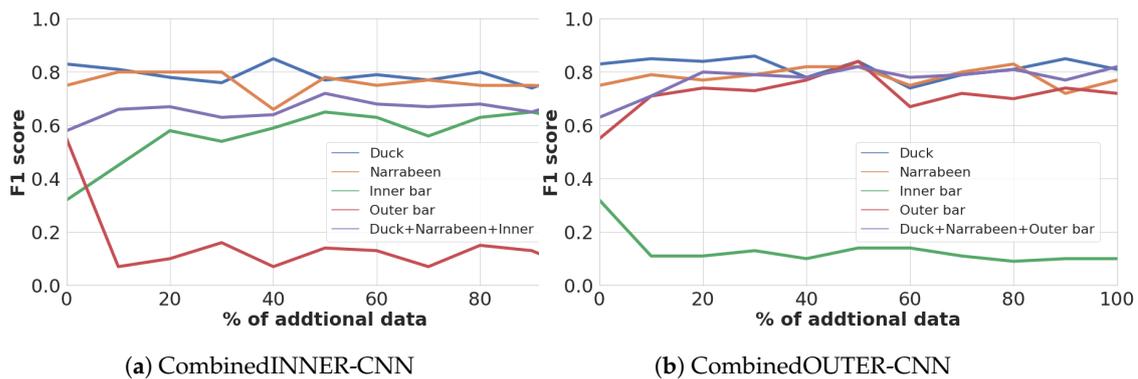
## 5.2. Double-Bar Beach Models Performances

Four CNNs were trained with the training data of the Gold Coast with either the inner or outer bar labels. The INNER-CNN, trained only with the Gold Coast inner bar labels, achieved an  $F1$  score of 0.73 on its self-test, which was a 128% increase as compared with CombinedSINGLE-CNN. The OUTER-CNN, trained with only the Gold Coast outer labels, achieved an  $F1$  score of 0.88, which is an increase of 60% as compared to the CombinedSINGLE-CNN.

Additionally, the CombinedINNER-CNN and CombinedOUTER-CNN were trained with data from the Gold Coast, added in increments of 10%. The  $F1$  values of both models, per increment of data added, are shown in Figure 11. The models' performances in terms of  $F1$  values and training and validation losses (Appendix A) varied with the different amounts of training data coming from the Gold Coast, with a maximum performance on the test data coming from all locations combined of  $F1$  scores of 0.72 and 0.82, for the CombinedINNER-CNN and CombinedOUTER-CNN, respectively. In both cases, these scores were achieved with 546 additional images from the Gold Coast. This corresponds to  $\frac{1}{3}$  of the total training dataset, with images coming from Duck, Narrabeen and the Gold Coast, resulting an equal amount of training images coming from each location.

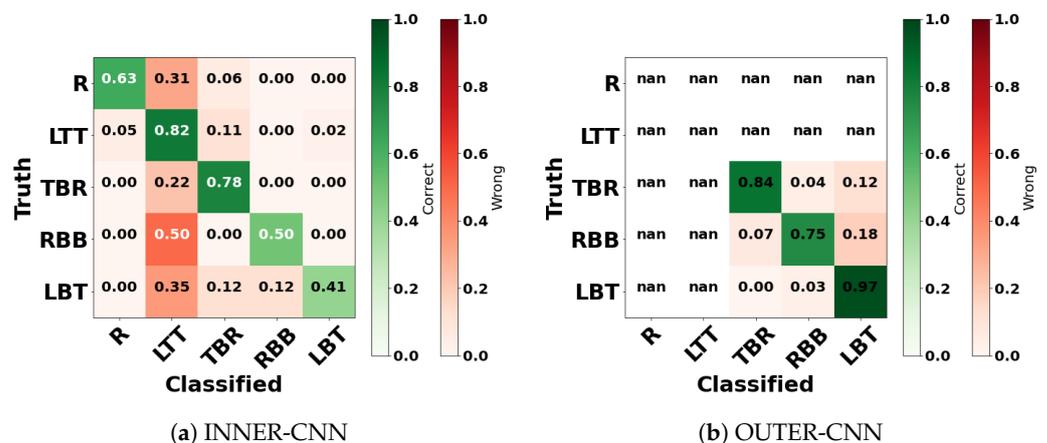
In Figure 11a, it can be seen that when trained with at least 20% of additional data with the inner bar labels,  $F1$  values within 20% of the  $F1$  score of the self-test of the INNER-CNN

(F1 of 0.73) were achieved. In Figure 11b, we see a similar case when trained with the outer bar labels. In this case only, 10% of additional data with the outer bar labels is required to achieve F1 scores within 20% of the self-test of the OUTER-CNN with an F1 score of 0.88. Furthermore, when data from the Gold Coast with the labels corresponding to the inner (outer) bar were added, the performance on the inner (outer) bar increased significantly. Simultaneously, the performance decreases substantially on the alternative bar (the location of which the labels were not included in the training). For example, when 10% of data were added from the outer bar to the CombinedOUTER-CNN, the F1 score increases at the outer bar from 0.58 to 0.71, while the F1 score at the inner bar decreases from 0.32 to 0.11.



**Figure 11.** F1 values of the CombinedINNER-CNN and CombinedOUTER-CNN with respect to percentage of data coming from the Gold Coast with labels coming from (a) the inner bar or (b) the outer bar and added to the combined training dataset of Duck and Narrabeen.

Figure 12 shows the confusion matrices corresponding to the per-state performance for the self-tests of the INNER-CNN and the OUTER-CNN. In the case of the INNER-CNN recall values ranging between 0.41 and 0.82 were achieved. There was a significant bias towards the LTT state as most of the confusions were made to this state. Specifically, 50% of the RBB, 31% of the R, 22% TBR and 35% of the LBT images were classified as LTT. The self-test of the OUTER-CNN resulted in recall values of 0.97 for LBT, 0.75 for RBB and 0.84 for TBR. The only significant confusion occurred in the LBT state, in which 18% of the RBB images and 12% of the TBR images were classified as LBT.



**Figure 12.** Normalized confusion matrices for the self-tests of (a) the INNER-CNN and (b) the OUTER-CNN.

Figures 13 and 14 show the confusion matrices corresponding to the tests of the CombinedINNER-CNN and CombinedOUTER-CNN, respectively, trained with 50% of additional data coming from the Gold Coast. The test of the CombinedINNER-CNN at the test data of all locations combined (Duck, Narrabeen, Gold Coast inner bar) reached

per-state recall values between 0.57 and 0.80, as shown in Figure 13a. The largest misclassifications were the 31% of the R and 26% of the LBT states classified as LTT and the 20% of RBB images classified as LBT. When the CombinedINNER-CNN was only tested at the Gold Coast inner bar, performance slightly decreased as recall values were slightly lower than for the combined beach test dataset (Figure 13b shows recall values between 0.41 and 0.76 for the Gold Coast inner bar dataset as compared with those between 0.57 and 0.80 for the combined dataset). In this case, almost all misclassifications happened in the LTT and TBR states.

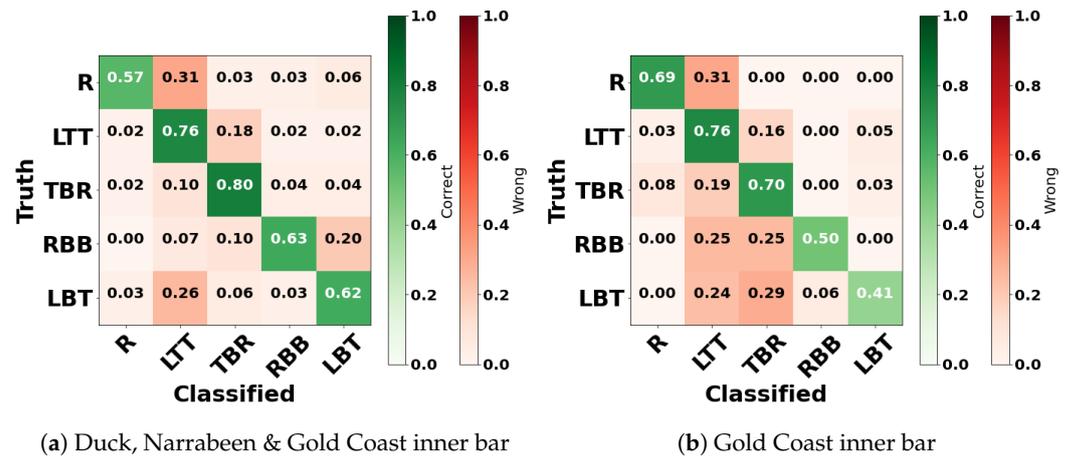


Figure 13. Normalized confusion matrices for the CombinedINNER-CNN tests: (a) the per-state performance of the self-test and (b) the per-state performance of the test at the data of the Gold Coast with the inner bar labels.

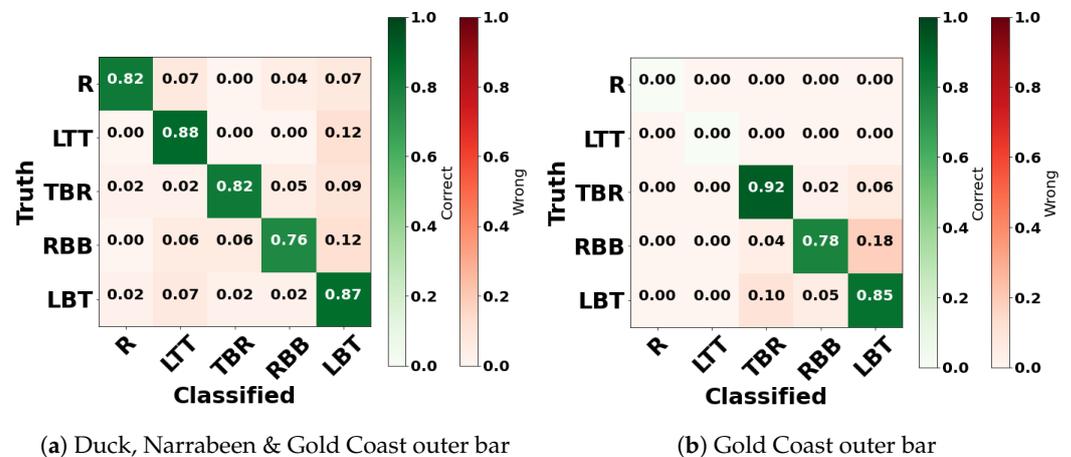


Figure 14. Normalized confusion matrices for the CombinedOUTER-CNN tests: (a) the per-state performance of the self-test and (b) the per-state performance of the test at the data of the Gold Coast with the outer bar labels.

The classifications of the test of the CombinedOUTER-CNN at the test data of all locations combined (Duck, Narrabeen, Gold Coast outer bar) resulted in recall values within the range of 0.76–0.88, as shown in Figure 14a. When the CombinedOUTER-CNN was tested at the outer bar of the Gold Coast, performance slightly increased as recall values were higher than for the combined beach test dataset. Testing the CombinedOUTER-CNN only at the Gold Coast with the outer bar labels (Figure 14b) gave recall values of 0.85 for LBT, 0.78 for RBB and 0.92 for TBR.

## 6. Discussion

### 6.1. Single-Bar Models

The misclassifications made by the CNNs trained with data of Duck and Narrabeen (Figures 7–9) in this study are consistent with [36]. Firstly, many misclassifications made by the CNNs can be attributed to states that are adjacent to one another in the classification scheme of [12]. Additionally, the misclassifications were made between the RBB and LBT states (states corresponding to an offshore sandbar with a distinct trough,) and between TBR and LTT states (states corresponding to bar welding and rip currents.) The fact that most misclassifications correspond to states with similar structural features that result in similar optical signatures indicates that the models have more difficulty classifying states with similar morphology.

As expected from the overall performances with *F1* values of 0.76 and 0.89 for the self-tests and 0.66 and 0.73 for the transfer-tests of NBN-CNN and DUCK-CNN, respectively, the number of misclassifications made at Narrabeen was typically higher than at Duck. This indicates that classifying beach states at Narrabeen is more complicated than at Duck. Ref. [36] stated that a possible explanation could be the difference in optical signatures of the shoreline. In Argus timex imagery, due to swash motions, Duck's shoreline is consistently identifiable by higher image intensities. In contrast, shoreline detection at Narrabeen happens in two ways: by higher image intensities associated with swash motions or lower intensities associated with wet, dark sand [66]. Due to the lack of a consistent optical signature of the shoreline at Narrabeen, the separation between shoreline and sandbar is less evident than at Duck. This could be one reason why the classification difficulty is enhanced.

When applied to the dataset of the Gold Coast, the CombinedSINGLE-CNN failed to accurately classify either the inner or outer bar in the double-barred system (Figure 10). We achieved *F1* values of 0.32 and 0.55 on the inner and outer bars, respectively. The per-state performance showed that for the inner bar, the images were primarily classified as LTT, TBR or LBT. For the outer bar, this was primarily the TBR and LBT states. The misclassifications at the inner bar may be due to difficulty in differentiating between the optical signature of the bar and the shoreline. At the Gold Coast, the Argus imagery consists only of low-tide images. In these low-tide images, the separation distance between the inner bar and the shoreline is more subtle compared to the separation distances in the single-barred imagery. When the models, trained with only the single-barred data were applied to identifying the inner bar, they were probably not sensitive enough to find those separation distances so it labelled most of the images into the shore-attached LTT and TBR states.

The lower performance of the CombinedSINGLE-CNN on the double-barred system of the Gold Coast as compared with the tests at Duck and Narrabeen (*F1* scores of 0.32 and 0.55 at the double-barred system compared with *F1* scores of 0.85 and 0.78 for the tests at Duck and Narrabeen) suggested that the algorithm required additional data from the Gold Coast to improve transferability performance.

### 6.2. Double-Bar Models

The *F1* values of 0.73 and 0.88 for the models trained with only the data of the Gold Coast with either the inner or outer bar labels, respectively, the INNER-CNN and OUTER-CNN, suggest that these models can distinguish between the inner and outer bar. We found that when the INNER-CNN was tested at the Gold Coast, it was biased towards labelling images as the LTT and TBR states (Figure 12a). In addition, the OUTER-CNN was biased towards the LBT state (Figure 12b). We found that these states occur more frequently in the training dataset. The training data of the inner bar were heavily imbalanced, with the frequencies of R/LTT/TBR/RBB/LBT being 13%/62%/18%/2%/5%. Additionally, the training data of the outer bar were moderately imbalanced, with the frequencies of TBR/RBB/LBT being 37%/23%/40%. This suggests that the biases towards these states were mainly caused by the imbalanced datasets (despite the effort to weigh

them more heavily during training using oversampling; see Section 3) rather than the confusion between similar looking morphologies.

In the second experiment, we tested the transferability of a model with data coming from both single and double-barred beach data. Notable is that when data from the Gold Coast were added, the performance on the bar from which the data were added increased up to recall values comparable to the self-tests of Duck and Narrabeen. On the other hand, the performance on the bar of which the data were not added decreased to an almost insignificant rate (Figure 11). Consistent with the results of the INNER-CNN and OUTER-CNN, this implies that the model can discriminate the inner and outer bar while training. Overall, the skill for the double-bar models (CombinedOUTER, CombinedINNER, INNER and OUTER) were highest for the outer bar labels, potentially indicating that the outer bar classification is ‘easier’ at the outer bar than the inner bar. This is consistent with our previous observations, showing that there are more difficulties to identify the inner bar compared to the outer bar.

Figure 11 showed that at least 20% of data coming from the Gold Coast with the inner bar labels or at least 10% of data with the outer bar labels was required for reasonable (within 20% of the self-test performances of the INNER-CNN and OUTER-CNN) transferability of the model to the new site. Moreover, we saw that for both cases, the best overall performance was achieved when an equal amount of training data came from all locations. Simultaneously, the performance on the single-bar datasets of Duck and Narrabeen remained quite similar compared to the performance of the CombinedSINGLE-CNN (with  $F1$  scores between 0.74–0.85 at Duck and  $F1$  scores between 0.66–0.83 at Narrabeen compared to the  $F1$  scores of 0.85 and 0.78 at Duck and Narrabeen, respectively). This indicates that increasing the diversity of images does not necessarily increase the overall skill at the original locations. Instead, it allows for classifications at an alternate location.

Depending on the dataset used for training, the bias towards certain beach states varied. For the CombinedSINGLE-CNN (trained on Narrabeen and Duck), there was a slight bias towards the LTT and TBR states, the INNER-CNN was biased towards the LTT state, and the CombinedINNER-CNN (trained on Narrabeen, Duck and Gold Coast inner datasets) was slightly biased towards the LTT and TBR states (Figure 13). This suggests that the inclusion of the Duck and Narrabeen datasets alleviated the bias of the INNER-CNN and spread the bias between the TBR and LTT states. In addition, the bias was less for the CombinedINNER-CNN than for the CombinedSINGLE-CNN, implying that the Duck and Narrabeen datasets contributed to the LTT/TBR bias when combined with the dataset of the Gold Coast with the inner bar labels. These results suggest that training with additional data from a new location does not cancel out confusions made by models the models without the data from a new location. Instead, these errors become relatively small, and therefore, the  $F1$  score becomes overall better as the total amount of test data from other locations increases.

In the cases where the models were trained with additional data from the Gold Coast, we stopped adding data when the training data consisting of data from Duck and Narrabeen doubled. Hence, only 36% of the total amount of the Gold Coast’s available training data was used to train the models. In most classification tasks, training the model on more data should improve the model’s performance [67]. However, in the case of using all three datasets, higher amounts of data coming from one location did not always positively affect the model’s performance. Reasons for this may be the site-specific features, such as tidal range, number of cameras, and wave climate. A majority of images coming from one specific site could result in the model training on such a specific feature. This would result in the fact that wrong features will be correlated with specific labels, thus decreasing the performance on the other sites.

In future research, more data from different locations would benefit the model’s performance at new locations and would be a step in the right direction for creating a universally applicable model for classifying beach states worldwide. Additionally, to classify beach states more accurately, the application of CNNs in object detection and localization tasks

could enable a CNN to localize multiple beach states alongshore and cross-shore [68]. Moreover, object tracking could enable a CNN to track sandbar evolution [69].

## 7. Conclusions

With our work, we extend the previous work by [36], on the classification of beach states in single-barred systems using a convolutional neural network, to the classification of beach states of double-barred beaches. The main findings of our work are that (1) a CNN trained with images from single-barred beaches shows poor performance when classifying double-barred beach states; (2) transfer learning, where limited data from a double-barred beach is added to the single-barred model, allows for the training of a well-performing model for classifying double-barred beach states; and (3) including outer-bar labels in the transfer learning has a larger impact on the resulting model performance than when labels from the inner bar are included.

Rather than training the model from scratch as in [36], we used transfer learning to fine-tune the pre-trained ResNet50 network. Datasets from three different beaches were used: the single-beach images from Duck, North Carolina, USA and Narrabeen-Collaroy, NSW, Australia, and double-bar beach images from the Gold Coast, Australia. The fact that we achieved comparable performance to [36] shows that the features learned from images within the ImageNet dataset can be applied to classify coastal imagery. Additionally, our work shows that we can classify both single and double-barred beaches in an automated way using a CNN. Hence, this is a step forward to better automatic beach state classification.

We conducted two experiments in which we trained and assessed seven different CNNs. Each model was tested on the test data from the locations where its training data came from, the *self-tests*, and on the test data of alternate, unseen locations, the *transfer-tests*. Three models were trained with only the single-barred data: one at Duck, one at Narrabeen, and one with data from both Duck and Narrabeen. The model trained on the data from both Narrabeen and Duck combined achieved  $F1$  values of 0.78 at Narrabeen, 0.85 at Duck and 0.78 for the combined test dataset. When the model trained on the data of both Duck and Narrabeen was tested on the Gold Coast, we achieved poor performance for both bars ( $F1$  score at inner bar = 0.32 and outer bar = 0.55). Consistent with [36], this suggested that the algorithm requires additional data from a new location to improve the transferability performance.

In a second experiment, additional data from the Gold Coast were used for training. Firstly, two models were trained with only the double-barred beach data. The skills of the self-tests of these models were comparable to the skills in the self-tests of the single-barred models of Duck and Narrabeen, with  $F1$  values of 0.73 and 0.88 for the models trained with either inner or outer bar labels, respectively.

Additionally, two models with data coming from Duck and Narrabeen and with incrementally added data of the Gold Coast, with either the inner or outer bar labels, were trained. The tests with these models showed that, with at least 20% of data with the inner bar labels or 10% of data with the outer bar labels,  $F1$  values within 20% of the  $F1$  scores of the self-tests of the INNER-CNN and OUTER-CNN (0.73 and 0.88) were achieved. Moreover, with an equal amount of the total training data coming from each location,  $F1$  values comparable with the self-test cases at each location were achieved. On the single-bar data, the CombinedINNER-CNN and CombinedOUTER-CNN achieved  $F1$  values at Narrabeen and Duck of, respectively, 0.80 and 0.85 for the model with labels from the inner bar and  $F1$  values of 0.82 and 0.86 for the model with the outer bar labels. On the double-barred data, the CombinedINNER-CNN and CombinedOUTER-CNN achieved  $F1$  values of 0.65 and 0.84 for the inner and outer bars, respectively.

For the models trained with data of the Gold Coast it mattered which of the labels were used for training; training with outer bar labels led to overall higher performances, with the exception of classifying the inner bar, on which the performance was higher when using the inner bar labels for training. Additionally, more data from one location did not

always positively affect the model's performance. However, the larger diversity of images allowed the transferability to more locations.

**Author Contributions:** Conceptualization, S.C.M.O., W.N., T.D.P. and A.N.E.; methodology, S.C.M.O. and A.N.E.; software, S.C.M.O. and A.N.E.; writing—original draft preparation, S.C.M.O.; writing—review and editing, W.N., T.D.P. and A.N.E.; supervision, W.N. and T.D.P.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding. We thank the Utrecht University Open Access Fund for reimbursing the article processing charge of this open access publication.

**Data Availability Statement:** Trained models are available at <https://github.com/StanOerlemans/DeepBeachStateV2>, accessed on 14 January 2022, and images upon request.

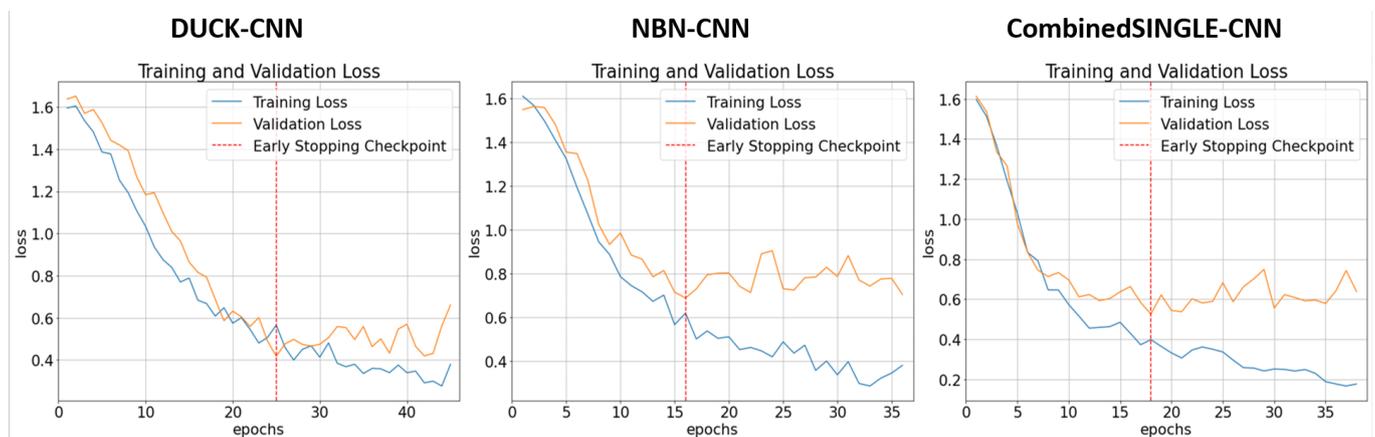
**Acknowledgments:** We thank Ton Markus from Communications & Marketing at the Faculty of Geosciences, Utrecht University for producing Figures 1 and 2. We also thank the anonymous referees for their valuable comments that improved the quality of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Performance in Terms of Loss

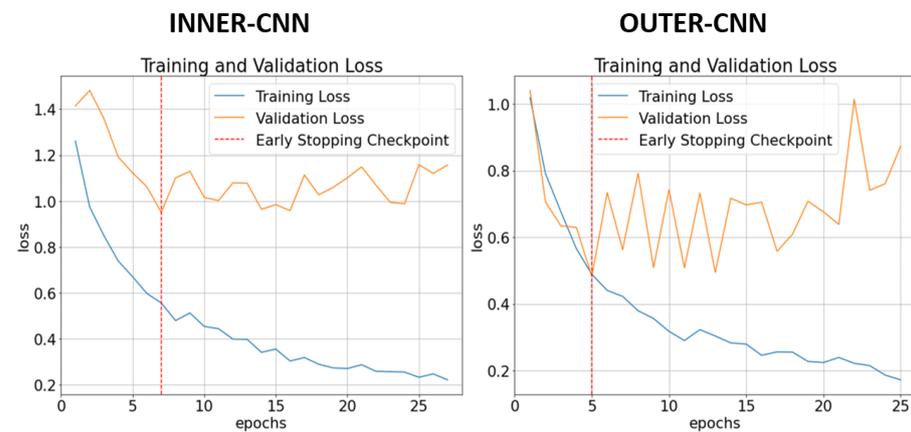
To assess the model's performance during training, the measure known as loss was used. The loss quantifies the error produced by the model. We used the training loss as a metric to assess the model's error on the training data. The validation loss was used to assess the model's error on the validation data. For each CNN, we visualized both the training and validation loss. Preferably, both losses are close together and decrease at the same rate. However, when they diverge the model starts to overfit, meaning that the model performs better on the data it 'knows' than on unseen data.

### Appendix A.1. Single-Bar Models

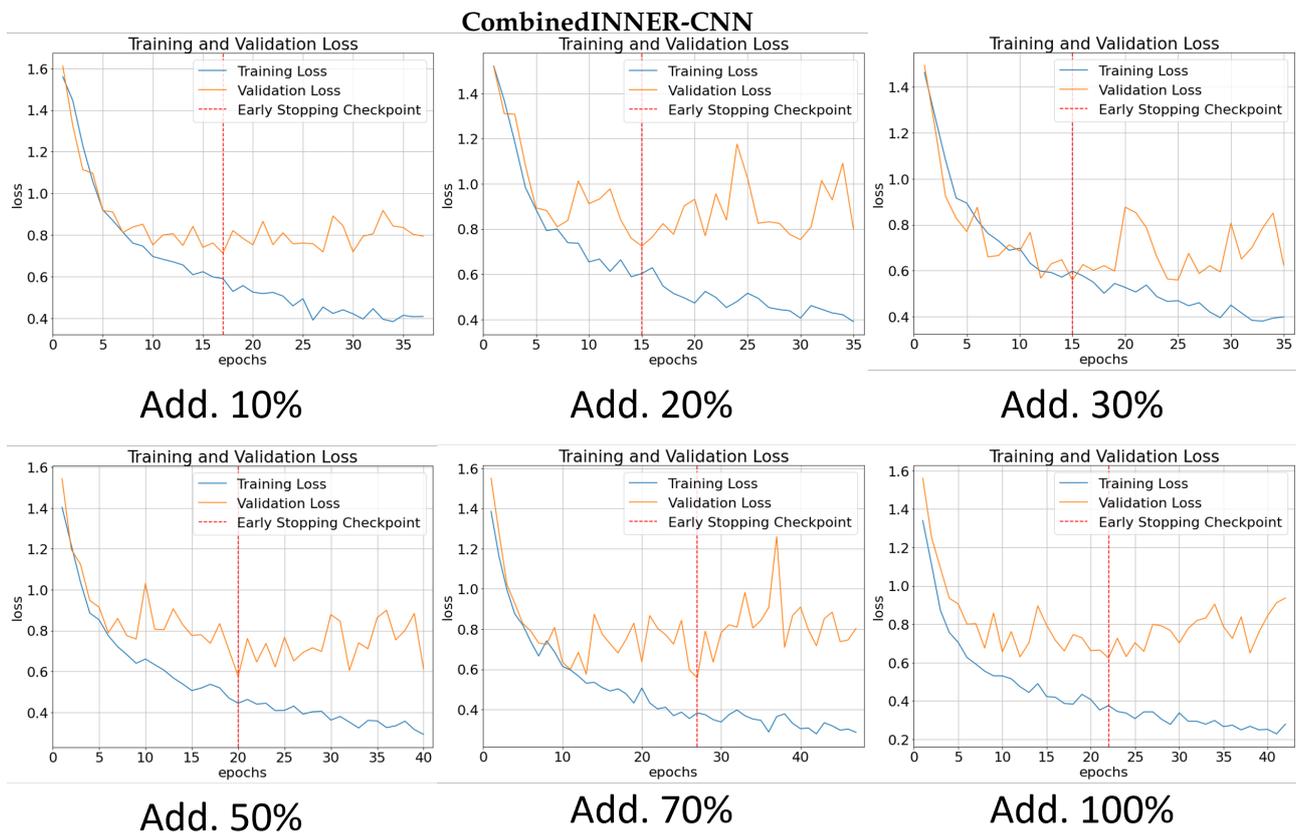


**Figure A1.** The training and validation losses of training the models: DUCK-CNN (left), NBN-CNN (middle) and CombinedSINGLE-CNN (right). On the  $y$ -axis is the loss, and on the  $x$ -axis is the number of epochs. The red dotted line indicates the early stopping point.

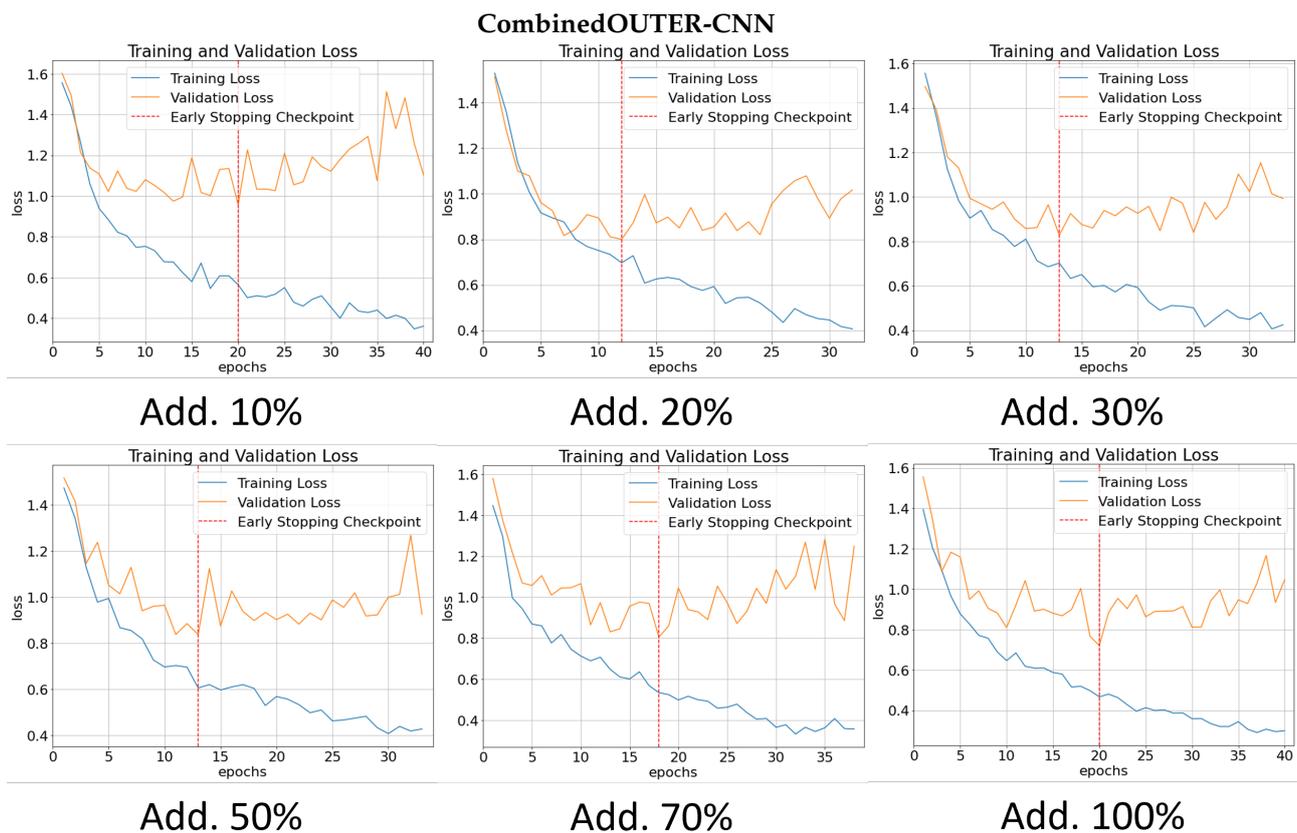
Appendix A.2. Double-Bar Models



**Figure A2.** The training and validation losses of training the models: INNER-CNN (left) and OUTER-CNN (right). On the *y*-axis the loss and on the *x*-axis the number of epochs. The red dotted line indicates the early stopping point.



**Figure A3.** The training and validation losses of the models fed with increments of 10% of additional double-barred beach data with the inner bar labels.



**Figure A4.** The training and validation losses of the models fed with increments of 10% of additional double-barred beach data with the outer bar labels.

## References

- Plant, N.; Holman, R.; Freilich, M.; Birkemeier, W. A simple model for interannual sandbar behavior. *J. Geophys. Res. Ocean.* **1999**, *104*, 15755–15776. [[CrossRef](#)]
- Walstra, D.J.R.; Wesselman, D.A.; Van der Deijl, E.C.; Ruessink, G. On the intersite variability in inter-annual nearshore sandbar cycles. *J. Mar. Sci. Eng.* **2016**, *4*, 15. [[CrossRef](#)]
- Alexander, P.S.; Holman, R.A. Quantification of nearshore morphology based on video imaging. *Mar. Geol.* **2004**, *208*, 101–111. [[CrossRef](#)]
- Holman, R.A.; Symonds, G.; Thornton, E.B.; Ranasinghe, R. Rip spacing and persistence on an embayed beach. *J. Geophys. Res. Ocean.* **2006**, *111*, C01006. [[CrossRef](#)]
- Tătui, F.; Constantin, S. Nearshore sandbars crest position dynamics analysed based on Earth Observation data. *Remote Sens. Environ.* **2020**, *237*, 111555. [[CrossRef](#)]
- Castelle, B.; Marieu, V.; Bujan, S.; Splinter, K.D.; Robinet, A.; Sénéchal, N.; Ferreira, S. Impact of the winter 2013–2014 series of severe Western Europe storms on a double-barred sandy coast: Beach and dune erosion and megacusp embayments. *Geomorphology* **2015**, *238*, 135–148. [[CrossRef](#)]
- Almar, R.; Castelle, B.; Ruessink, B.; Sénéchal, N.; Bonneton, P.; Marieu, V. Two- and three-dimensional double-sandbar system behaviour under intense wave forcing and a meso–macro tidal range. *Cont. Shelf Res.* **2010**, *30*, 781–792. [[CrossRef](#)]
- Grant, S.B.; Kim, J.H.; Jones, B.H.; Jenkins, S.A.; Wasy, J.; Cudaback, C. Surf zone entrainment, along-shore transport, and human health implications of pollution from tidal outlets. *J. Geophys. Res. Ocean.* **2005**, *110*, C10025. [[CrossRef](#)]
- Castelle, B.; Scott, T.; Brander, R.; McCarroll, R. Rip current types, circulation and hazard. *Earth-Sci. Rev.* **2016**, *163*, 1–21. [[CrossRef](#)]
- Short, A.D.; Aagaard, T. Single and multi-bar beach change models. *J. Coast. Res.* **1993**, *15*, 141–157.
- Price, T.; Ruessink, B. State dynamics of a double sandbar system. *Cont. Shelf Res.* **2011**, *31*, 659–674. [[CrossRef](#)]
- Wright, L.D.; Short, A.D. Morphodynamic variability of surf zones and beaches: A synthesis. *Mar. Geol.* **1984**, *56*, 93–118. [[CrossRef](#)]
- Van Enckevort, I.; Ruessink, B.; Coco, G.; Suzuki, K.; Turner, I.; Plant, N.G.; Holman, R.A. Observations of nearshore crescentic sandbars. *J. Geophys. Res. Ocean.* **2004**, *109*, C06028 [[CrossRef](#)]
- Abessolo Ondo, G. Response of Sandy Beaches in West Africa, Gulf of Guinea, to Multi-Scale Forcing. Ph.D. Thesis, Université Paul Sabatier—Toulouse III, Toulouse, France, 2020.

15. Gallagher, E.L.; Elgar, S.; Guza, R.T. Observations of sand bar evolution on a natural beach. *J. Geophys. Res. Ocean.* **1998**, *103*, 3203–3215. [[CrossRef](#)]
16. Holman, R.A.; Stanley, J. The history and technical capabilities of Argus. *Coast. Eng.* **2007**, *54*, 477–491. [[CrossRef](#)]
17. Holman, R.; Haller, M.C. Remote sensing of the nearshore. *Annu. Rev. Mar. Sci.* **2013**, *5*, 95–113. [[CrossRef](#)]
18. Aarninkhof, S.; Ruessink, G. Quantification of surf zone bathymetry from video observations of wave breaking. In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, CA, USA, 6–10 December 2002; Volume 2002, p. OS52E-10.
19. Aleman, N.; Certain, R.; Robin, N.; Barousseau, J.P. Morphodynamics of slightly oblique nearshore bars and their relationship with the cycle of net offshore migration. *Mar. Geol.* **2017**, *392*, 41–52. [[CrossRef](#)]
20. Palmsten, M.L.; Brodie, K.L. The Coastal Imaging Research Network (CIRN). *Remote Sens.* **2022**, *14*, 453. [[CrossRef](#)]
21. Jackson, D.W.; Short, A.D.; Loureiro, C.; Cooper, J.A.G. Beach morphodynamic classification using high-resolution nearshore bathymetry and process-based wave modelling. *Estuar. Coast. Shelf Sci.* **2022**, *268*, 107812. [[CrossRef](#)]
22. Janowski, L.; Wroblewski, R.; Rucinska, M.; Kubowicz-Grajewska, A.; Tysiac, P. Automatic classification and mapping of the seabed using airborne LiDAR bathymetry. *Eng. Geol.* **2022**, *301*, 106615. [[CrossRef](#)]
23. Pape, L.; Ruessink, B. Neural-network predictability experiments for nearshore sandbar migration. *Cont. Shelf Res.* **2011**, *31*, 1033–1042. [[CrossRef](#)]
24. Kingston, K.; Ruessink, B.; Van Enckevort, I.; Davidson, M. Artificial neural network correction of remotely sensed sandbar location. *Mar. Geol.* **2000**, *169*, 137–160. [[CrossRef](#)]
25. Pape, L.; Ruessink, B.G.; Wiering, M.A.; Turner, I.L. Recurrent neural network modeling of nearshore sandbar behavior. *Neural Netw.* **2007**, *20*, 509–518. [[CrossRef](#)]
26. Collins, A.M.; Geheran, M.P.; Hesser, T.J.; Bak, A.S.; Brodie, K.L.; Farthing, M.W. Development of a Fully Convolutional Neural Network to Derive Surf-Zone Bathymetry from Close-Range Imagery of Waves in Duck, NC. *Remote Sens.* **2021**, *13*, 4907. [[CrossRef](#)]
27. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)]
28. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
29. Yu, J.; Wang, Z.; Vasudevan, V.; Yeung, L.; Seyedhosseini, M.; Wu, Y. Coca: Contrastive captioners are image-text foundation models. *arXiv* **2022**, arXiv:2205.01917.
30. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [[CrossRef](#)]
31. Hoshen, Y.; Weiss, R.J.; Wilson, K.W. Speech acoustic modeling from raw multichannel waveforms. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, Australia, 19–24 April 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 4624–4628.
32. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
33. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. *arXiv* **2015**, arXiv:1804.06655.
34. Niu, S.; Liu, Y.; Wang, J.; Song, H. A decade survey of transfer learning (2010–2020). *IEEE Trans. Artif. Intell.* **2020**, *1*, 151–166. [[CrossRef](#)]
35. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land use classification in remote sensing images by convolutional neural networks. *arXiv* **2015**, arXiv:1508.00092.
36. Ellenson, A.N.; Simmons, J.A.; Wilson, G.W.; Hesser, T.J.; Splinter, K.D. Beach State Recognition Using Argus Imagery and Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 3953. [[CrossRef](#)]
37. Birkemeier, W.A.; DeWall, A.E.; Gorbics, C.S.; Miller, H.C. *A User's Guide to CERC's Field Research Facility*; Technical Report; Coastal Engineering Research Center: Fort Belvoir VA, USA, 1981.
38. Lee, G.h.; Nicholls, R.J.; Birkemeier, W.A. Storm-driven variability of the beach-nearshore profile at Duck, North Carolina, USA, 1981–1991. *Mar. Geol.* **1998**, *148*, 163–177. [[CrossRef](#)]
39. Stauble, D.K.; Cialone, M.A. Sediment dynamics and profile interactions: Duck94. In Proceedings of the Coastal Engineering 1996, Orlando, FL, USA, 2–6 September 1996; pp. 3921–3934.
40. Turner, I.L.; Harley, M.D.; Short, A.D.; Simmons, J.A.; Bracs, M.A.; Phillips, M.S.; Splinter, K.D. A multi-decade dataset of monthly beach profile surveys and inshore wave forcing at Narrabeen, Australia. *Sci. Data* **2016**, *3*, 160024. [[CrossRef](#)]
41. Splinter, K.D.; Harley, M.D.; Turner, I.L. Remote sensing is changing our view of the coast: Insights from 40 years of monitoring at Narrabeen-Collaroy, Australia. *Remote Sens.* **2018**, *10*, 1744. [[CrossRef](#)]
42. Harley, M.D.; Turner, I.; Short, A.; Ranasinghe, R. A reevaluation of coastal embayment rotation: The dominance of cross-shore versus alongshore sediment transport processes, Collaroy-Narrabeen Beach, southeast Australia. *J. Geophys. Res. Earth Surf.* **2011**, *116*, F04033. [[CrossRef](#)]
43. Harley, M. Coastal storm definition. In *Coastal Storms: Processes and Impacts*; Wiley–Blackwell: Hoboken, NJ, USA, 2017 pp. 1–21.
44. Short, A.D.; Trenaman, N. Wave climate of the Sydney region, an energetic and highly variable ocean wave regime. *Mar. Freshw. Res.* **1992**, *43*, 765–791. [[CrossRef](#)]

45. Strauss, D.; Mirferendesk, H.; Tomlinson, R. Comparison of two wave models for Gold Coast, Australia. *J. Coast. Res.* **2007**, *50*, 312–316.
46. Allen, M.; Callaghan, J. *Extreme Wave Conditions for the South Queensland Coastal Region*; Environmental Protection Agency: Brisbane, Australia, 2000.
47. Jackson, L.A.; Tomlinson, R.; Nature, P. 50 years of seawall and nourishment strategy evolution on the gold coast. In Proceedings of the Australasian Coasts & Ports conference, Cairns, Australia, 21–23 June, 2017; pp. 640–645.
48. Jackson, L.A.; Tomlinson, R.; Turner, I.; Corbett, B.; d’Agata, M.; McGrath, J. Narrowneck artificial reef; results of 4 yrs monitoring and modifications. In Proceedings of the 4th International Surfing Reef Symposium, Manhattan Beach, CA, USA, 12–14 January 2005.
49. Turner, I.L.; Whyte, D.; Ruessink, B.; Ranasinghe, R. Observations of rip spacing, persistence and mobility at a long, straight coastline. *Mar. Geol.* **2007**, *236*, 209–221. [[CrossRef](#)]
50. Khan, A.; Sohail, A.; Zahoor, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. *Artif. Intell. Rev.* **2020**, *53*, 5455–5516. [[CrossRef](#)]
51. Shrestha, A.; Mahmood, A. Review of deep learning algorithms and architectures. *IEEE Access* **2019**, *7*, 53040–53065. [[CrossRef](#)]
52. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
53. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
54. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, pp. 2818–2826.
55. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, pp. 770–778.
56. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional neural networks: An overview and application in radiology. *Insights Imaging* **2018**, *9*, 611–629. [[CrossRef](#)] [[PubMed](#)]
57. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
58. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [[CrossRef](#)]
59. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A survey on deep transfer learning. In Proceedings of the International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Springer: New York, NY, USA, 2018; pp. 270–279.
60. Smith, L.N.; Topin, N. Super-convergence: Very fast training of neural networks using large learning rates. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications, Baltimore, MD, USA, 15–17 April 2019; International Society for Optics and Photonics: Bellingham, WA, USA, 2019; Volume 11006, p. 1100612.
61. Zhang, T.; Yu, B. Boosting with early stopping: Convergence and consistency. *Ann. Stat.* **2005**, *33*, 1538–1579. [[CrossRef](#)]
62. Yang, J.; Yang, G. Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer. *Algorithms* **2018**, *11*, 28. [[CrossRef](#)]
63. Menardi, G.; Torelli, N. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Discov.* **2014**, *28*, 92–122. [[CrossRef](#)]
64. Cubuk, E.D.; Zoph, B.; Mane, D.; Vasudevan, V.; Le, Q.V. Autoaugment: Learning augmentation strategies from data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–19 June 2019, pp. 113–123.
65. Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1.
66. Pianca, C.; Holman, R.; Siegle, E. Shoreline variability from days to decades: Results of long-term video imaging. *J. Geophys. Res. Ocean.* **2015**, *120*, 2159–2178. [[CrossRef](#)]
67. Lei, S.; Zhang, H.; Wang, K.; Su, Z. How Training Data Affect the Accuracy and Robustness of Neural Networks for Image Classification. In Proceedings of the International Conference on Learning Representations (ICLR 2019), New Orleans, LA, USA, 6–9 May 2019.
68. Jiang, B.; Luo, R.; Mao, J.; Xiao, T.; Jiang, Y. Acquisition of localization confidence for accurate object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 784–799.
69. Bashir, F.; Porikli, F. Performance evaluation of object detection and tracking systems. In Proceedings of the 9th IEEE International Workshop on PETS, New York, NY, USA, 18 June 2006; pp. 7–14.