GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Identifying open and closed stratocumulus clouds by their reflectance using a convolutional neural network

A thesis submitted in fulfilment of the requirements for the academic degree of
Master of Science (M.Sc.)
in
Meteorology

Johann Wolfgang Goethe-University, Frankfurt am Main
Institute for Atmospheric and Environmental Sciences

prepared by
Christopher Reichel
born on 11 April 1994 in Limburg a. d. Lahn

Supervisor: Dr. Anna Possner
Co-supervisor: Prof. Dr. Bodo Ahrens

Submitted on 11 April 2022

Statutory Declaration

I herewith declare that I have composed the present thesis myself and without use of any other than the cited sources and aids. Sentences or parts of sentences quoted literally are marked as such; other references with regard to the statement and scope are indicated by full details of the publications concerned. The thesis in the same or similar form has not been submitted to any examination body and has not been published. This thesis was not yet, even in part, used in another examination or as a course performance.

Frankfurt am Main, 11 April 2022 .....

Abstract

Stratocumulus clouds cover vast areas of the ocean and thus have a high impact on Earth's climate due to their ability to reflect solar shortwave radiation. Past studies show that different morphologies of stratocumulus clouds reflect solar radiation differently. For example, closed mesoscale cellular convective (MCC) clouds have been found to reflect more solar radiation than open MCC clouds, even though both MCC morphologies have the same cloud fraction, and therefore a closer investigation of these morphologies is necessary.

In order to achieve this, during this thesis two years of Aqua satellite observations are used to train a convolutional neural network (CNN) so that it is able to recognize open, closed, and no MCC clouds by itself.

The data available for the training is augmented using two augmentation methods and then run through six different CNN model configurations to see how the configurations affect the two different augmented datasets. The data available for the training is augmented using two augmentation methods and then run through six different CNN model configurations to see how the models are affected by the two different augmented datasets. The first augmentation method is already effective with a simple model configuration and shows no further improvement when used with extended configurations, whereas the second augmentation method only delivers the desired result with more complex configurations.

The most promising models reach an overall accuracy of 80 % and even 88 % for the open and closed MCC categories. Only the classification of the no MCC scenes lacks accuracy, due to the complexity of some no MCC images, which partly have open and closed MCC clouds structures.

Contents

List of Figures

List of Tables

List of Abbreviations

CNN	Convolutional Neural Network
DJF	December/January/February
FN	False Negative
FP	False Positive
JJA	June/July/August
LC	Layer Combination
MCC	Mesoscale Cellular Convection
MBL	Marine Boundary Layer
MCAO	Marine Cold Air Outbreak
MODIS	Moderate Resolution Imaging Spectroradiometer
NA	North Atlantic
NEP	North East Pacific
NN	Neural Network
PDF	Probability Densitiy Function
POC	Pocket of Open Cells
SEI	South East Indian
SEP	South East Pacific
SST	Sea Surface Temperature
SP	South Pacific
TN	True Negative
TP	True Positive

1 Motivation

Marine boundary layer clouds, especially stratocumulus clouds, are very important for Earths climate. They have a high radiative impact, which leads to a strong cooling effect, especially for solar radiation (Muhlbauer et al., 2014; McCoy et al., 2017). Stratocumulus clouds are especially likely to occur in the upper region of the shallow planterar boundary layer (Wood, 2012). In the layer around stratocumulus clouds, important key processes prevail that strongly influence their formation and course (Klein and Hartmann, 1993; Wood, 2012).

These marine low-level clouds self-organize into various mesoscale types, for example into open, closed or cellular but disorganized mesoscale cellular convective (MCC) clouds, which have different radiative properties (McCoy et al., 2017) and underlying physical processes (Cahalan et al., 1994; Stevens et al., 2005).

Global climate models have large uncertainties in model predictions for these clouds, because they do not resolve their physical processes sufficiently (Stocker, 2014; Wood and Hartmann, 2006). However, these types of low level clouds are very important, because they highly influence the radiative impact and the climate feedback. (Wood and Hartmann, 2006). Open MCC, for example, reflect significantly less shortwave radiation than closed MCC for the same cloud cover (McCoy et al., 2017). This means that the radiation contribution differs depending on their organizational regime . However, in order to statistically explore the relationships between cloud morphology and their radiative properties one has to be able to objectively quantify different organisational regimes in an automated manner.

To better investigate these morphological types, they need to be classified first. Usually experienced experts are needed to classify them manually, but this method is time-expensive and the classification depends strongly on the experience of the operator and can therefore be subjective or biased. This means, there is a high demand for a cloud classification method, which is automatic and precise (Zhang et al., 2018).

Wood and Hartmann (2006) for example trained a neural network (NN) to classify low-level cloud scenes from Moderate Resolution Imaging Spectroradiometer (MODIS)

satellite data by their morphology. As input data they used probability densitiy functions (PDFs) and power spectra analysis of the liquid water path (LWP). They classified each cloud scene into four morphological categories: stratiform, open MCC, closed MCC, and disorganized MCC. Although the method is good for liquid phase clouds, LWP retrievals in mixed-phase clouds are less reliable and thus other quantities may need to be used. Therefore, they effect the radiative energy balance, cloud electrification, precipitation formation, as well as the general life cycle of clouds on a regional and global scale (Korolev et al., 2017). Hence, input data which includes those type of clouds would be beneficial, for example the reflectance.

In another work, by Yuan et al. (2020), a training dataset was first created by having six different people classify thousands of scenes of satellite images into six different categories. These categories are: stratus, closed cellular convection, disorganized convection, open cellular convection, clustered cumulus convection and suppressed cumulus convection. Then, these data were used to train a deep convolutional neural network model and subsequently tested using a test dataset. An average cross-type precision of about 93 % was achieved for the trained model. Furthermore, a pre-trained model was used in the work to achieve better results. Possibly by using other satellite data or applying a slightly modified model, the precision could also be increased (Yuan et al., 2020).

For my thesis I introduce a convolutional neural network (CNN) to classify marine low clouds by their morphology type. 224 x 224 pixels. For this purpose, several thousand scenesfrom satellite data with a resolution of 224x224 pixels each are used and divided into four different categories using five classifiers. The resulting dataset is then divided into a training and test dataset. After the CNN has been trained with the training set, the accuracy will be checked with the test dataset. The goal is to create an automated mechanism for cloud classification to detect open, closed and disorganized mesoscale cellular convective clouds by their reflectance.

This thesis is divided into six chapters. In chapter two, the theoretical background is explained in order to develop an understanding of MCC clouds and neural networks. In the following third chapter, the methods for data acquisition, as well as the basic network architecture are explained. Chapter four presents the results of this thesis, which are then discussed in chapter five. Finally, chapter 6 summarizes the main findings of this thesis.

2 Theoretical background

2.1 Clouds

Clouds are formed by the condensation of water vapor on condensation nuclei in the atmosphere (Klose, 2015). A precondition for condensation is that the air is saturated with water vapor. In other words, the saturation vapor pressure must be equal to the partial pressure (Malberg, 2007). This point is reached by adiabatic cooling of rising warm air (Malberg, 2007; Klose, 2015). The cooling of the warm air packets occurs mainly through two processes. First, by turbulent lifting of air parcels, which is called **free convection** and second, by large-scale sliding of air, also named by **forced convection** (Schönwiese, 2013).

Free convection: An air movement in which a particle constantly changes its speed and direction is called turbulent. Due to this turbulence, air parcels move upward and cool adiabatically. In the case of free convection, for example, an static unstable stratification exists due to heating of the ground. There, the temperature of the air parcel is higher than the surrounding temperature and is thus forced to rise. At some point, this air parcel reaches the condensation level and water droplets begin to form. The air parcel continues to rise and condense until it reaches equilibrium with the surrounding temperature. The clouds that result from this process are mostly cumulus clouds. (Schönwiese, 2013).

Forced convection: Here, cloud formation occurs due to forced lifting of air masses in a static stable stratification. In this case, the air mass is colder than the surounding temperature and would therefore not rise by itself. If an air parcel hits a mountain, for example, it is forced to rise. Due to this forced convection, the air cools down. If the air is saturated, clouds are formed by the condensation of water droplets. Such forced uplift can also occur when warm air glides over cold air during warm fronts. The resulting clouds are stratiform clouds such as stratocumulus clouds (Schönwiese, 2013).

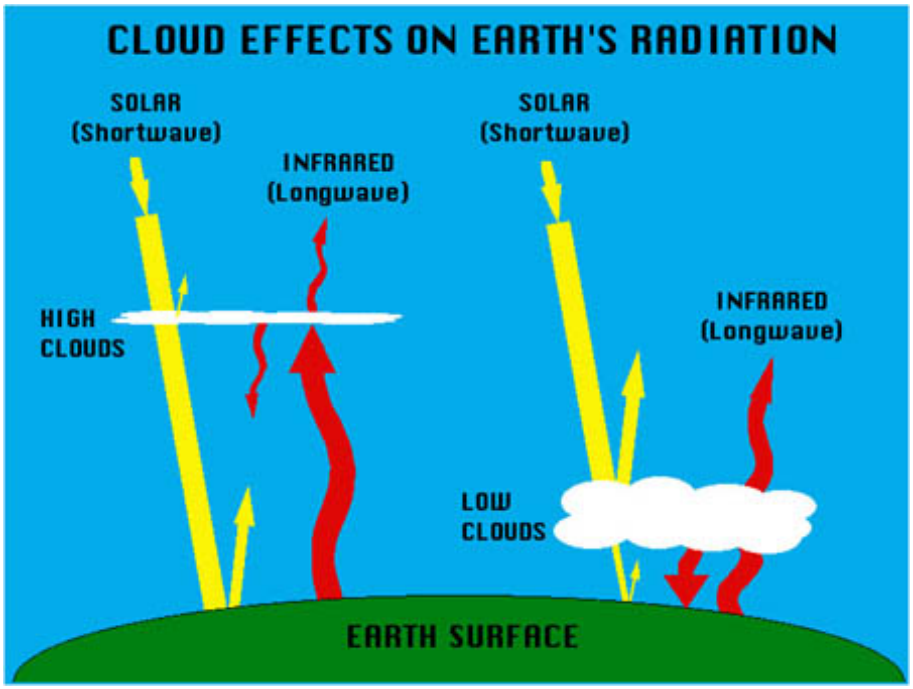


Figure 2.1: Illustration of the radiative properties of high and low clouds on the basis of short and long wave radiation (`https://www.nasa.gov/audience/forstudents/5-8/features/F_The_Role_of_Clouds.html`, accessed on 14.03.2022).

Clouds are extremely important for Earth's climate balance (Stocker et al., 2013). Due to their ability to reflect and absorb radiation, they have a cooling as well as a warming effect on the atmosphere. The cooling effect occurs due to the reflection of shortwave solar radiation and depends mainly on the optical thickness, as well as the particle size and phase. A warming effect is caused by absorbing and emitting longwave radiation, which in turn depends largely on the cloud top temperature (Arking, 1991). Depending on the type of clouds, they have a different effect on the radiation budget of Earth, as can be seen in Figure 2.1. High clouds have a cloud top height of about 5 km (Kraus, 2004) or higher and are optically thinner and let more shortwave radiation through, whereas the amount of absorbed longwave radiation is larger compared to reflected shortwave radiation. Thus, they are more likely to contribute to warming. Low clouds, on the other hand, are located at altitudes below 2 km (Kraus, 2004) and are optically very thick (Hahn et al., 2001) and therefore have a high reflective property with respect to solar radiation. They also absorb and emit in the long-wave range to a not too small extent, but this is significantly smaller (Chen et al., 2000). However, globally the cooling effect of low clouds dominate by around -20 $\mathrm{W\,m^{-2}}$ (Loeb et al., 2009). As this work deals with mesoscale forms of stratocumulus clouds, this low cloud type is discussed in more detail below.

2.1.1 Stratocumulus clouds

Because of the fact, that they have the highest occurrence of all cloud types, they cover large parts of Earth and are largely distributed over the ocean due to the favorable conditions (Wood, 2012). Consequently, marine stratocumulus clouds have a great influence on the climate due to their frequency and their radiative characteristic of low clouds (Hartmann et al., 1992; Chen et al., 2000; Wood, 2012). The illustration of the mesoscale forms will be explained in more detail in section 2.1.2. First, the general radiative, climatological and micro/macro physical properties of stratocumulus clouds will be discussed in more detail.

Climatology, radiative properties and physical processes

As shown in the Figure 2.2, stratocumulus clouds occur especially frequently in subtropical regions at the eastern edge of the oceans. (Wood, 2012; Klein and Hartmann, 1993). Sea surface temperatures (SST) are particularly cool in these regions, which is a favorable condition for stratocumulus clouds (Klein and Hartmann, 1993). Due to strong subsidences of warm dry air caused by large-scale circulations such as the Walker Circulation, and with an underlying inversion limiting a shallow well-mixed shallow layer benath, the development of stratocumulus clouds is further enhanced (Klein and Hartmann, 1993).

Important physical processes of the stratocomulus topped boundary layer responsible for the formation, development and maintenance of marine stratucumulus clouds are shown in Figure 2.3 (Wood, 2012). As mentioned before, a transport of warm moist air is very important for the formation of clouds. This transport is provided by a turbulent well-mixed boundary layer (Nicholls, 1984; Wyant et al., 1997). One of the physical processes responsible for turbulent mixing is longwave cooling. Longwave cooling at the top of the cloud is the most important process in terms of convection (Lilly, 1968). By absorbing shortwave radiation at the cloud top, it heats up and then radiates out as longwave radiation. The resulting temperature gradient in the form of high longwave radiation upwards from the cloud top and low longwave radiation from the clear sky area above the clouds downwards, causes a strong cooling in the uppermost meters of the stratocumulus cloud. This temperature gradient again favors the rise of warm air into the upper end of the cloud, which in turn strongly promotes convective growth during the day (Roach et al., 1982; Wood, 2012). Latent heating and evaporation are two other important processes that can increase turbulence (Moeng et al., 1992). In latent heating, the condensation of water droplets causes the air to heat up, creating

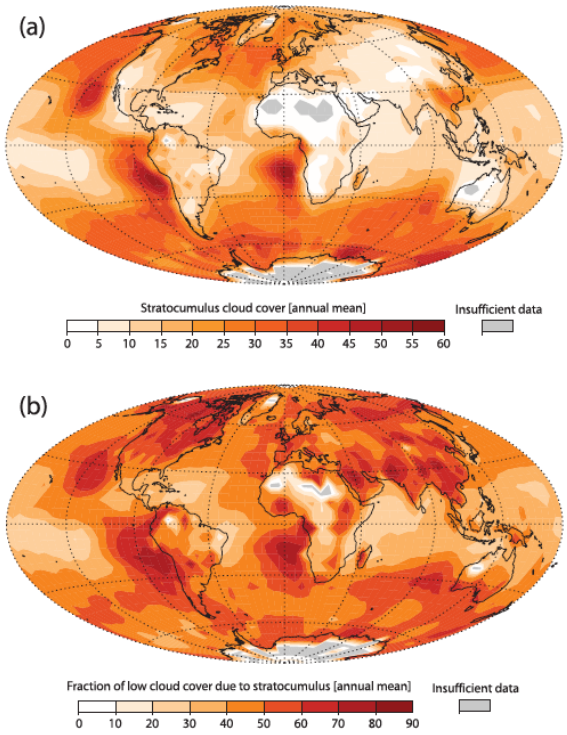


Figure 2.2: Figure of a) annual mean stratocumulus cloud cover and b) fraction of annual mean low cloud cover with stratocumulus clouds. Gray areas represent unknown data points (Fig. 4 in Wood (2012)).

updrafts. Downdrafts also occur, because the evaporation of water droplets cools the air. As a result of these processes, surface fluxes, such as moisture flux, reach the cloud layer more easily, which favors the convective developement of straocumulus clouds further (Atkinson and Wu Zhang, 1996).

Furthermore, the occurrence of stratocumulus clouds is also strongly temporally dependent (Wood, 2012). The temporal variability relates to seasonal, synoptic and diurnal scales. For the latter, mainly the solar radiation during the day and its absorption play a role. As mentioned before, absorption causes longwave cooling at the cloud top. As there is no solar radiation at night, this process takes place only to a very limited extent at night. Consequently, there is less circulation in the mixed layer at night and thus less exchange with the surface moisture, which is essential for the formation of clouds (Rogers and Koracin, 1992; Caldwell et al., 2005). Seasonal observations confirm a preferential formation of stratocumulus clouds under a well defined inversion (Klein and Hartmann, 1993). Furthermore, it can be observed that if a region has a strong stability of the lower troposphere, a high number of stratocumulus clouds can be observed in this region, and vice versa (Klein and Hartmann, 1993). Synoptic observations showed increased stratocumuli occurrence in connection with marine cold

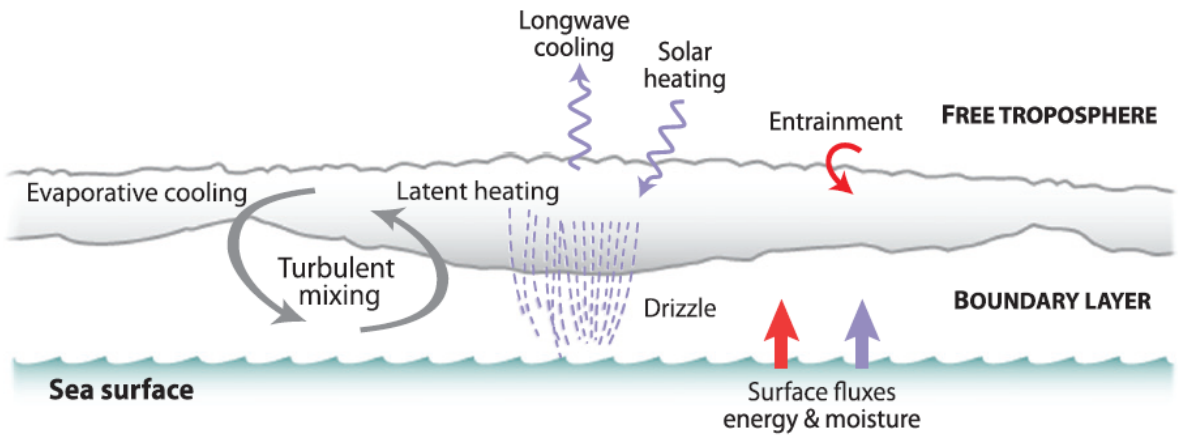


Figure 2.3: Illustration of the key processes that occur in and around stratocumulus clouds (Fig. 2 in Wood (2012)).

air outbreaks (MCAO) in the middle latitudes (Muhlbauer et al., 2014). MCAOs are defined as cold air movements moving from the poles toward the equator (Abel et al., 2017; Fletcher et al., 2016). At higher latitudes they mostly occur in cold air regions moving towards the equator and usually have a shorter lifetime (Norris and Klein, 2000). It was also found that a change in the amplitude and location of the suptropical high is coupled with the low cloud cover due to large-scale subsidence and upwelling of cold ocean waters (Klein et al., 1995; Klein, 1997; Muhlbauer et al., 2014).

The dissipation of stratocumulus clouds is associated with the decoupling of the cloud layer with the moisture transport from the surface (Wyant et al., 1997). Decoupling can be associated with advection of stratocumulus clouds over warmer waters, as exemplified by movements of air masses from the suptropics towards the equator (Bretherton and Wyant, 1997; Wyant et al., 1997). This movement over warmer surfaces enhances the upward draft of air and thus surface moisture flux, which in turn promotes turbulence in the boundary layer and thus convective growth. The enhanced updrafts also increase the inflow of dry warm air from the free atmosphere above. Consequently, there is a lower temperature gradient between the cloud and the surface layer and thus a weakening of the moisture transport, which in turn promotes the dissipation of the stratocumulus clouds (Bretherton and Wyant, 1997; Wood, 2012). Because of the increased convection due to large-scale advection over warm areas, cumulus clouds are formed, but at the same time, due to increased tropospheric entrainment, the large-scale stratocumulus clouds disappear (Wyant et al., 1997). This process is therefore also called a transition from stratocumulus sheets to trade cumulus (Wyant et al., 1997). This transition can also be considered as the passage through different stages of

mesoscale convection of a stratocumulus cloud. The mesoscale stages are referred to as closed MCC clouds (stratocumulus-like) and open MCC clouds (cumulus-like) (Wood, 2012) and are discussed in more detail in the following chapter.

2.1.2 Mesoscale cellular convection

The study of mesoscale morphologies is very important because each morphology type has different degrees of cloud cover and radiative properties. These observations are due to the different micro and macro physical properties that take place within these cloud systems (Wood and Hartmann, 2006; Wood et al., 2011). In this section, we will discuss in more detail the structures, occurrences, and micro and macro physical properties of the different morphologies and their transition from closed to open MCC.

Previous studies have shown that low clouds, especially over oceans, tend to form large-scale MCC clouds (Agee et al., 1973; Atkinson and Wu Zhang, 1996). In this regard, they basically exist in three different morphologies: open, closed and cellular but disorganized (Wood and Hartmann, 2006). Open and closed MCC both have hexagonal features in shape but are opposite in the type of cloud arrangement. As can be seen in Figure 2.4, open cells are transparent in the center, whereas the edges are cloudy. Closed cells are the other way around. They have an opaque center, but dissolve at the edges, making them more transparent there. The reason for the hexagonal shape of these two morphologies is due to the upward flow of warm air. This process is known as Rayleigh Bernard convection (Getling, 1998). However, the air flow is different in both cases. In the case of closed cells, the moist warm air rises in the center and sinks again at the edges. With open cells, on the other hand, the air rises at the edges and sinks in the center. Open cells are significantly more unstable than closed cells (Atkinson and Wu Zhang, 1996). Disorganized cells can be seen in the last row. They also have cellular clouds, but in contrast to closed and open MCC they are not organized on such a large scale (Wood and Hartmann, 2006).

Frequency of occurence of MCC

Muhlbauer et al. (2014) show that each MCC type has different regions where it is more prevalent, and all MCC types also exhibit a strong seasonal cycle (see Figure 2.5). This distribution is represented as frequency of occurrence, which is the proportion of a MCC type in relation to the total number of all MCC types.

The annual frequency of occurrence of closed cells shows that they are particularly prevalent at higher latitudes, as well as in the subtropics and in mid-latitude areas associated with storm tracks such as the west coasts of South America and South

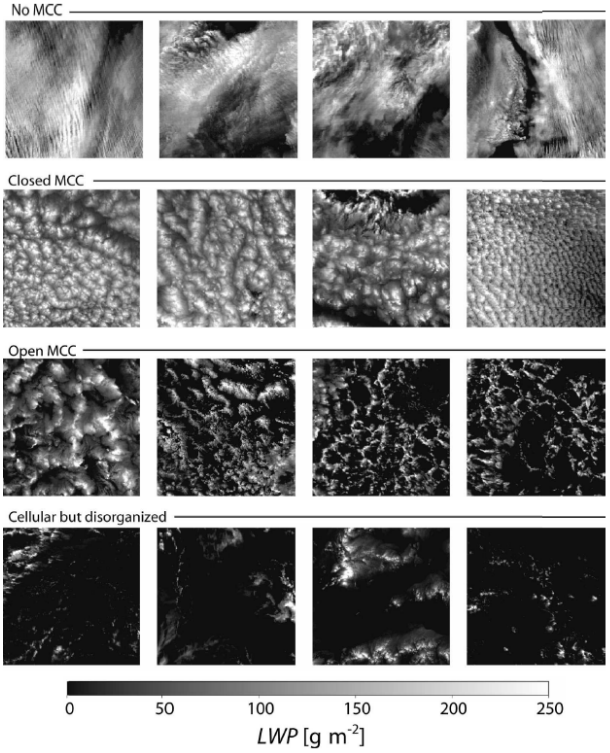


Figure 2.4: Illustration of the different MCC types by their liquid water path (in Fig. 5 in Wood and Hartmann (2006)).

Africa (McCoy et al., 2017; Muhlbauer et al., 2014). Especially in the summers of the Northern and Southern Hemisphere noticeable seasonal cycles can be seen (Muhlbauer et al., 2014). In the months of December, January and February an increased occurrence in more southern regions can be observed, whereas in the summer months of June, July and August a clearly increased occurrence in the Northern Hemisphere can be found. Furthermore, a correlation of the seasonal change of the closed cells with the seasonal change of the lower tropospheric stability is shown (Muhlbauer et al., 2014). However, recent studies have also shown that closed cells at higher latitudes have a lower frequency of occurrence than previously thought (Rampal and Davies, 2020; Mohrmann et al., 2021; Lang et al., 2022).

Open cells, on the other hand, prefer warmer SST's and a less stable boundary layer (Wyant et al., 1997), therefore they show a higher frequency of occurrence in the west of subtropical high pressure areas (Muhlbauer et al., 2014). They also often appear near the coast of subtropical areas, but further away from the coast than closed MCC (Muhlbauer et al., 2014). This can also be attributed to a higher SST. Of the three MCC types, open MCC have the lowest frequency of occurrence globally, but possess significantly fewer latitude-dependent trends than closed and disorganized MCC. Only

in subtropical areas is a tendency evident (Muhlbauer et al., 2014; McCoy et al., 2017). The seasonal variability of open MCC shows that they also have a higher frequency of occurrence in the winter months at higher latitudes like closed MCC. However, slightly shifted towards equator (Muhlbauer et al., 2014). This is probably due to MCAO's (McCoy et al., 2017).

Disorganized MCC are the most widespread of all MCC and are associated with regions with high SSTs in the tropics. Seasonal variability is much less pronounced in contrast to closed and open MCC (Muhlbauer et al., 2014)

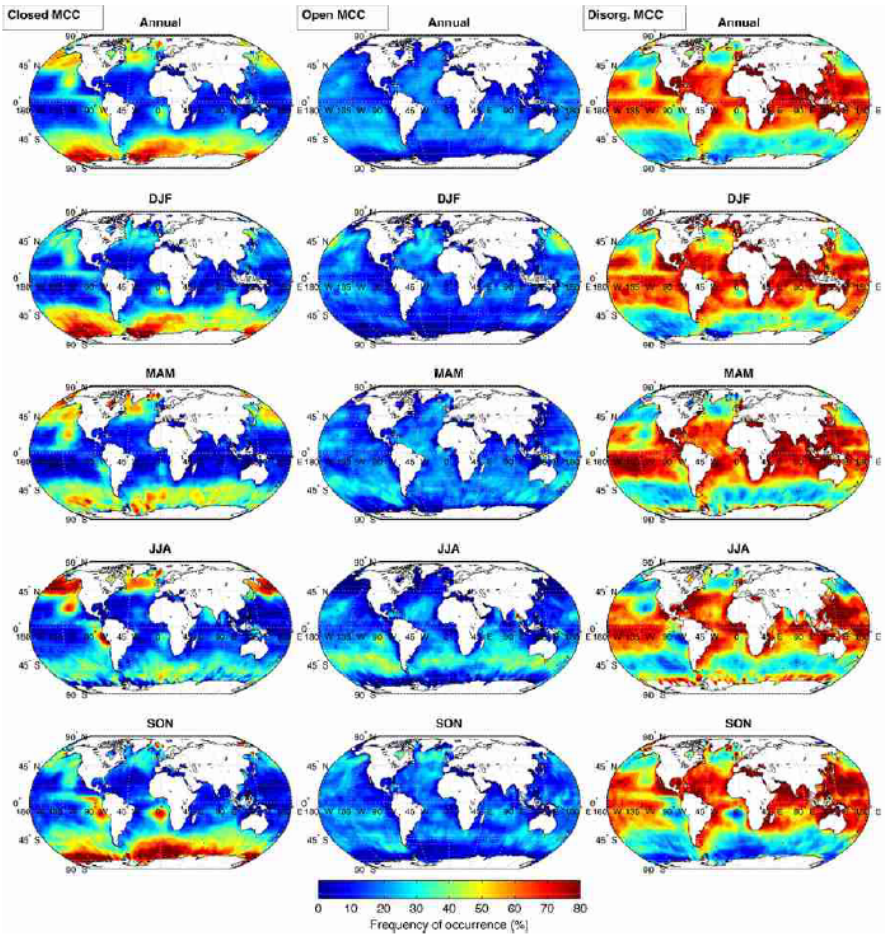


Figure 2.5: Plot of the annual, as well as seasonal, frequency distribution of closed, open and disorganized MCC based on the fraction of one MCC type to the full number of MCC types, expressed as frequency of occurrence. (in Fig. 5 in Muhlbauer et al. (2014)).

Micro and macro physical differences of open and closed MCC

Several studies have shown that significant differences and similarities in the physical processes that exist within the cloud system of open and closed MCC and these also

partially influence the transition from closed to open MCC (Wyant et al., 1997; Xue et al., 2008; Wood et al., 2011; Sandu and Stevens, 2011; Berner et al., 2013; Yamaguchi and Feingold, 2015).

According to Wood et al. (2011), similarities and differences of closed and open MCC can be identified in precipitation, aerosol concentration, marine boundary layer structure (MBL) and LWP. These findings confirm studies previously conducted on this subject (vanZanten and Stevens, 2005; Petters et al., 2006; Sharon et al., 2006; Comstock et al., 2007) and, together with Wood et al. (2011), yield the following conclusions: The cloud cover of overcast regions consisting of closed MCC is significantly larger than the cloud cover of areas consisting of pockets of open cells. In precipitation there are also clear differences to be seen (Wood et al., 2011). The only common feature is that almost all clouds precipitate in the form of drizzle. In closed MCC, however, more light drizzle prevails, not even reaching the surface. In the open region, drizzle is up to three times stronger than in the overcast region (Wood et al., 2011) and is thus significantly stronger, so that a large proportion also reaches the surface.

There are also strong differences in aerosol concentration and distribution with respect to accumulation und aitken mode (Wood et al., 2011). Aitken mode refers to the growth of very small aerosol particles by condensation of gases in a very short time. Accumulation mode then describes the somewhat slower growth that follows by liquid particles flowing together and coagulating with other particles. In pocket of open cells (POC), there are significantly lower concentrations of aerosols in the accumulation mode, with overall more aerosol particles in the Aitken mode than in the overcast region. This observation of accumulation mode particles also correlates with the lower cloud droplet concentration in POC. The lower concentration of Aitken mode particles can be explained by their accretion by condensation or coagulation with other particles (Wood et al., 2011). The MBL from the overcast region is significantly more mixed than that from the POC region, which is due to the decoupling of the surface layer with the cloud layer, thus interrupting moisture transport from the surface (Wood et al., 2011). The LWP is distributed homogenously flat in the region with closed MCC, whereas the LWP in the area of POCs is strongly fluctuating and distorted. The high proportion of LWP in POCs in the form of drizzle is striking (Wood et al., 2011).

One of the most important climatological differences of open and closed cells is that closed MCC clouds have a larger impact on shortwave radiation due to a higher albedo compared to open MCC clouds for the same cloud fraction (McCoy et al., 2017). Therefore, it is even more important to understand the transition from closed to open MCC to minimize model uncertainties.

Research on the closed-to-open transition has revealed two important physical processes that may be responsible. One is the advection of closed MCC over warmer SST and the other is precipitation associated with aerosols in closed cells (Yamaguchi and Feingold, 2015).

The first process has already been explained in chapter 2.1.1 with the dissipation of stratocumulus clouds to shallow cumulus and can be equally associated with the transition from closed MCC to open MCC. McCoy et al. (2017), for example, find that a strong correlation exists between MCAO's and the frequency of open and closed MCC. As the SST decreases with lower latitudes, strong temperature differences arise and the transition from closed to open MCC occurs.

The second process involves the almost cloud free areas created by the precipitation of closed cloud cells (see Figure 2.6). By precipitation so called cold pools are formed beneath closed MCC, which ensure that warm and humid air is forced to rise at the borders of these cold pools. This leads to the convection of new clouds at these locations. As soon as the region between the clouds is rained out, a gap is formed, which explains the shape of open cells (Wood et al., 2011; Berner et al., 2013). Also, a study shows that aerosols influence the shape of open cells by suppressing precipitation (Xue et al., 2008).

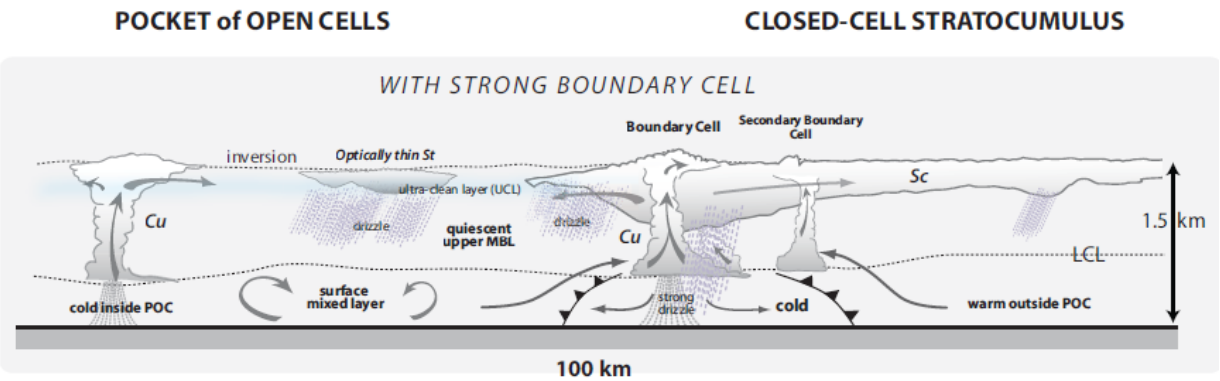


Figure 2.6: Illustration of the transition from closed MCC to pocket of open cells. The arrows represent the direction of the air flow and the lines with triangles represent the boundaries of the cold pool (upper drawing of Fig. 21 in Wood et al. (2011)).

However, precipitation does not guarantee the formation of open cells. Yamaguchi and Feingold (2015) for example show with the help of a model that isolated local weak precipitation regions are not sufficient for a transition. The implementation of several local precipitation regions, in contrast, shows better model results. However, the distance between the regions plays an important role aswell. Larger distances between

the regions have a significantly weaker effect on closed stratocumulus areas than shorter distances.

2.2 Neural network

The history of artificial neural networks goes back to the work of Warren S. McCulloch & Walter Pitts (McCulloch and Pitts, 1943). They showed that any arithmetic or logical function could be computed by simple classes of neural networks.

Then, in 1974, the backpropagation method was developed by Paul Werbos and is still considered the basis for artificial neural network training methods (Werbos, 1994).

Furthermore, special types of neural networks have been developed over time, such as the convolutional neural network (LeCun et al., 1998). In 2012, such a network also won for the first time at the ImageNet Large-Scale Visual Recognition Challenge (Krizhevsky et al., 2017).

An artificial neural network consists of a large number of interconnected neurons that are modeled on the human brain. They are used to solve complex tasks from various scientific fields such as physics, statistics and economics (Goodfellow et al., 2016).

The best known neural network structures are the feedforward neural network and the recurrent neural network. They differ in that in feedforward networks, unlike recurrent neural networks, the neuron connections do not form a direct cycle, but are only connected in one direction (Habibi Aghdam, 2017). Convolutional neural networks are a special type of neural networks and since only this one is discussed in this thesis, only this one is explained in more detail in Chapter 4.2.

In general, typical neural networks consist of three layers, where the input and output are connected to each neuron from the hidden layer and are therefore called fully connected layer. Convolutional layers are typically placed before those fully conncected layer to filter important features in advance. (Goodfellow et al., 2016). A typical schematic of a fully connected layer is shown in Figure 2.7.

Input Layer: The data to be processed is given to the input layer and this is weighted and given to the next layer. Passing the input with weights is also called feature transformation.

Hidden Layer: Hidden layers are located between the input layer and the output layer and can occur several times. Here, the weighted data from the input layer can be weighted again and passed on to another hidden layer with a different weighting or to the output layer.

Output Layer: Once the data has completed all feature transformations, it is passed to the output layer. The output layer then classifies the final data passed by the hidden layer.

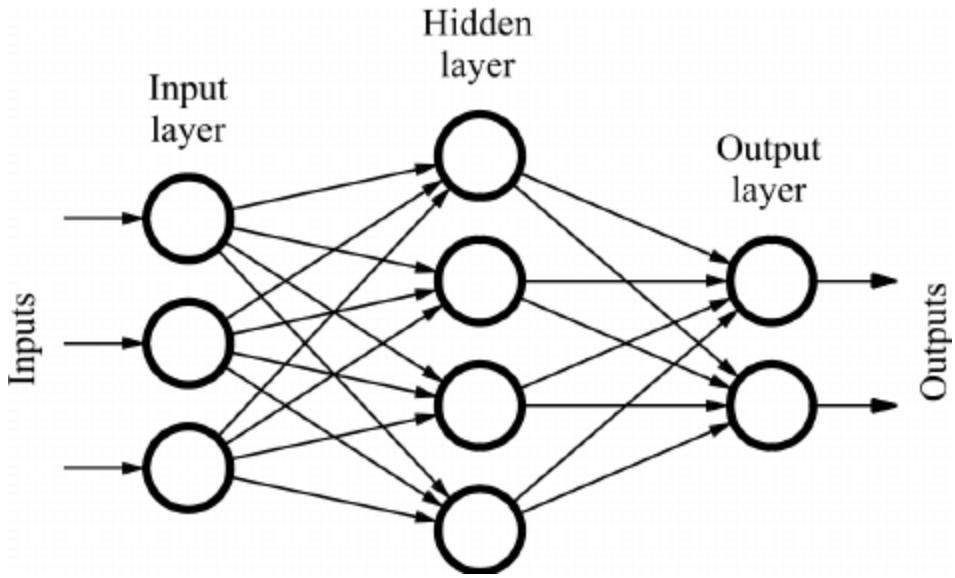


Figure 2.7: A typical illustration of a feedforward network with one input, hidden and output layer (Fig. 6.1 in Davim (2011)).

Figure 2.8 shows an exemplary structure of a single artificial neuron in a hidden layer. X_i represents the input of the data. These are first weighted with W_i. Then the weighted input is fed into the neuron and summed up into a linear function z. Part of this function is also often a so-called bias term b. This value is a constant and can speed up or slow down the transfer to the next neuron (Bishop, 2016). Thus, the summation function can be described mathematically as follows:

$$z = \sum_i W_i X_i + \Theta \tag{2.1}$$

Here, i stands for each input and weight given to the neuron.

Biological neurons normally charge up to a certain threshold value and only then pass on the information to the next neuron. This threshold is imitated in artificial neurons through activation functions f(z) (Towards Data Science, 2021). Neurons thus pass on the processed data (neuron output) only as soon as the linear function has reached this threshold. The network then uses a loss function to check the output value of the activation function against the expected value. The goal with each training session, also called epoch, is to minimize the loss function. This is achieved by optimizing the input

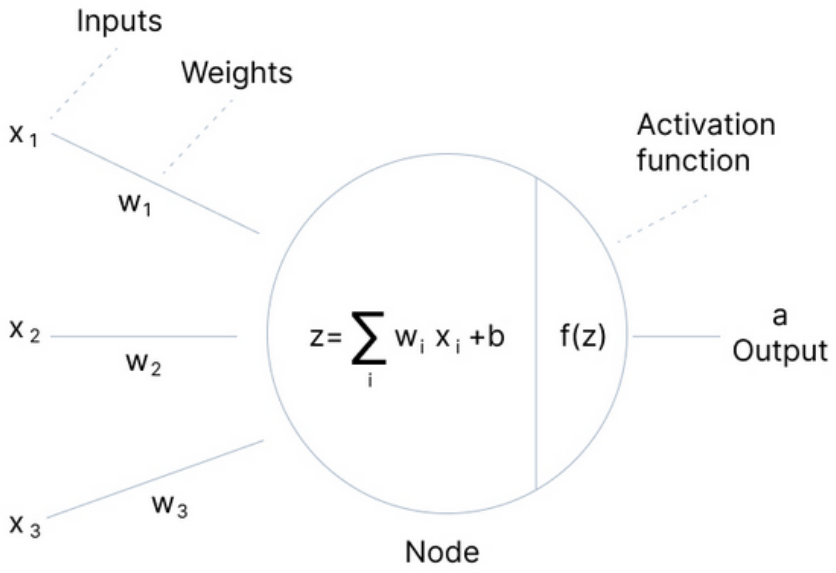


Figure 2.8: Simplified representation of the structure of an artificial neuron (v7, 2022).

weights. In addition, it is possible that a model is overtrained on the input, but this can be prevented with various regularization methods. These processes and methods are described in more detail in the following sections.

2.2.1 Activation functions

As mentioned before an activation function inspects the output of the summation function $z = x$ and checks if a certain threshold is reached. If this is reached, then the neuron is considered activated and the output of the neuron is forwarded to the next layer (Habibi Aghdam, 2017).

Basically, the activation functions of neural networks must satisfy two important properties. Namely **non-linearity** and **continuous differentiability**. At least one activation function should be non-linear so that the whole network is non-linear. Differentiability is important because neural networks mainly use gradient-based networks (Habibi Aghdam, 2017). In the following, a few typical activation functions are explained in more detail:

Sigmoid

The sigmoid activation function and its derivative are displayed in Figure 2.9. The sigmoid activation function can be described as follows:

$$\zeta_{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

and its derivation as:

$$\zeta'_{sigmoid}(x) = \zeta(x)(1 - \zeta(x)) \tag{2.3}$$

This activation function was very popular in the past, but it brings a big problem with it. Especially with very deep networks, it happens that the weight change is very small or even almost disappears. Consequently, the network comes to a standstill and it can no longer learn anything. This problem is also known as *vanishing gradients problem*. Mathematically, the problem is that if $|x|$ is not close to the origin, the gradient becomes very small (Habibi Aghdam, 2017).

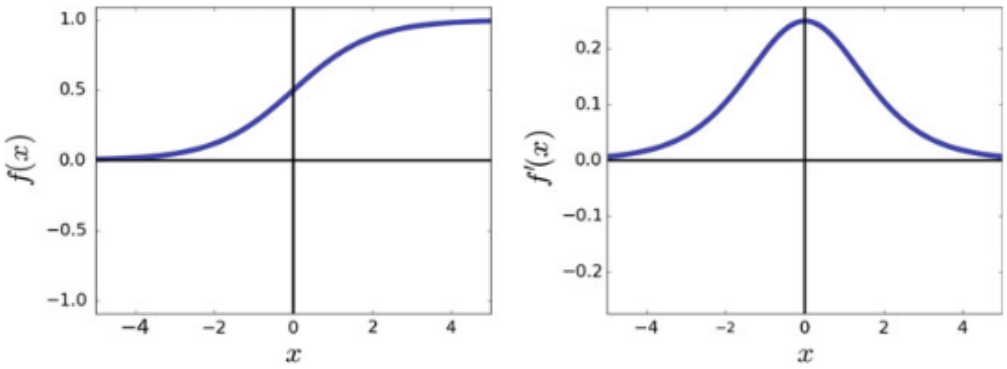


Figure 2.9: Display of the sigmoid function and its derivative (Fig. 2.39 in Habibi Aghdam (2017)).

Hyperbolic tangent

The hyperbolic tangent activation is shown in Figure 2.10. Its function and derivative are described by the following equations:

$$\zeta_{tanh}(x) = \frac{e^{x} + e^{-x}}{e^{x} + e^{-x}} = \frac{2}{1 + e^{-2x}} - 1 \tag{2.4}$$

$$\zeta'_{tanh}(x) = 1 - \zeta(x)^2 \tag{2.5}$$

This activation function has an important advantage due to the properties $\zeta_{tanh}(0) \approx 0$ and $\zeta'_{tanh}(0) \approx 1$. Namely, it increases the convergence speed of the gradient descend algorithm. It is preferable to the sigmoid activation function, but still has the same problem. Here, too, the *vanishing gradient problem* occurs, since saturation occurs when $|x|$ is increased (Habibi Aghdam, 2017).

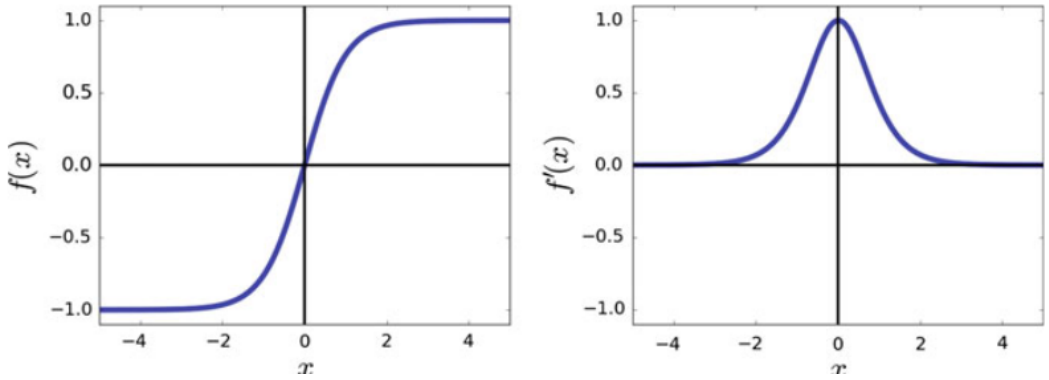


Figure 2.10: Display of the hyperbolic tangent actiaction function and its derivative (Fig. 2.40 in Habibi Aghdam (2017)).

Rectified linear unit

The rectified linear unit (ReLU) activation function is described as follows and is mapped in 2.11:

$$\zeta_{relu}(x) = max(0, x) \tag{2.6}$$

$$\zeta'_{tanh}(x) = \begin{cases} 0 & x<0 \\ 1 & x\geq 0 \end{cases} \tag{2.7}$$

In contrast to the activation functions explained so far, the ReLU offers some advantages. First, it is a very simple linear function and therefore computationally very efficient. Furthermore, unlike the activation functions sigmoid and tanh, it does not suffer from the *vanishing gradient problem*, since the function is not saturated at high numerical values. This makes it very popular especially for deep neural networks. The

disadvantage of this function is that if a neuron has a negative value, this is always output as zero. This results in so-called *dead neurons*, which can worsen the overall accuracy. By removing these dead neurons, the network can also become more efficient (Habibi Aghdam, 2017).

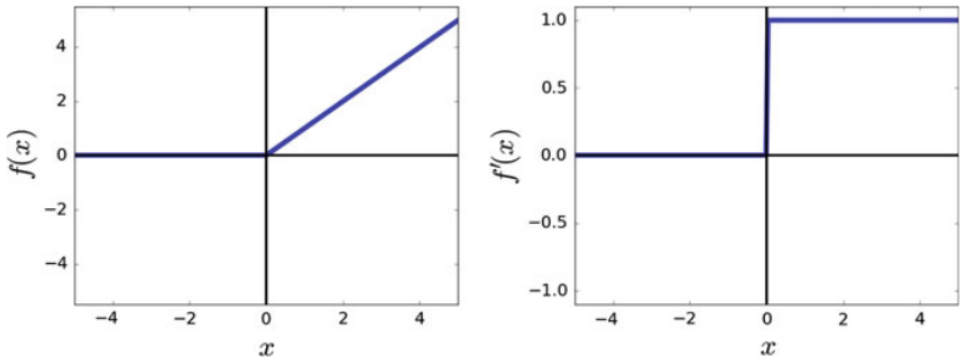


Figure 2.11: Display of the rectified linear unit function and its derivative (Fig. 2.42 in Habibi Aghdam (2017)).

Softmax

The softmax activation function is particularly useful at the end of CNN's because it gives the calculated values as probabilities relative to each other (Wood, 2019).

The softmax function is a multiclass variant of the sigmoid function and can be represented according to Analytics Vidhya (2021) as follows:

$$\zeta_{softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{2.8}$$

2.2.2 Loss functions

In general, the smaller a loss function is, the better the predictions of a model are. The value of model outputs will be compared with the expected value of a model. This comparative value is called error. Using the result of the loss function, the weights of the previous layer are updated (A Blog on Building Machine Learning Solutions, 2021). In the following, the cross entropy based loss function is explained in more detail.

Cross entropy

The cross entropy loss is a logarithmic based loss function and increases exponentially as the prediction deviates from the actual result (Hastie et al., 2017). Two popular cross entropy based functions are the binary cross entropy (see Equ. 2.9) and the categorical cross entropy function (see Equ. 2.10). They differ in that the former is applied to classifications with two classes and the cross entropy function is applied to multi-class classifications. They are described mathematically as follows:

$$Loss = -(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \tag{2.9}$$

$$Loss = \sum_{j=1}^{M} y_j \log(\hat{y}_j) \tag{2.10}$$

y stands for the expected value and $\hat{y}$ for the predicted probability.

2.2.3 Optimizer

Optimizers are used in neural networks to minimize the loss calculated by the loss functions so that the model performs better overall. The minimization of the loss is achieved by adjusting the weights and learning rate of the network. Especially the **Adaptive Moment Estimation (Adam) Optimizer** is very popular in neural networks, because it shows the most promising results of minimizing the loss (Towards Data Science, 2020a) The formulas for the gradients of the Adam optimizer are shown as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{2.11}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{2.12}$$

m_t and v_t are moving averages and g_t represents the gradient. They are respectively composed of the averages m_{t-1} and v_{t-1} and the hyperparameters β_1/β_2 (KDnuggets, 2020).

2.2.4 Regularization

In a neural network, problems can occur that include underfitting and overfitting. Figure 2.14 shows the loss function of the training and validation dataset to illustrate what is the meaning of underfitting and overfitting. The dashed blue line represents the desired point at which the loss has found its minimum. To the left of this point is the area of underfitting, where the loss tries to minimize itself. Accordingly, the model has not yet been trained with enough data or has not had enough time to train and has not yet reached the desired point. The overfitting area is on the opposite side. The model has learned the features of the training set too well and is therefore less able to assign the features of the validation set (IBM, 2021; Bishop, 2016). Therefore, it is important to avoid overfitting and in the following are a few regularizatrion methods to prevent this.

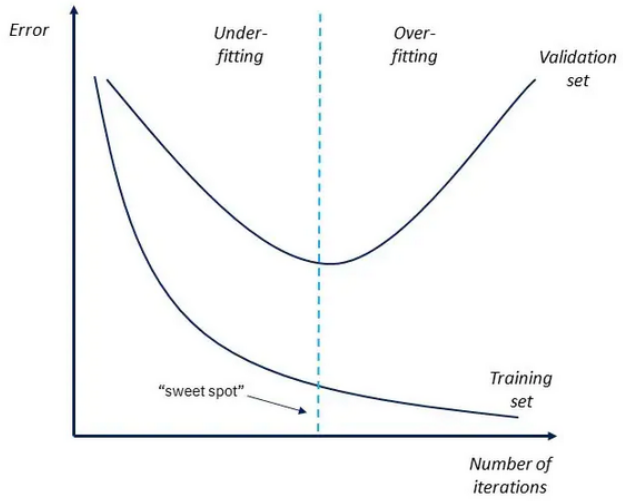
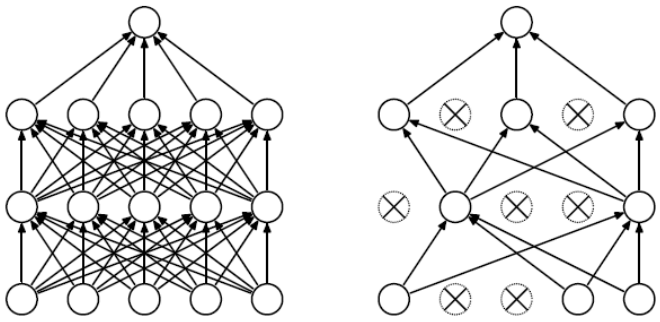


Figure 2.12: Schematic of the course of a loss function of the training and validation set is shown and on the basis of these it is made clear how underfitting and overfitting is to be understood (IBM, 2021).

Dropout

Dropout is an effective method to improve a model and can be described by the Figure 2.13. On the left you can see an ordinary neural network. There, each value of one layer is connected to a value of the next layer. On the right, however, you can see a network where dropout has been applied. As the name suggests, certain values are removed from the network with a certain probability. This thins out the network and should be able to better ignore random correlations between the data. (Srivastava et al., 2014)



(a) Standard Neural Net (b) After applying dropout.

Figure 2.13: Illustration of a neural network with and without dropout. (Fig. 1 in Srivastava et al. (2014)).

L1 & L2 regularizer

The loss functions of the two regularizers are as follows:

$$L1 = Loss(y, \hat{y}) + \lambda \sum_{i=1}^{N} |W_i| \tag{2.13}$$

$$L2 = Loss(y, \hat{y}) + \lambda \sum_{i=1}^{N} |W_i^2| \tag{2.14}$$

Here, y stands for the actual value and $\hat{y}$ for the predicted value. The regularization parameter is given by λ and N gives the number of features. They differ from each other in the last term. The L1 regularizer penalizes the absolute values of the weights (W), whereas the L2 regularizer penalizes the sum of squares of the weights. Both regularizers are intended to improve the model, but each has a different effect. The L1 regularizer is particularly well suited for feature selection, provided that the input data also has a high number of features. The reason for this is that it sets the coefficient of less important features to zero (Towards Data Science, 2017). The L2 regularizer penalizes the weights less severely because it does not set their values to zero, but lets them run towards zero. The advantages of the L2 regularizer is that it is computationally cheaper than L1 and is particularly well suited for overfitting. (neptuneblog, 2021; Goodfellow et al., 2016).

Batch normalization

Batch normalization is a normalization method which takes place during the learning of a model (Baeldung on Computer Science, 2021). The normalization can be described with the following formula.

$$z^N = \frac{z - m_z}{s_z} \tag{2.15}$$

The mean value of the output of a neuron (z) is represented by m_z and its standard deviation is represented by s_z. Batch normalization allows taking advantage of higher learning rates and therefore makes the model more efficient (Baeldung on Computer Science, 2021).

Data augmentation

In general, the higher the dataset, the better the model. If the existing dataset is too small, it can be increased by data augmentation and thus reduces overfitting. The existing images are slightly changed by rotation, zoom or cropping. A more detailed description of the methodology used in this study is given in Chapter 3.2.1.

2.2.5 Convolutional neural network

A CNN is a special kind of neural network. The beginnings go back to the work of LeCun et al. (1998). There the first developed CNN was applied to handwritten numbers.

A CNN is especially used in pattern recognition, for example in image or speech recognition. Basically, such a network consists of three typical layers (Goodfellow et al., 2016). These three layers are the convolutional layer, pooling layer and the fully connected layer. The last one was already explained in Chapter 2.2 at the beginning. The first two layers are typically set before the fully connected layer and in the following their function will be described in more detail.

Convolutional layer

Convolutional layers have their name because of the mathematical operation "convolution". They are used instead of general matrix multiplications in CNN's like the fully connected layer and are therefore an important tool in data processing (Goodfellow et al., 2016).

The operation convolution is an integral consisting of two functions expressing the amount of overlap of one function (f) with another function (g) (Goodfellow et al., 2016) and can be represented as follows:

$$s(t)=[f * g](t)=\int x(a)\omega(t-a)da \tag{2.16}$$

$[f * g](t)$ stands for the convolution of f and g at a given time t and point a.

In terms of CNN, the function f is often referred to as input. In image recognition, this is represented in the form of a multidimensional array. The function g, on the other hand, represents the kernel and is also a multidimensional array consisting of parameters that are adjusted by the algorithm. These arrays can be called tensor. The output $s(t) = [f * g](t)$ is represented as a feature map (Goodfellow et al., 2016).

Figure 2.14 illustrates how a typical 2D convolution is performed. In the process, the kernel slides over the input matrix and is multiplied by it. The result of this matrix multiplication is represented as output. Once this step is completed, the kernel moves one box to the right and performs another matrix multiplication with the new values. This process is repeated until the boundary values are reached. Then the kernel shifts to the left edge again, but is then shifted down one step. The kernel repeats this pattern until the values at the lower right edge of the matrix have been reached.

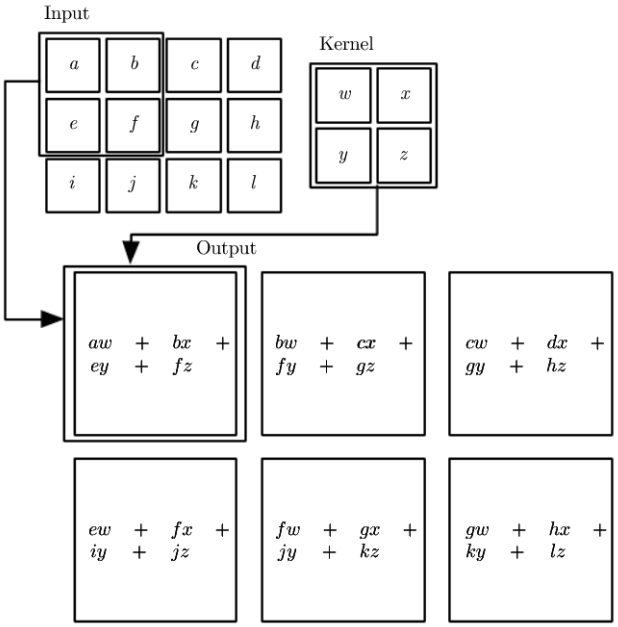


Figure 2.14: Illustration of a typical convolution. This shows how the kernel is applied to the input and how the output results from it Goodfellow et al., 2016.

In Figure 2.14, it is noticeable that the kernel is smaller relative to the input and output. However, this is a decisive advantage over other neural networks. This advantage is also called *sparse interaction*. Despite the small kernel, it is possible to recognize important features of an image. Therefore, less data has to be stored in the output. This results in an increased storage capacity and is statistically more efficient (Goodfellow et al., 2016).

Another advantage of convolution is that only one parameter (content of the kernel) is used for several functions. This is also called *parameter sharing*. This means, each value of the kernel is usually also used for one value of the input, see Figure 2.15. This leads to the relief of the memory capacity (Goodfellow et al., 2016).

Pooling

A CNN usually contains a so-called pooling layer. It is used to compress and reduce the resolution of the output of the convolution layer in order to recognize certain features of images. The aim is to make these features invariant to small changes in the input (Goodfellow et al., 2016). Different types of pooling can be used. Two of them are **max pooling** and **average pooling** (Making Developers Awesome at Machine Learning, 2020). In the following, we will take a closer look at max pooling, which is explained with the help of the Figure 2.16. Here a 2 x 2 max pool matrix is applied to a 4x4 matrix. No region will be overlapped and this means that the 2 x 2 matrix can only be

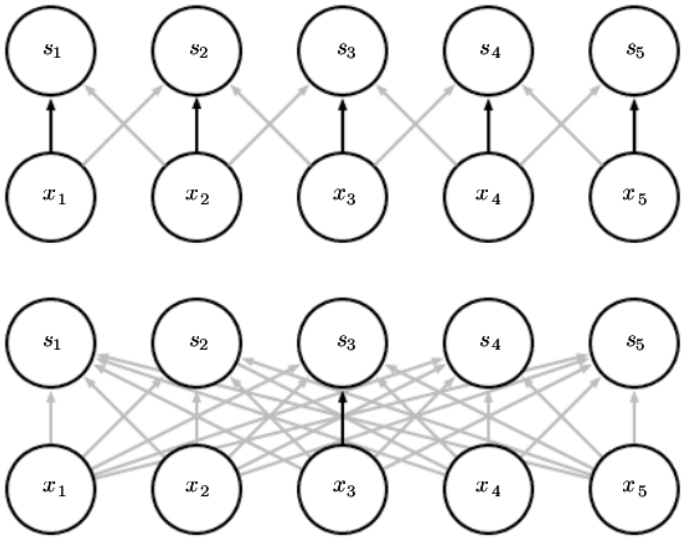


Figure 2.15: Example of paramter sharing. Above, paramter sharing is applied. The black arrows represent the application of the kernel to each input. The lower part of the figure shows the use of a weighted matrix in a fully connected model. There the parameter is used only once (Goodfellow et al., 2016).

applied four times. For each application it goes over a 2 x 2 value area of the 4x4 matrix and picks the largest value from each area, hence the term "max pooling". These values are then given as 2 x 2 output.

The resulting output from pooling is then either passed to another convolutional layer and pooling layer or it is passed to the fully connected layer. Here the values are represented as probabilities and the model assigns a class to the data (see Chapter 2.2).

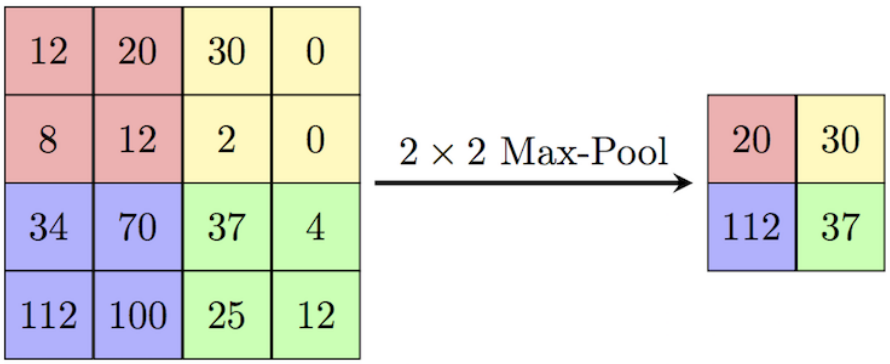


Figure 2.16: Illustration of Max Pooling. (*Max-pooling / Pooling - Computer Science Wiki* 26.02.2022).

3 Methods and data

The methods and data chapter is divided into three sections. In the first subchapter, the data collection, the selection of scenes and the categorization of these are described in more detail. There, questions have to be answered such as: Which satellite is probably best suited for data acquisition? What data from the satellite are needed? From which areas should the data best be taken? How large should the scenes be? How should the scenes be classified and which conditions should be set?

The next section explaines how the acquired scenes are expanded in order to fit into the required format for the CNN. Then, a listing and detailed explanation is given about the different model configurations that have been used in this work. In addition, it is briefly mentioned how the results of the models are statistically evaluated.

The last part of methods and data describes how the results of this work have been compared with another dataset from another NN algorithm.

3.1 Data acquisition

3.1.1 Data source

To generate enough scenes for the classification, satellite data will be used. We decided to use data from the Aqua satellite, which is a polar orbiting satellite, in order to get a global coverage. Geostationary data is not used because it would require using multiple satellites with different bands and data formatting to get a global image. Furthermore, no temporal development of clouds is needed for this project, which would be the advantage of geostationary data. The Aqua satellite was launched in May 2002 and in June the first data was received (Miller and McKee, 2004). There is also the Terra Satellite, which contains the required data aswell, but does not have the advantage of the Afternoon Constellation (A-Train Constellation).

The special feature of the A-Train is that several satellites fly in a row at short distances from each other, thus collecting data at short intervals with a wide variety of

instruments (Schoeberl, 2002), such as the MODIS sensor. The MODIS data is available different processing levels. For this work, data from the level 1 "MOD 03 Geolocation" and the level 2 "MOD 06 Cloud Product" dataset are taken. The first dataset contains information about longitude and latitude, solar zenith angle and land/sea mask and has a spatial resolution of one kilometer at nadir (Parkinson and Greenstone, 2000). The MOD 06 dataset includes cloud optical and physical parameters such as cloud top height and reflectance. The MODIS sensor uses 36 different bandwidth channels with different spectral resolutions. The first two bands have a spatial resolution of 250 m. Bands 3 - 7 have a resolution of 500 m and the remaining 31 bands 1000 m (Miller and McKee, 2004; Parkinson and Greenstone, 2000). For this work the second reflectance band with a spectral range of 0.86 μm is chosen, because it is shown that clouds over the ocean can be displayed very well with this band (Platnick et al., 2003). With this reflectance band, the neural network is trained. The spatial resolution of this dataset is 1 km and 5 km at nadir. (Parkinson and Greenstone, 2000).

For this work, the data is used with a spatial resolution of 1 km at nadir, which corresponds to a grid resolution of 1354 pixels in width and 2030 pixels in length. One pixel represents 1 km^2.

3.1.2 Scene selection

In order to find as many scenes as possible, certain regions on Earth are selected, that according to Muhlbauer et al., 2014, have a high accumulation of MCC clouds. Figure 3.1 shows the global distribution of annually averaged low cloud fraction. Noticeable is a high distribution of low cloud fraction in subtropical regions in the eastern parts of the oceans. The observations seen here are associated with persistent decks of subtropical marine stratus. In addition, subtropical regions are chracterized with upwelling of cold ocean waters near the coast, strong subsidences in subtropical high pressure areas, and weaker tropospheric stability. Based on this, regions with higher low-cloud fraction are selected: North Atlantic (NA), South East Pacific (SEP), South East Indian (SEI), North East Pacific (NEP) and South Pacific (SP) (Muhlbauer et al., 2014). Furthermore, Muhlbauer et al., 2014 show, the frequency of the different morphologies is strongly seasonal dependent. Figure 3.2 shows the frequency of occurrence of the three MCC morphologies for each region selected by Muhlbauer et al. (2014). Especially in the winter months of the respective hemisphere an increase in frequency of occurence of open MCC can been seen. Because the number of open and closed cell images should be about the same, so the network is not biased towards one class which has high training numbers, the months of December, January, and February (DJF) are selected for the

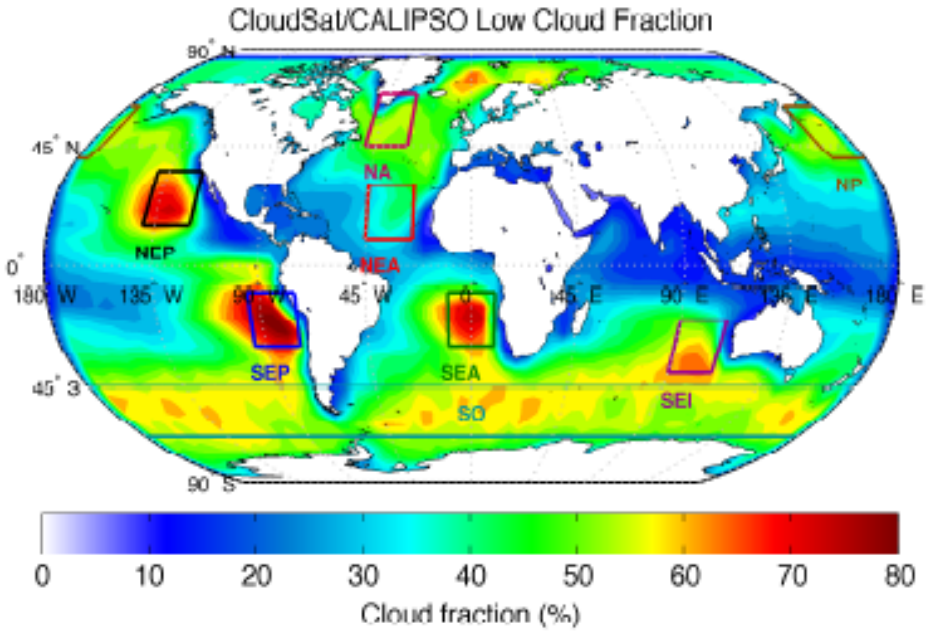


Figure 3.1: Illustration of the globally displayed low cloud fraction (Fig. 2 in Muhlbauer et al. (2014)).

Northern Hemisphere regions NA and NEP, and the months of June, July, and August (JJA) are selected for the Southern Hemisphere regions SEI, SEP, and SP. In order to get even more data, the respective summer months are also added for the regions NA, SEP and SEI. The regions and their respective ranges of longitudes and latitudes can be seen in Table 3.1.

The secence are selected based on the following procedure. For the years 2008 and 2009 the MODIS data MYD 03 and MYD 06 with the spatial resolution of 1 km natdir are used and the resulting scenes have a resolution of $1354 \,\mathrm{x}\, 2030\,\mathrm{km}^2$. The pixel size of the scenes given into the model should have a resolution of $224 \,\mathrm{x}\, 224$ pixels. We chose this certain size for the secenes, because on the one hand the images must be as small as possible for a good computational performance of the model, but on the other hand still large enough, so that the cloud structures of the individual morpholgies are well recognizable. Furthermore, this size is often used in deep convolutional neural networks, for example with pre-trained models like VGG 16 (Simonyan and Zisserman, 2014). The $1354 \,\mathrm{x}\, 2030$ scenes are then cut into as many $224 \,\mathrm{x}\, 244$ scenes as possible under certain conditions. These preconditions are related to the selected regions, cloud top height and land sea mask. The latter limits the local restriction in terms of land and sea properties. For the sea properties, the scenes had to be either over moderate or continental ocean and deep ocean. The first is ocean which is at least 5 km from the

coast and 50/500 m deep and deep ocean means ocean which is deeper than 500 m. In addition, clouds within the scene had to be below a cloud top height of 3500 m, as we are only interested in low clouds. If these requirements are fulfilled, the corresponding scene is cut out in the respective regions and saved as a PNG and h5 file with all associated data. The resulting scenes are then used for the scene classification to assign a category to them.

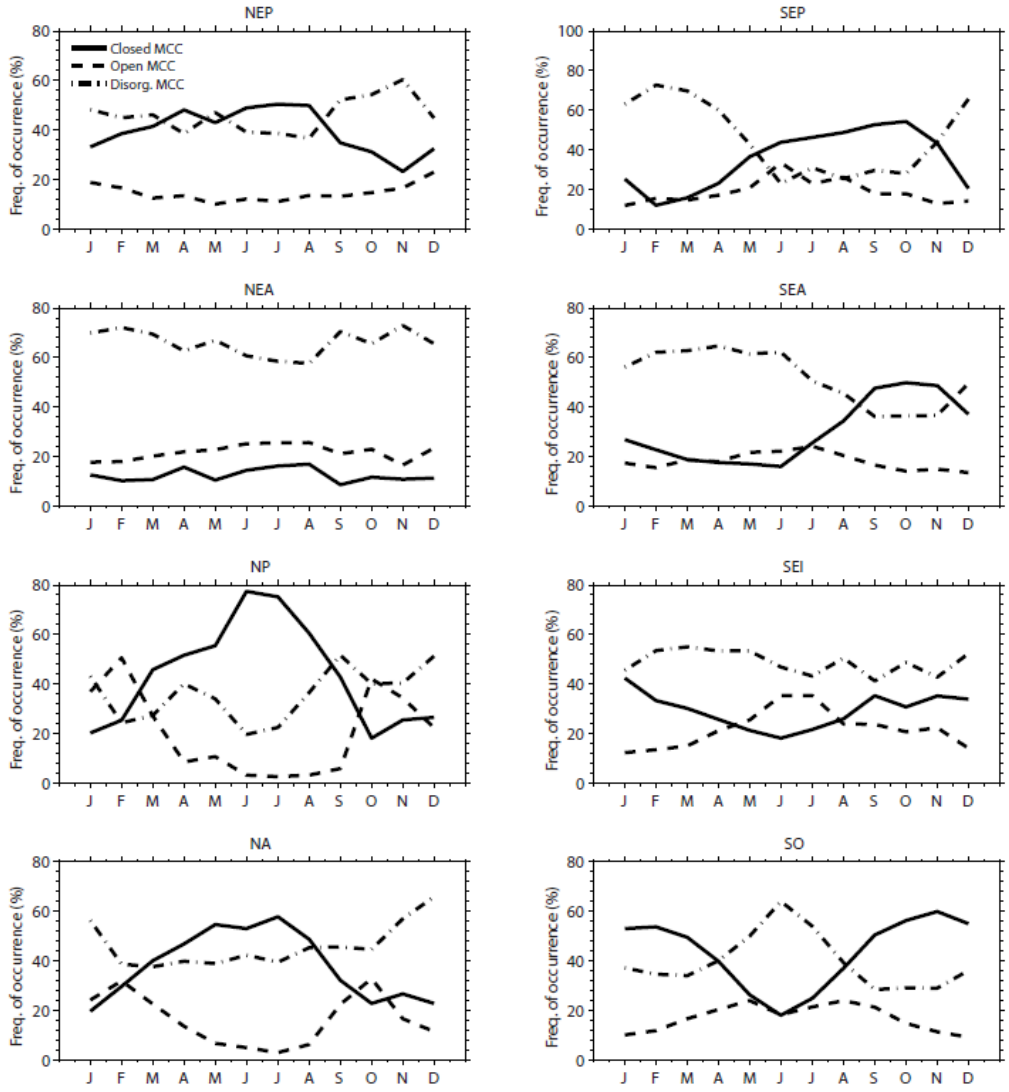


Figure 3.2: Trajectories of the seasonal cycle of frequency of occurrence of open, closed and disorganized MCC for each region (Fig. 6 in Muhlbauer et al. (2014)).

Table 3.1: Coordinates of the selected regions.

Coordinates		
Region	Latitude	Longitude
NA	40 ° to 60 °	-45 ° to 25 °
SEP	-40 ° to -20 °	-90 ° to -70 °
SEI	-60 ° to -40 °	70 ° to 90 °
NEP	15 ° to 40 °	-150 ° to -120 °
SP	-60 ° to -30 °	-180 ° to -90 °

3.1.3 Classification procedure

The scenes are classified by visuall inspection of five classifiers, who are familiar with the morphology of MCC organization and received a training presentation to reduce errors as much as possible. The images are classified on three consecutive days. To keep errors to a minimum, the working time is divided so that there are enough breaks to ensure stable concentration. The prepared scenes are classified according to one of five labels. These are "Open", "Closed", "NoMCC", "Disorganized" and "ERROR". The "ERROR" label is used for scenes which show the swath border and therefore create additional non-real edges.

The classification is done using a labeling tool from Schnelle (2020). Within this tool, all prepared scenes are labeled by each classifier. On the right side of the tools interface (Fig. 3.3), you can always see the image that needs to be classified. Directly above are the labels to be selected, which are confirmed either with the mouse or with the keys 1-5 of the keyboard. As soon as a classifier has selected a label, this is displayed at "Current rating". In the red marked field the path can be selected, in which the images are stored. Once this is done, all available images are displayed in the blue marked area. By using the arrow keys or by pressing the "previous" or "next" button, you can iterate through the dataset.

Images are considered "classified" if at least three out of five classifiers agreed on the same class. Otherwise, the scene will not be used. Furthermore, a scene is considered open or closed only if at least 70 % is covered accordingly with the resprective structure. More detailed information about the available data, as well as the results and statistical analysis of the scene classification are presented and analyzed in Chapter 4.1.

The data obtained from the classification appointment is divided into 70 % training dataset and 30 % test dataset. The test dataset is used to check the accuracy of the trained models after. The training dataset is further extended to improve the model with data augmentation.

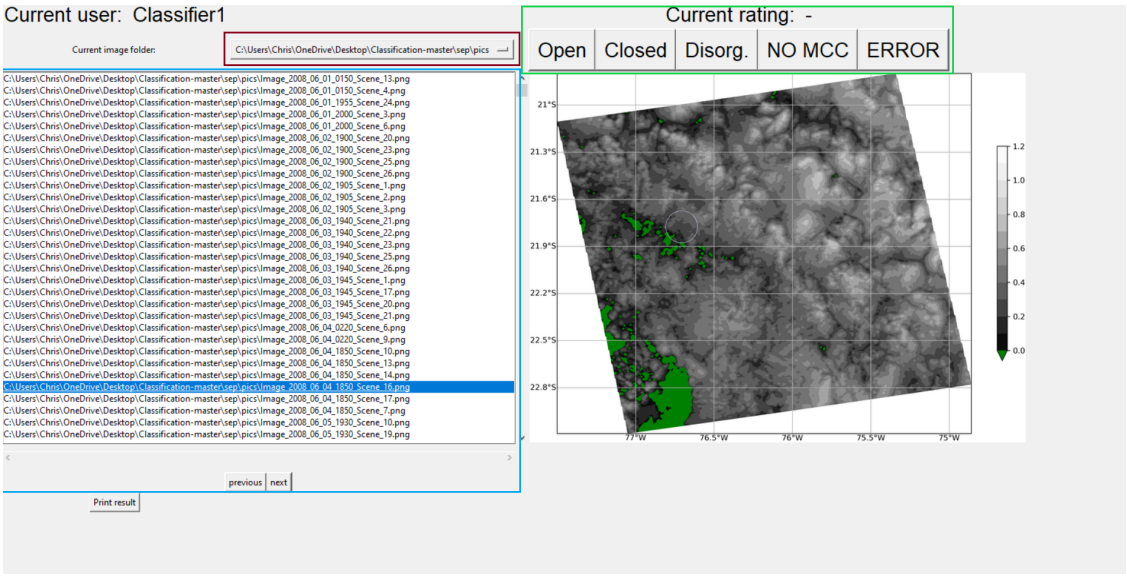


Figure 3.3: Image of the labeling tool used to classify the images. The category window is outlined in green. The file path to the images is outlined in red and in the blue outline you can see the different images to be selected.

3.2 Data augmentation & network configurations

3.2.1 Data augmentation

To avoid the overfitting mentioned in the section 2.2.4, data augmentation is applied to the available training dataset of the 224 x 244 images used in this work. Data augmentation is used to increase the existing dataset by making small changes to the image. Past deep learning models such as CNN's have shown that larger datasets often improve the performance of the model (Perez and Wang, 2017; Sun et al., 2017). In this work, we mainly focus on the more traditional transformations such as cropping, zooming, flipping, and rotation, as these have also been successful in past scientific work (Halevy et al., 2009). The dataset is augmented "manually" via Python and via a function of the module "tensorflow" (Abadi et al., 2016), where you can preset which augmentations should be applied by an "augmentation layer" and the augmentations run synchronously with the remaining layers of the model. In the following the two different augmenation configurations are explained in more detail.

Manuel data augmentation

Table 3.2 shows the different configurations for the manually augmented images. The respective augmentation is applied to each of the 1155 images and added to the dataset.

Table 3.2: Listing of the different configurations of the manuel data augmentation.

Augmentations	Factor
Flip	horizontal and vertical
Crop	horizontal and vertical 44 pixel
Rotation	60 °, 120 °, 240 ° and 300 °
Zoom	0.3
Brightness	0.8

For the change in brightness, all values are multiplied by a factor of 0.8. For the crops, 44 pixels are cropped horizontally and vertically and then resized back to 224 x 224. For the augmentation zoom a negative zoom of 0.7 was applied. This means that the image has been zoomed in by a factor of 0.3. For the flips the rows and columns of pixels for each image have been swapped horizontally and vertically respectively. And finally in the last augmentation the images have been rotated by 60 °, 120 °, 240 ° and 300 ° and then cropped to 163x163 images. The reduction to 163 pixels is done because of the gaps on the sides caused by the rotation. Afterwards, the image is scaled back to 224 x 244 pixels. Finally, the augmentation brightness has been applied to all augmented images except for itself, resulting in a dataset of 21945 images in total.

Data augmentation layer

Table 3.3 shows the different augmentations of the data augmentation layer. All of them occur randomly during the training of the model with the original 1155 images. The images are flipped horizontally and vertically, rotated by an angle dividable by 40 °, zoomed in in the range of 10 % - 20 %, changed in contrast by a factor 0.7, cropping of the images to the size 163 x 163 and being translated in width and height by a factor of 0.2.

Table 3.3: Listing of the different augmentations of the data augmentation layer and their factor

Augmentation	Factor
Random Flip	horizontal and vertical
Random Rotation	40 °
Random Zoom	height = (-0.1 ,-0.2)
Random Contrast	0.7
Random Crop	height = 163, width = 163
Random Translation	height = 0.2, width = 0.2

The following Table 3.3 shows a list of the augmentations used for the two augmentation methods. The resulting augmentations from both methods are illustrated

and explained in more detail in Chapter 4.2.1. The datasets obtained through the data augmentations are then run through different CNN models with different configurations to find out which configuration produced the best result. These configurations are described in the following.

3.2.2 CNN

For the use of the CNN, the framework tensorflow is chosen, which can be operated via Python. The framework is one of the most popular for machine learning and deep learning and is open source software and therefore accessible to everyone (Abadi et al., 2016). With tensorflow, it is very easy to integrate and configure the layers required for a CNN. Also, the different regularizations methods can be implemented at various points in the model, which is why the decision is made in favor of this framework.

Model configurations

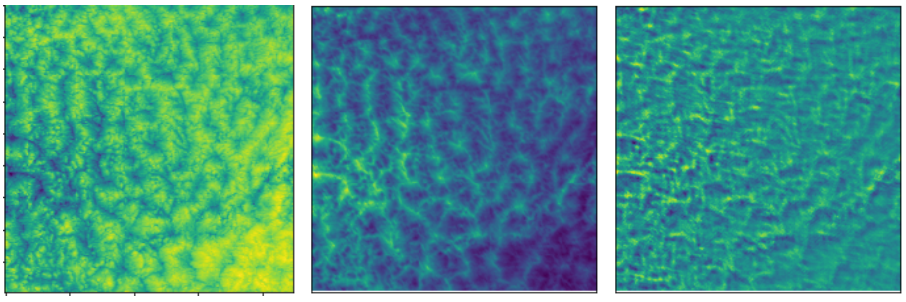
For the models, we basically worked with different typical configurations for CNN's. These possible configurations and their settings can be seen in Table 3.4. Since we work with a CNN, there are also convolutional layers. In the different runs different filters and kernel sizes are used for the convolutional layer. The number of filters varied between the sizes 32, 64, and 128 or these are combined with each other by adding several layers one after the other. The reason for the runs with different sizes is that the model can take up more features by increasing the number of filters during training and thus also improve the prediction (Brownlee, 2020). The choice for kernel size is 3 x 3, as CNN's have been shown to be more efficient and better at detecting features in

Table 3.4: Listing of the different configurations and their settings.

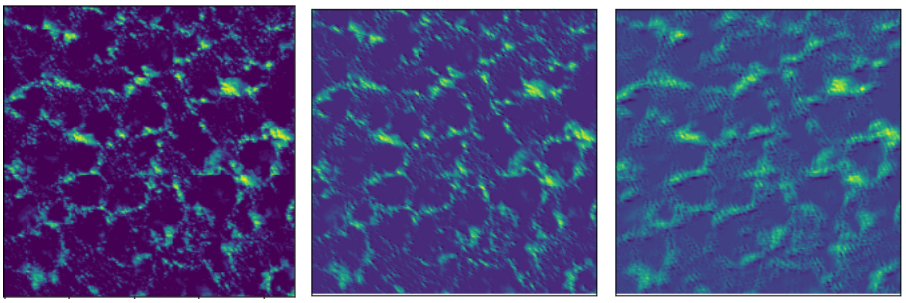
Configuration	Setting Values
Convolutional Layer	Filter: 32/64/128, Kernelsize: 3 x 3
Max Pooling Layer	Size: 2 x 2
Dense Layer	Size: 64 and 3
Activation Functions	RELU and Softmax
Dropout	0.5 and 0.3
L2 Regularizer	Bias/Kernel = 0.01
Batch Normalization	Default
Adam Optimizer	Learning Rate = 0.001
Loss Function	Categorical Crossentropy
Batch Size	20
Epochs	50

different situations (Simonyan and Zisserman, 2014; Analytics Vidhya, 2020a; Analytics Vidhya, 2020b). Also the plotting of example images (see Fig. 3.4) at the kernel sizes 3 x 3 and 7 x 7 has shown that the structures are better to recognize by the smaller 3 x 3 kernel size, which is why this configuration is chosen. Behind each convolutional layer there is a max pooling layer, which reduces the output of the convolutional layer by probing for the highest value with a 2 x 2 matrix. When the input image has passed all convolutional layers it is passed to two dense layers. The first dense layer filters the previous output using an activation function and is connected to all neurons of that layer. A typical size for a dense layer is 64 (Towards Data Science, 2020b). The second dense layer is the last one of the model and has the size corresponding to the number of classes to be classified. In this case it is three, and outputs the probability for each class. As activation function of this dense layer the softmax function is used, since it is a multiclass problem. And for all other layers the relu activation function is used, since this has achieved the best results in past applications (Brownlee, 2021).

Furthermore, the regularization methods dropout, L2 regularizer and batch normalization are used for some settings. The L2 regularizer is used in a convolutional layer as kernel regularizer and as bias regularizer. The kernel regularizer reduces the weights, whereas the bias regularizer, as the name suggests, reduces the bias term.



(a) Closed (b) Kernel: 3x3 (c) Kernel: 7x7



(d) Open (e) Kernel: 3x3 (f) Kernel: 7x7

Figure 3.4: Image of an (a) original closed and (b) original open image and their respective representation after filtering by a 3 x 3 and 7 x 7 kernel.

Dropout with the value 0.5 is used behind the dense layer (64), because in the original paper of Srivastava et al. (2014) it is used only for this kind of layer. However, a recent study also shows that dropout after the convolutional layer with lower values may well perform better (Park and Kwak, 2017). Therefore, model runs are also carried out in which dropout is placed behind convolutional layer with a lower value of 0.3.

In the model, the adam optimizer is used with the learning rate set to 0.001. The learning rate indicates the degree to which the weights must be corrected for a given error. The batch size indicates how many samples are trained by the network at once. The batch size is set to 20 for all model runs. As soon as all batches of the training dataset are run through once, this is called an epoch. In total, 50 epochs are defined for all model runs.

Of the dataset available for training, 0.3 of it is used as the validation dataset. This is not taken for the training process, but is used to determine the model capability while the model is setting parameters. Based on these configurations, six different models are created, all using different settings.

Model runs

Six different model runs, which have all different configurations, with both augmentation datasets are compared by looking at the loss and accuracy curves of the training and validation dataset. These model configurations for each model run can be seen in the Figures 3.5 and 3.6

In Figure 3.5 different layer combinations (LC) can be seen. The shortcuts are used to make the model runs clearer, because these layer combinations are often repeated. The base is LC1. There the model layers 1) convolution layer, 2) the corresponding activation function and 3) the max-pooling layer are summarized. In LC2 a batch normalization layer is added and in LC4 a dropout layer. LC3 contains both dropout and batch normalization.

In Figure 3.6 the structure of the model for the respective runs are shown. For Run 1, for example, the input to the model is the training dataset. The dataset first encounters the layer combination LC1 $(32, 3 \times 3)$. The first number, in this case 32, indicates the number of filters and the second entry (3×3) indicates the kernel size of the convolutional layer. After the dataset is filtered by the convolutional layer with the relu activation function, the output is then done by sorting out the max pooling layer to the dense layer (64). Before this, the dataset goes through a process called flattening which is marked as a red arrow. In this process, all three-dimensional arrays are converted into one-dimensional arrays so that they can be processed by the dense

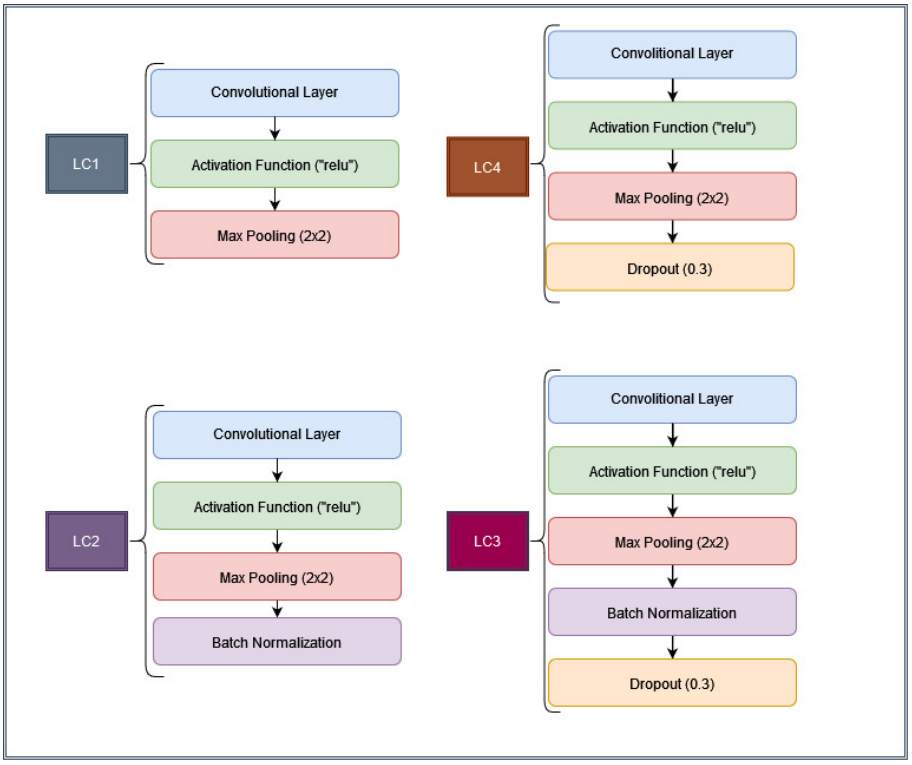


Figure 3.5: Representation of layer combinations for better illustration of Figure 3.6.

layer (64) with the activation function relu and passed on as output to the last layer of the model with the activation function softmax. This then finally gives the probability of a category for each image as output.

In Run 2, two layer combination LC1 is added. But with a different number of filters (32, 64, 128). In Run 3 the model is extended with a dropout layer behind the first dense layer. Model Run 4 has the same structure as Run 3, except that L2 regularizers are added to the convolutional layers of each layer combination. In Run 5 the first layer combination is replaced with LC2, which includes batch normalization. And last but not least in model Run 6, LC3 is used for the first layer combination and LC4 for the following two. This added a dropout layer to each layer combination compared to Run 5.

The models then predicted the test dataset. The best model of each augmentation method is then evaluated using a confusion matrix. The illustration of this matrix and the statistical procedure is explained in Chapter 3.1. Finally, the test dataset and the two predicted datasets are compared with another dataset.

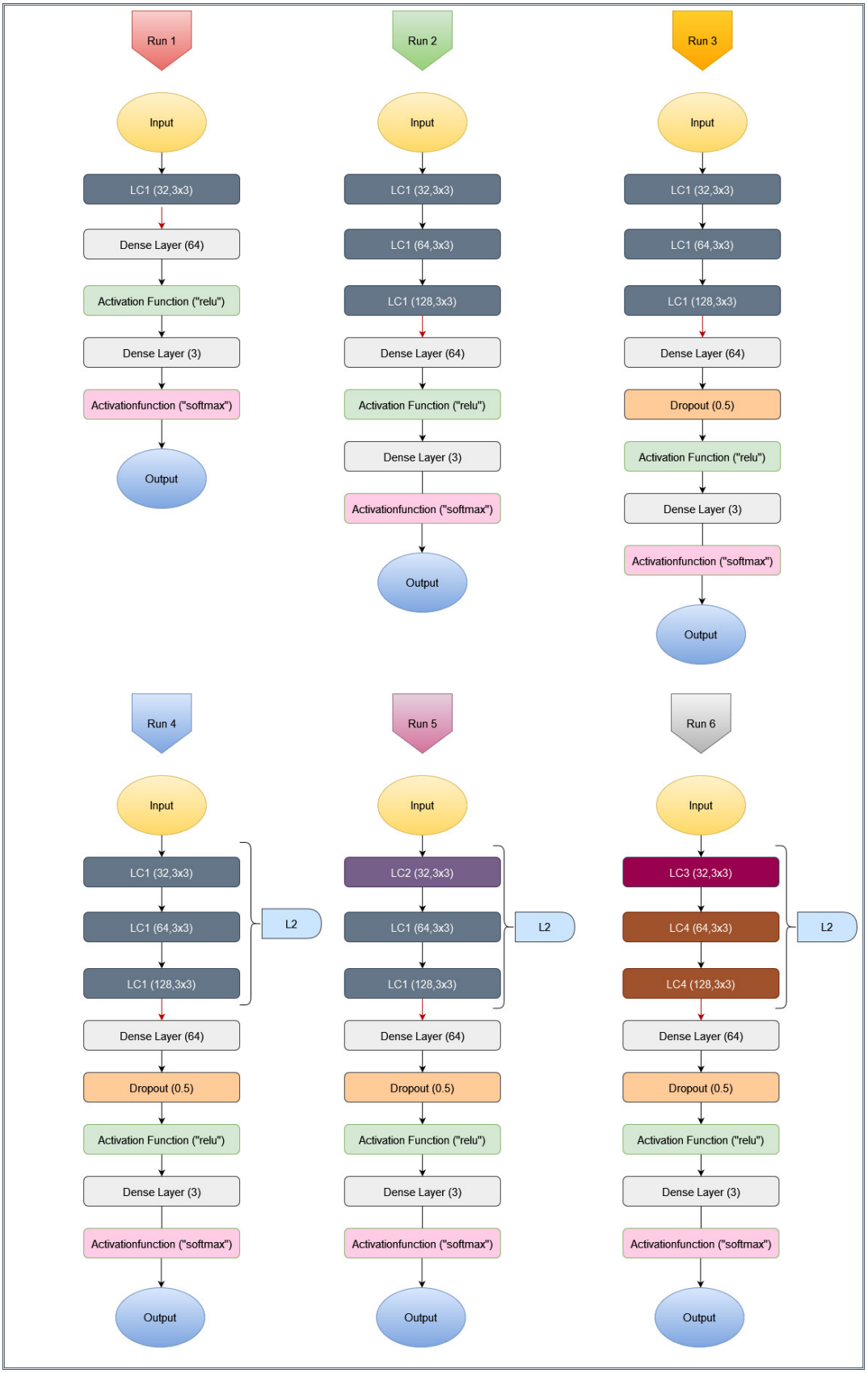


Figure 3.6: Flowshart to illustrate the structure of the six different model runs.

3.3 MCC dataset

The test dataset, as well as the predicted datasets from the best model runs from both augmentation methods, is compared to an enhanced MCC dataset from McCoy et al. (2017). The dataset includes tens of thousands of scenes over the years 2008 and 2009 that are globally distributed. The scenes are originally from the MODIS Aqua collection 6.1 and are $256 \times 256\,\mathrm{km}^2$ in size and offset 128 km apart so that they overlap halfway. To classify these scenes, they used the NN algorithm of Wood and Hartmann (2006). This algorithm then classifies the scenes using the LWP probability density function and the LWP power spectrum into the categories open, closed, and cellular but disorganized. Each scene is thus assigned to a category and represented as a coordinate in the center of their scences.

Since the scenes available for this work are of a different size and slightly offset with respect to the center coordinate, the scenes could not be compared with each other with pinpoint accuracy. Therefore, the coordinate points of the corresponding category are compared with each other using following criterion.

A circle with a radius of 152 km is set around the points of this work. Afterwards we check whether in this circle scene points of the MCC dataset are and if, which categories these are assigned. The given radius could contain up to five scene points of the MCC dataset. If the absolute majority of the scene center points has the same category as our dataset, it is considered a match. If there is not an absolute majority, either by tie or a majority of another category, the point is considered as "No Match". If the circle around the scene did not include any points from the MCC dataset, they are categorized as "No Datapoint".

4 Results

The results chapter is divided into three sections. In the first section, the results from the classification appointment are presented. In addition, these are statistically evaluated and compared with the results of recent studies of the climatology of open and closed MCC clouds.

In the second part, the training dataset obtained from the scene classification is extended via two data augmentation methods. Each of the two resulting datasets is then trained with six different model configurations. Then the best models of the respective augmentations method predicted the test dataset, which the model has never seen before. The two best predictions are compared statistically using a confusion matrix.

The last part deals with the comparison of the dataset obtained from 1) the scene classification and 2) the predicted test dataset by one of the best models with the MCC dataset.

4.1 Scene classification

Of the original dataset of 28645 MODIS sensor images from 2008 and 2009, not all images could be classified due to time constraints. This concerns especially scenes from the SP region from 2009. This number amounted to 6944. In addition, all images are removed, where three or more agreed with "ERROR" (1594 images). These are images in which non-real edges exist and therefore not meaningful. Furthermore, only three images are assigned to the category "Disorganized", thus this category is also not considered due to lack of quantity. Consequently, 20104 images are available and 19771 of them are successfully classified to one of the classes Open, Closed and NoMCC with the condition that at least three or more classifiers have agreed.

4.1.1 Classification statistics

The statistical analysis was inspired by the paper of Stevens et al. (2020). Table 4.1 shows the fraction of classified images for which k or more classifiers are in agreement.

It also shows what the fraction of classified images would be if they were randomly selected for the cases where at least three and four classifiers must match a class. This value is determined with the Equations 4.1 and 4.2:

$$Random = l \sum_{i=k}^{n} Pr(i; n, p) \tag{4.1}$$

$$Pr(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{4.2}$$

p stands for the probability that a class is randomly assigned. The number of classifiers is expressed as n, and k is the number at which a certain number of classes must be identical. The number of classes is expressed by l. In the case of three classes must match, these values are ($p = 1/3$, $n = 5$, and $k = 3$) and for the other case where four classes must match, the values are ($p = 1/3$, $n = 5$, $k = 4$).

Table 4.1 shows that for the proportion where three or more are assigned the same class. There is a very good agreement of 98 %, which is well above the probability of 63 % if the classes were randomly selected. Even in the case of $k = 4$, there is still a very good agreement of 86 %. The probability here that a class was randomly classified four times in the same way is significantly lower with 14 % and again illustrates the strength of the overall agreement of the classifiers.

Table 4.1: Representation of the fraction of images that matched $k = 3$ and $k = 4$ times with the number of available images (20104), as well as the probability if the three classes were selected k times at random.

k	3	4
Actual	0.98	0.86
Random ($p = 1/3$)	0.63	0.14

To take a closer look at the strength of the agreement, Table 4.2 shows the fraction of the matches between the three classes for the cases $k = 3$ and $k = 4$. First, it can be seen that the fraction of the no MCC class is the highest for both cases, since a large part of the images are classified as such. However, it is interesting to note that for $k = 4$, the agreement between open and closed is about half as large as for $k = 3$. For no MCC, the difference is only roughly 10 %. Thus, overall, there seems to be much greater agreement among the no MCC class.

Table 4.2: Fraction of images of each class for which the five classifiers matched three and four times. The sum of the individual class proportions gives the values 0.98 and 0.86 from Table 4.1 at "Actual".

k	Open	Closed	No MCC
3	0.042	0.027	0.914
4	0.020	0.013	0.831

The small number of open and closed classes compared to no MCC classes is further demonstrated by Table 4.3. There you can see the absolute number of images of the respective category with k = 3 and k = 4. Closed is the least categorized label. Open are categorized about 35 % more often and no MCC dominates by far in both cases with 18374 and 16703 respectively. This is probably because all images that were not categorized as MCC morphologie were classified the no MCC category and explains the high fraction in Table 4.2 aswell.

Table 4.3: Absolute number of images of each class for which the five classifiers matched three and four times.

k	Open	Closed	No MCC
3	847	550	18374
4	405	255	16703

To investigate annual variations between the categories, it is worth looking at the distribution of classified categories between the years 2008 and 2009. These can be seen in Figure 4.1. It is clear that fewer images are classified overall in 2009 than in 2008, primarily because the 2009 images are removed for the SP region. Therefore, it is difficult to make a statement about the annual variation. However, as can be seen from Table 4.5, the SP region had a very large part of the total images. The fraction of open images in 2008 is about three times larger than in 2009, while the number of no MCC images is twice as large, and the closed category has about a third more images. The large contrast between open and closed can be explained by the large amount of open images in Table 4.5 compared to closed in the SP region.

Table 4.4 shows the assignment of categories and their sum for each of the five classifiers. The sum of classifiers (hereafter named as C) 1-4 is about the same for each with 19700. Only C5 has significantly fewer images with 14666. This is due to the fact that from C5 the file with the stored data for the summer regions is damaged and therefore the data could not be extracted. But this is not so important for the categories open and closed, because they are hardly classified in the summer time anyway (see Tab. 4.5). C4 categorized the open class the most with 1683 images and C3 the least

with 710. For the closed class, C2 (1038 images) had the most images and C3 (482 images) the least. And in general, for open and closed, the number seems to vary for each classifier. For no MCC, this variation is weaker. This can also be confirmed using Table 4.2. There, the proportion of open and closed has decreased by about half at k = 4 matches compared to k = 3 in contrast to the no MCC images with 8.5 % less. Nevertheless, overall at k = 3 with 98 % there still seems to be a large agreement among the classifiers.

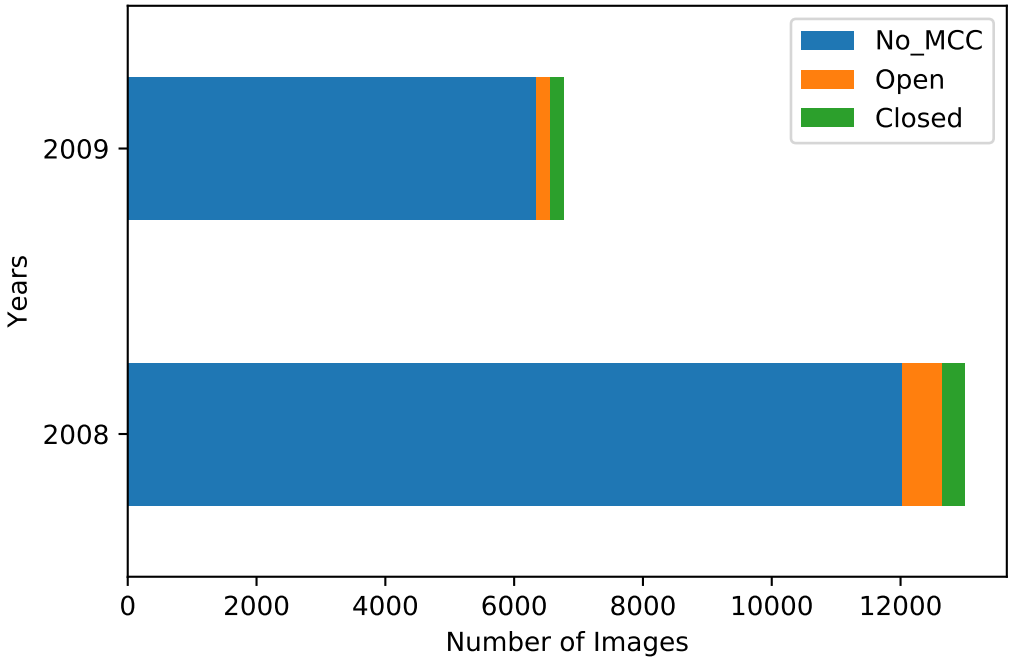


Figure 4.1: Histogram of the distribution of the three categories over the years 2008 and 2009 for k = 3.

Table 4.4: Number of classified categories divided among the respective classifiers.

Classifier List				
Classifier	Open	Closed	NoMCC	SUM
Classifier 1	1031	663	18246	19940
Classifier 2	1381	1038	17425	19844
Classifier 3	710	482	18455	19647
Classifier 4	1683	839	17195	19717
Classifier 5	940	705	13021	14666

4.1.2 Climatology of MCC

In the following, the k = 3 categorized images are compared with the data from other studies. Table 4.5 shows the classified images for k = 3, sorted by region according to the respective season.

Summer labels

It is noticeable that in the regions NA, SEP and SEI hardly any open and closed images are categorized in the summer months. In the subtropical regions SEP and SEI this is probably related to the seasonal cycle of the cloud fraction, which is also related to the seasonal cycle of lower tropospheric stability and SST (Muhlbauer et al., 2014). As a result, warmer SST'Ss occur in the summer months, leading to a weaker temperature gradient between the sea surface and the boundary layer top and thus to a weakened convection formation of clouds. Muhlbauer et al. (2014) also show a simultaneous anti-correlation in the seasonal cycle with disorganized MCC and closed MCC. This would explain why no open MCC were found in the summer months at lower cloud fraction, since in the scene classification of this study probably all disorganized images are sorted into the no MCC category.

Only in the SEP region are high values of closed images (88) registered. According to Muhlbauer et al. (2014), there tends to be a lower frequency of occurrence for closed and open images during the months of DJF, but there is a small peak of 28 % during

Table 4.5: Number of classified cateogries with the condition that three or more classifiers agree divided by the regions and the seasons, as well as their sum of the respective category.

Winter Labels				
Region	Open	Closed	NoMCC	SUM
NA (DJF)	150	41	1219	1410
SEP (JJA)	76	129	1406	1611
SEI (JJA)	109	71	1084	1264
NEP (DJF)	68	133	3706	3907
SP (JJA)	429	70	5869	6368
Summer Labels				
NA (JJA)	3	9	1900	1912
SEP (DJF)	3	88	1767	1858
SEI (DJF)	9	9	1423	1441
Sum Labels				
Sum	847	550	18374	19771

January, which also correlates with the increased cloud fraction there at that time, which in turn may explain why more closed MCC clouds are found there.

Furthermore, it is surprising that so few closed images have been detected in the NA region, since a frequency of occurence of about 55 % is seen there at (Muhlbauer et al., 2014). This could be due to the fact that in the NA region the cloud fraction is with values of around 50 % generally lower than in the other regions (see Figure 3.1) according to Muhlbauer et al. (2014). However, recent studies also show that at higher latitudes, due to MCAO's, there are significantly more no MCC clouds than closed MCC clouds than previously thought (Lang et al., 2022; Mohrmann et al., 2021; Rampal and Davies, 2020). Furthermore, in the algorithm of Wood and Hartmann (2006) used by (Muhlbauer et al., 2014), all stratiform clouds are included in the closed category, but we classified them as no MCC, which in turn could explain the increased number of closed images in the NA region in the study of Muhlbauer et al. (2014), compared to other studies.

Winter labels

Significantly more open and closed images are classified in the winter months of the regions, which is propably due to a higher gradient of air and ocean temperature that favors stronger energy and moisture fluxes due to MCAO's and advection of cold air from continents due to cyclons. (Muhlbauer et al., 2014; McCoy et al., 2017)

It is remarkable that in the regions NA, NEP and SP, which are located at higher latitudes, more open MCC clouds are classified than closed MCC clouds in comparison with Muhlbauer et al. (2014). However, particularly striking here is the six times higher number of 429 images classified as open in the SP region in contrast to the 70 closed images. This lower tendency towards closed MCC clouds at higher latitudes is also observed in recent studies such as Lang et al. (2022), Yuan et al. (2020), and Rampal and Davies (2020). There it is assumed that the MCAO's occurring at higher latitudes favor significantly less closed MCC clouds than previously thought. This in turn is contrary to the observations of Muhlbauer et al. (2014) and McCoy et al. (2017), which observerd higher occurences of closed MCC clouds and that more MCAO's favor a closed to open transition due to stronger temperature gradients at higher latitudes.

In the subtropical regions SEP and NEP open MCC clouds are about 50 % less represented than closed MCC clouds. As already mentioned, the higher occurrence of closed MCC clouds in the subtropics is probably due to the high anti-correlation between disorganized and closed MCC clouds and the correlating seasonal cycle of cloud fraction and SST with closed MCC clouds. The occurrence of open MCC clouds

in the regions can be explained by the preferential conditions that prevail further away from the coasts. There, higher SST's and a less stable boundary layer prevail, thus favoring a transition from closed to open MCC clouds.

Among all regions, in the SP region the most images are classified. Thus it can be assumed that in this region especially many clouds with the scene conditions defined in the methods chapter are found. If the year 2009 had been additionally classified, there would probably be even more open and closed images available. The main reason is probably the larger range of latitude and longitude of the SP region compared to the other regions (see Tab. 3.1). Also the NEP region with a slightly larger range has a larger number of hits (3907) than the NA, SEP and SEI regions. The SEI region has the lowest number of images for both seasons. The reason for this could be the generally lower cloud fraction observed there at Muhlbauer et al. (2014) compared to the NEP and SEP regions, for example.

The images categorized with the condition $k = 3$ (see Table 4.3) are used as the dataset for the CNN because of the higher number compared to $k = 4$ and are expanded by using data augmentation.

4.2 CNN models

4.2.1 Data augmentation

From the available dataset, the closed images are first randomly divided into 385 (70 %) and 165 (30 %) images. Since an equal number of images should exist for each category for an equivalent training of the model, 385 images are randomly selected for training and 165 for testing from the open and no MCC datasets. This results in a training dataset of 1155 images and a test dataset of 495 images. The training dataset is extended using two different data augmentation approaches to reduce overfitting.

As can be seen in Figure 4.2, data augmentation can increase the variability of the relatively small dataset, since augmentations are related to different physical properties of the clouds. For example, by changing the brightness or the contrast, a cloud scene with different reflectance values can be mapped. Similary by reducing brightness and contrast they can resemble cloud scenes with lower water content. Spatially variable images can be recreated by cropping, flipping or rotating scenes for example. Since the cell sizes of MCC also vary between 10 - 50 km, it is useful to set different zoom levels. Thus cloud fields can be imitated, which would have formed on shallower boundary layers for example.

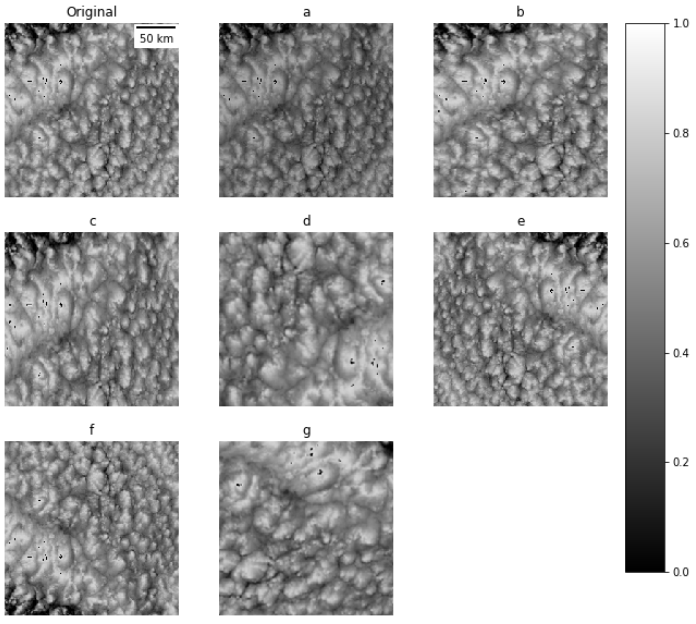
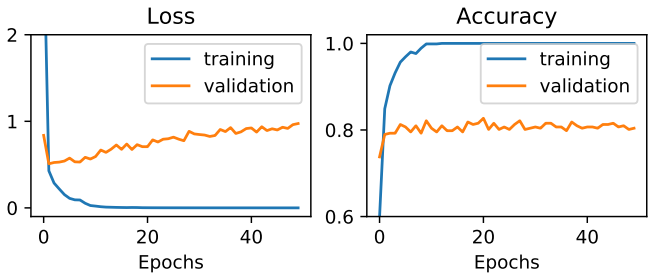


Figure 4.2: Display of the most individual data augmentations (a-g) that have been augmented and the original image (top left). a) Brightness b) Horizontal Crop c) Vertical Crop d) Zoom e) Horizontal Flip f) Vertical Flip g) Rotation and Cropping.

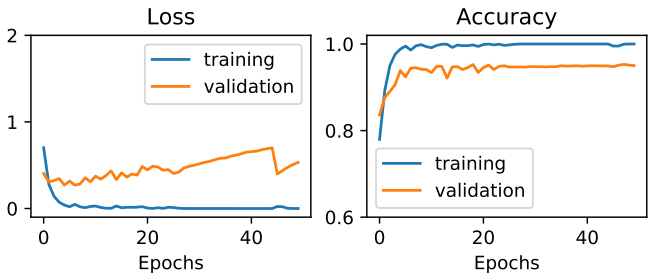
4.2.2 Model runs

Run 1: With and without data augmentation

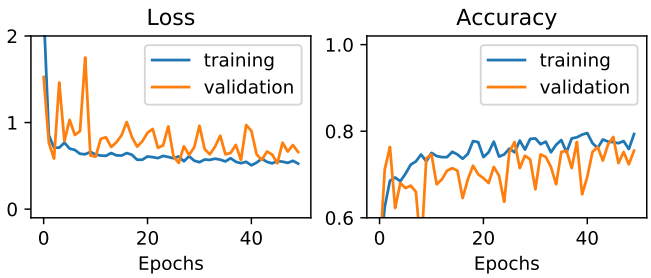
In the first run, a simple model consisting of only one convolutional layer is used to compare the training dataset with and without the two augmentation methods (see Fig. 4.3). Basically, the configuration shows rather less satisfying results. For the dataset without augmentation and manual augmentation a clear overfitting can be seen. With each further epoch the validation loss moves further away from the training loss and



(a) Wihout Augmentation



(b) Manuel Augmentation



(c) Augmentation Layer

Figure 4.3: Run 1: Display of loss and accuracy curves with and without data augmentation

the training accuracy is significantly higher than the validation accuracy.

The latter has a difference of about 0.2 to the training accuracy for model a) without augmentation, whereas the validation loss has a mean difference of about 0.67 to the training loss. In model b) manual augmentation, a small improvement can be observed. There is only an average difference of 0.05 between training and validation accuracy and only 0.43 between validation and training loss, which is an improvement of 75 % and 35 % respectively.

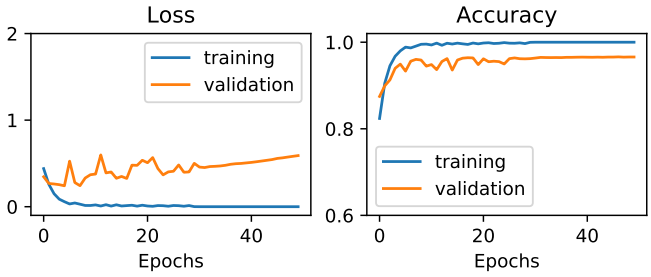
In the model c) with the augmentation layer dataset, the respective loss and accuracy curves seem to converge. The validation loss has an average difference of about 0.15 to the training loss and the validation accuracy has a mean difference of about 0.04 to the training accuracy and is thus the strongest model among the three. However, the validation values show a very high variance between the epochs. This could be due to a too high learning rate for this dataset, which updates the weights with a too high amplitude. In addition, it is possible that the variance in the training dataset is too high due to the augmentations, so that in one epoch the validation dataset matches the training dataset more or less. However, by augmenting the data, an improvement can be seen in both methods. The next runs will test whether an increase in the number of layers and the use of regularizers can improve the models (b) and (c).

Run 2: Three convolutional layer

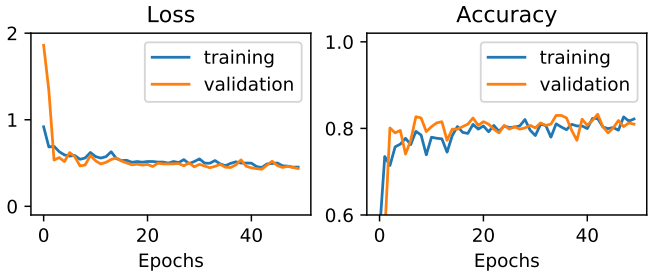
In Run 2, both augmentation methods are tested with three convolutional layers with an increasing number of filters 32, 64 and 128. There is still strong overfitting for the manual data augmentation and there is only a small improvement in validation loss and validation accuracy by around 0.014 for both parameters. The accuracy and loss curves from the validation set on the augmentation layer are significantly less noisy compared to Run 1 and have also converged somewhat more around a mean loss and accuracy of around 0.53 and 0.79, respectively. This could be due to the fact that by increasing the number of filters, more combinations of features could be recognized. Here, too, a slight improvement can be seen for both parameters. So far, it seems that data augmentation with the augmentation layer is significantly more effective than manual augmentation. With the following runs we check if an improvement of the models is possible by adding the regularizers: dropout, L2 and batch normalization.

Run 3-6: Dropout, L2 regularizer and batch normalization

Figure 4.5 shows the loss and accuracy curves of the manual data augmentation dataset for the Runs 3-6 and Figure 4.6 shows those of the data augmentation layer.



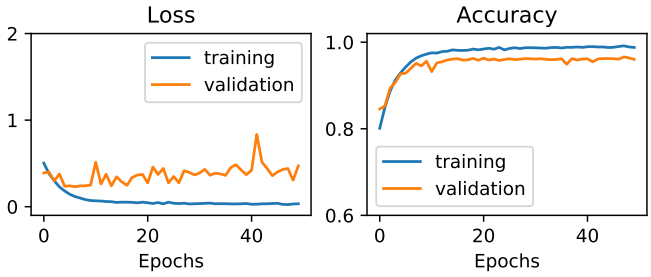
(a) Manuel Augmentation



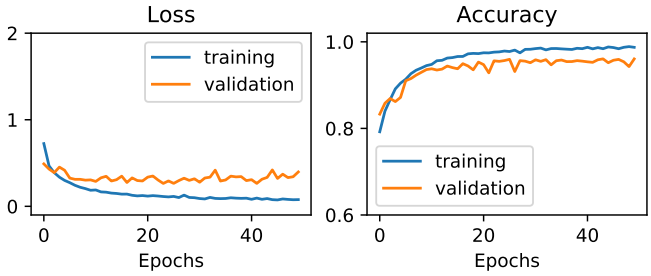
(b) Augmentation Layer

Figure 4.4: Run 2: Display of loss and accuracy curves with three convolutional layers.

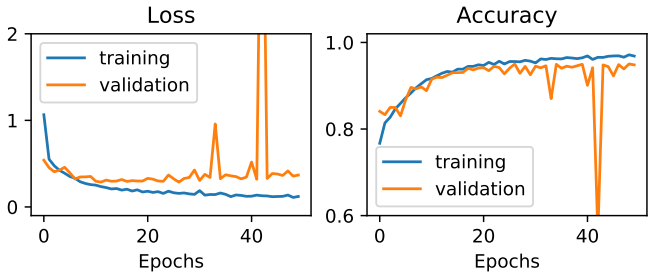
For the manual dataset, it is noticeable that for Run 3, just by adding dropout after the dense layer the accuracy curves are getting closer, and the validation loss curve is no longer rising as steeply. However, the overfitting still exists, but has decreased for the loss curves by a factor 28 % and for the accuracy curves by a factor of 37,7 % compared to Run 2. In Run 4, the overfit is reduced again slightly by adding the L2 regularizer. Stabilization can be seen in the validation loss in particular. The biggest change so far is the addition of batch normalization in Run 5. The accuracy curves are now much closer earlier to each other, but seem to diverge somewhat with increasing epoch, which again speaks for overfitting. In the loss, the distance has remained about the same, but it seems as if there are much higher peaks to be observed at the higher epochs for the validation curves. Especially in epoch 42 a very high deviation can be seen. Run 6 delivers so far the best result for the manual dataset. Here, dropout with the value 0.3 was added after each max pooling layer. This removes some complexity from the model by removing neurons and the training dataset has become closer to the validation dataset with a difference of (<0.01) between training and validation curves, which is a very good result and has eliminated the problem of overfitting.



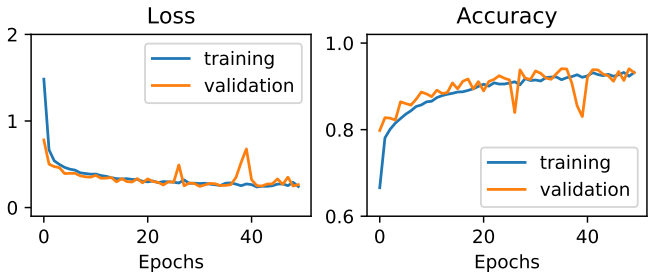
(a) Run 3



(b) Run 4



(c) Run 5



(d) Run 6

Figure 4.5: Run 3-6: Display of loss and accuracy curves of the manuel augmentation dataset for four different configurations.

The data augmentation layer dataset shows fewer changes and improvements between these runs. Only in the Runs 5 and 6 is it noticeable that the validation curves, especially the accuracy, are more fluctuating. This could be due to the fact that batch normalization is added from Run 5 onwards and therefore this variation can be observed. However, since the regularizations have hardly changed anything overall, it could be that the problem lies in the training dataset itself.

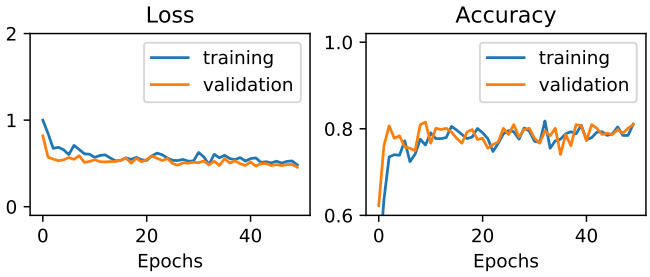
Choose best runs

Based on the results of the runs, only model runs that did not exhibit significant overfitting are considered for evaluation. At the epoch with the highest validation accuracy, these models predicted the test dataset and then output the test accuracy. Those model runs and their test accuracy are shown in Table 4.6.

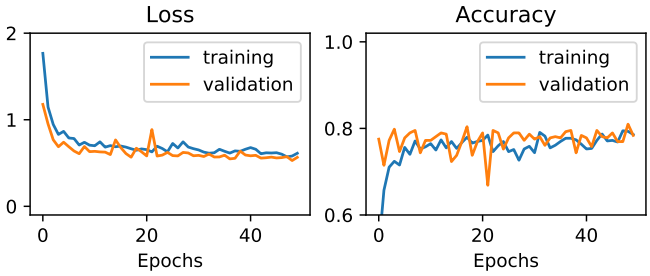
For the manual augmentation dataset, the Runs 5 and 6 are selected because all others show significant overfitting. Even though Run 5 still has some overfitting, this value was taken as a comparison value, because of all other model runs this one has the least overfitting. As expected, Run 6 shows the best test accuracy with 80.67 % as the configuration with the lowest overfitting. For the augmentation layer dataset, there are significantly more model runs to choose from. Here it is shown that the model from Run 2 has the best test accuracy with 80.61 %. This shows again that the regularization methods have no real influence on the result. Nevertheless, the other runs are not much worse. Since each setup is run only once, it is possible that Run 2 is the best by chance. However, this would not change the fact that all accuracys are between 75 % and 80 %. Overall, it is also very noticeable that in both cases, no much higher accuracy than 80 % can be achieved. This again maybe suggests a problem with the dataset itself.

Table 4.6: Test accuracy values for the best model runs for a) manual augmentation and b) augmentation layer.

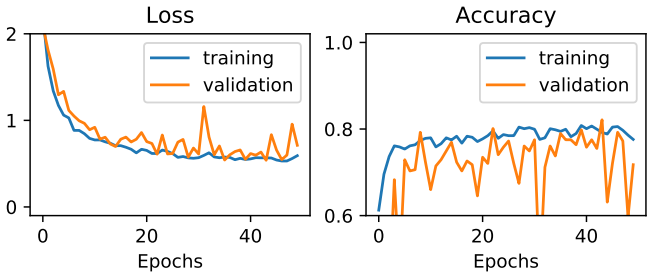
	a)		b)				
Runs	Run 5	Run 6	Run 2	Run 3	Run 4	Run 5	Run 6
Accuracy(%)	77.73	80.67	80.61	78.78	76.00	77.33	76.55



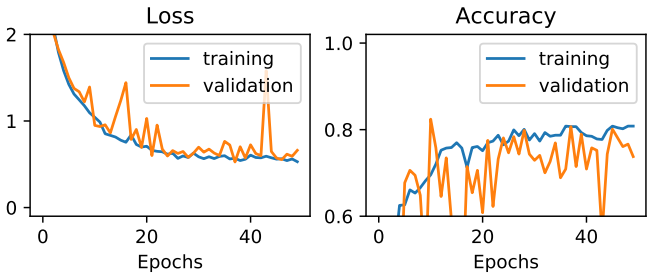
(a) Run 3



(b) Run 4



(c) Run 5



(d) Run 6

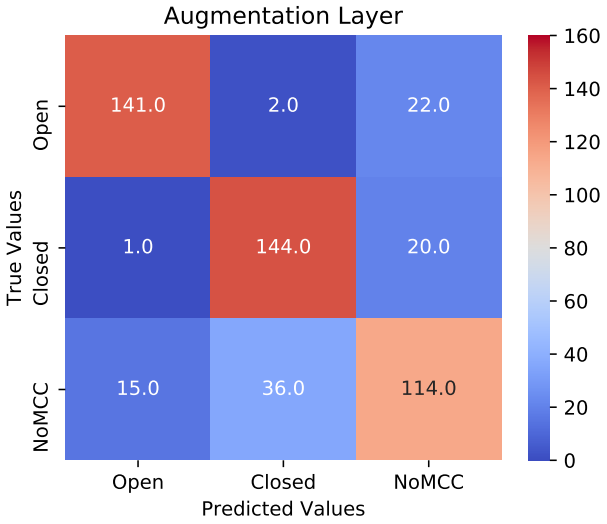
Figure 4.6: Run 3-6: Display of loss and accuracy curves of the augmentation layer dataset for four different configurations.

4.2.3 Confusion matrix evaluation

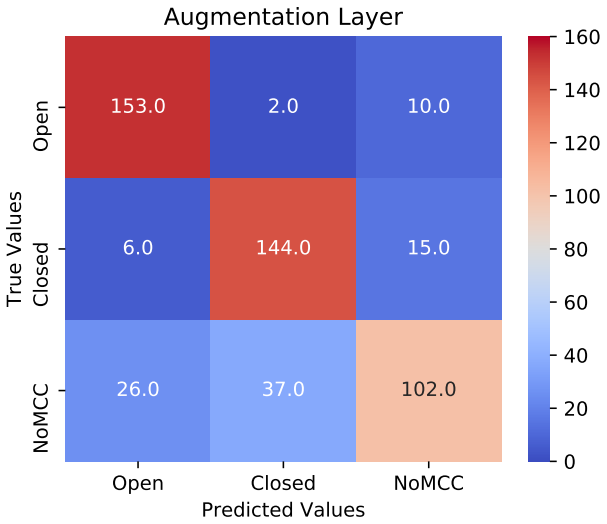
Figure 4.7 shows the confusion matrices for the best runs from the two augmentation methods. A confusion matrix clearly indicates the extent to which the model classified the images differently than the actual test dataset is classified by hand. By looking at the upper confusion matrix of Figure 4.7 the diagonal from top left to bottom right indicates the images correctly classified by the model. These are images that are also classified by the original test dataset as open, closed or no MCC. These are also called True Positive (TP). Each class has its own TP. The TP of the open images (141) can be seen in the upper left corner. There, in the same row, the open images are displayed that are incorrectly recognized by the model as closed (2) and no MCC (22). These are called false negatives (FN). The values in the same column are closed (1) and no MCC (15) images that are incorrectly assigned as open by another category and are called False Positive (FP). All remaining values in the confusion matrix are not related directly to the open category and are therefore named True Negative (TN). This procedure is analogous for the other categories starting from their TP point.

It is noticeable that most of the open and closed images are correctly assigned for both cases and the largest uncertainty is for the no MCC images. It can also be seen that both cases have different numbers when assigning the images, except for the closed images. There, both models predicted the same number. For further statistical analysis, both matrices are examined using the parameters precision, recall and the F1 score (Grandini et al., 2020).

The precision indicates for the respective class the ratio of its correctly assigned images (TP) to the total of all images of this class declared as positive. This includes those images that are incorrectly assigned to this class (FP) from other categories. Therefore, it says how precisely this class has been predicted, hence the name Precision. The recall, on the other hand, indicates how high the proportion of correctly classified images (TP) is in relation to the total number of images of the respective class. The total also includes images that were incorrectly assigned to a different class by the model (FN). If both parameters have high values, it also means that the model is good. A maximum high precision indicates that no image of another category is falsely predicted as such. A maximum high recall, on the other hand, indicates that no image has been incorrectly assigned to one of the others categories. The F1 score combines these two parameters and gives a statement about their harmonic mean. Again, a high value means a well-trained model. The respective formulas are described as follows.



(a)



(b)

Figure 4.7: Confusion matrix for (a) Augmentation Layer (80.61%) and (b) Manuel Augmentation (80.67%).

$$Precision_k = \frac{TP_k}{TP_k + FP_k} \tag{4.3}$$

$$Recall_k = \frac{TP_k}{TP_k + FN_k} \tag{4.4}$$

$$F1 - Score_k = 2 \cdot \frac{Precision_k \cdot Recall_k}{Precision_k + Recall_k} \tag{4.5}$$

Here, k stands for the respective category.

In Table 4.7 the precision, recall and F1 score for the respective class of each method can be seen. Basically, it can be said that for both cases the F1 score is almost equally distributed. The open images seem overall to have the best prediction, with precision being higher for a) and recall being higher for b). This means model a) assigns other categories less to the open category, but incorrectly assigns more open images to other categories. The closed category has exactly the same distribution of parameters for both cases and is slightly lower than the open images with an F1 score of 0.83. The no MCC F1 score is the worst of all the categories at about 0.70. However, the category is also distributed differently here for precision and recall. The models seem to have prediction problems especially in this category, which reinforces the assumption that no better model result can take place due to the dataset. Due to the weaker result in the no MCC category, the misattributed images is investigated in more detail.

Table 4.7: Presentation of the statistical evaluation of the two models and their categories through precision recall and F1 score.

	Precision	Recall	F1-Score
	a) Augmentation Layer		
Open	0.9	0.85	0.88
Closed	0.79	0.87	0.83
NoMCC	0.73	0.69	0.71
	b) Manuel Augmentation		
Open	0.83	0.93	0.87
Closed	0.79	0.87	0.83
NoMCC	0.80	0.62	0.70

A closer look at some images reveals certain patterns in these scenes which can be assigned to the "incorrectly" predicted classes open and closed. Figure 4.8 shows the no MCC images that are classified as closed and Figure 4.9 shows no MCC scenes that are assigned to the open cateogry.

For example, in the assigned closed category, images (a) and (e) show strong evidence of closed MCC, as typical bumps with margins can be seen. However, these margins are somewhat faint, so it may be that the classifiers have categorized this image as no MCC. Furthermore, in images (b), (c), (d) and (f), closed MCC structures can also be seen. For example in image (c) in the upper left half or in (b), where even

half of the image is covered with closed MCC clouds. The fact that these images have been categorized as no MCC is due to the fact that one of the classification conditions was that at least 70% of the image must be covered with open or closed in order to be categorized as such and since there is no special category for this kind of images, they are assigned to the no MCC category. Nevertheless, this apparently leads to a misprediction of the model, since these images still have open and closed structures, unlike the conventional no MCC images (see Fig. 4.10) that dominate the dataset.

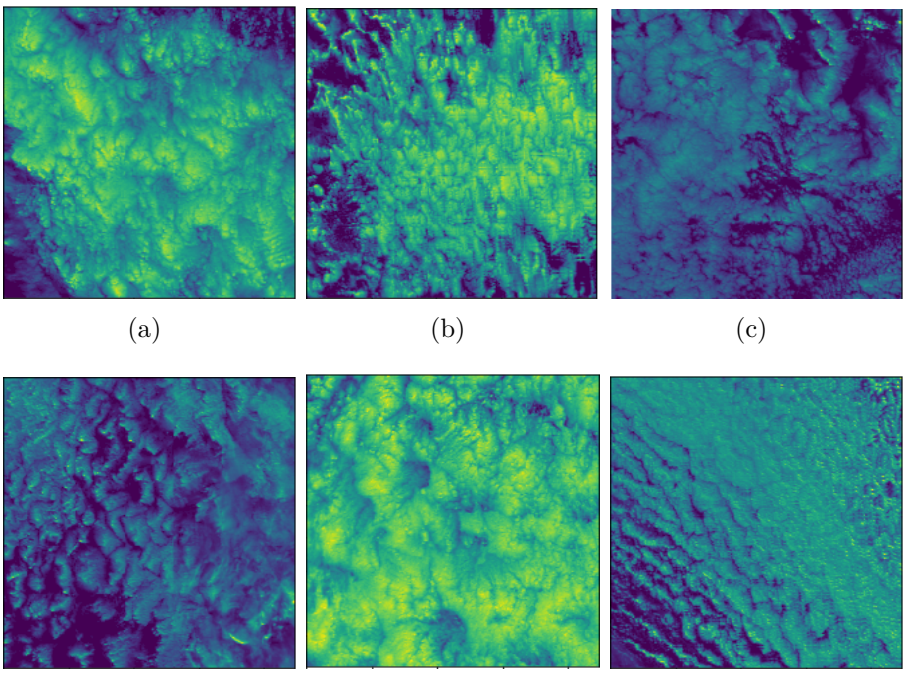
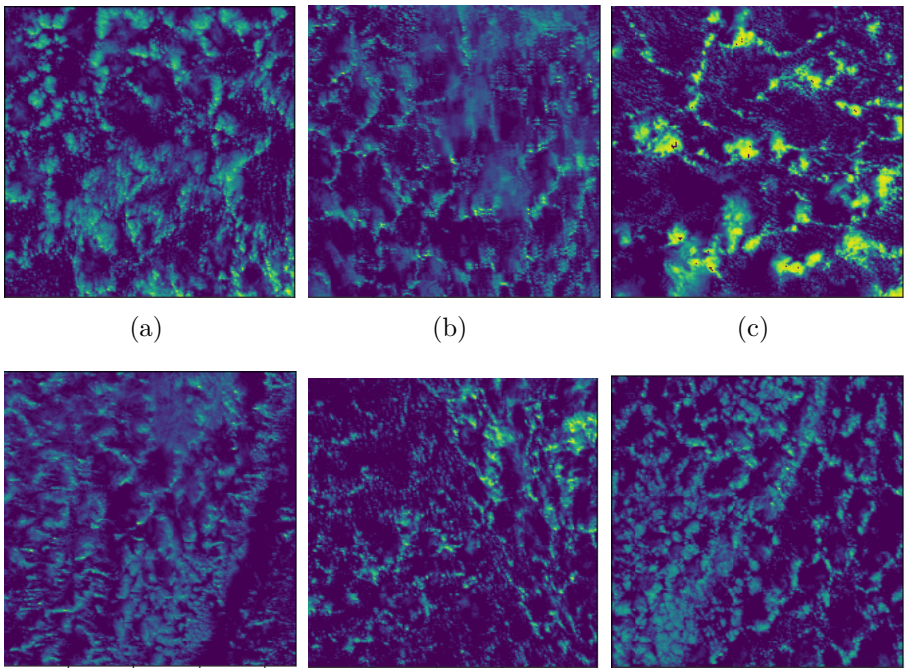


Figure 4.8: Display of some incorrectly predicted NoMCC images to the Closed category by both models.

The same can be observed when looking at the open images. Some of the structures appear somewhat blurred, but they also exhibit characteristics of open MCC, such as in (a), (b) and (d). In some cases, the cells also appear somewhat more separated, which suggests that they may belong to the disorganized category, which are by far the least classified category by the classifiers. Due to this uncertainty in classification, these structures are placed into the no MCC category, which nevertheless still contain open cell structures and therefore are "incorrectly" predicted by the model.

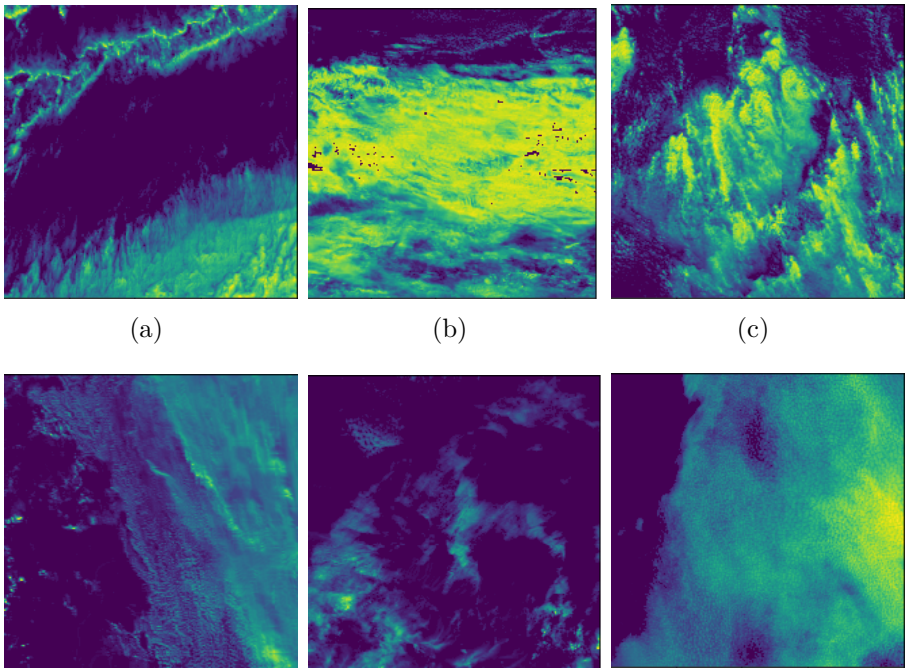
By adding more categories that include these ambiguous images, by removing them from the no MCC category or by reducing the image size of the input data in general, these mispredictions could be avoided and result in a better model performance.



(a) (b) (c)

(d) (e) (f)

Figure 4.9: Illustration of some incorrectly predicted NoMCC images to the category Open by both models.



(a) (b) (c)

(d) (e) (f)

Figure 4.10: Illustration of some corectly predicted NoMCC images by both models.

4.3 Comparison with MCC dataset

Table 4.8 shows the respective matches of the original test dataset and the predicted datasets with those of the MCC dataset. Basically, the open categories have the fewest matches for all datasets and the original dataset had the lowest number of all three. This is probably because the other predicted datasets have some no MCC images categorized as open, which in turn were also considered open by the other model. Accordingly, the manual augmentation dataset has the most matches to the other dataset with 14 open images. Likewise, the general low open match could also be due to the fact that the scenes classified as open by this work are partially classified as disorganized by the other algorithm.

Clearly more agreements are found with the closed category with numbers of roughly 40 matches. Here, too, slight differences can be seen in each dataset, which in turn can be attributed to the no MCC images that are predicted as closed, as well as to the different closed images of the respective predicted datasets that are classified as open or no MCC, for example.

The predicted dataset from the augmentation layer has the highest overall agreement with the MCC dataset with 54 matches. The other two are not far behind with 48 and 52, which is to be expected, since the predicted datasets match the test dataset by about 80 %.

Despite this, the number of points that did not have a data point within the radius and the number of points that did not meet the criterion to match outweighs for all

Table 4.8: Listing of the number of matches which were in the selected radius of the a) test dataset, b) predicted augmentation layer dataset and c) the predicted manuel augmentation dataset with MCC dataset, as well as the number of points that do not have a datapoint (No Datapoint) and the number of matches that did not satisfy the majority criterion (No Matches). The number of possible matches is derived from the available open and closed images with subsequent subtraction of the No Datapoints of the respective dataset. The number of available images for b) and c) results from the sum of the columns of the Confusion matrixes of the categories open and closed from Figure 4.7.

	a)	b)	c)
Open	8 (3.0 %)	11 (4.1 %)	14 (4.7 %)
Closed	40 (15.0 %)	43 (16.0 %)	38 (12.8 %)
Matches	48 (18.0 %)	54 (20.1 %)	52 (17.5 %)
No Datapoints	64	71	71
No Matches	218	214	245
Possible Matches	266	268	297

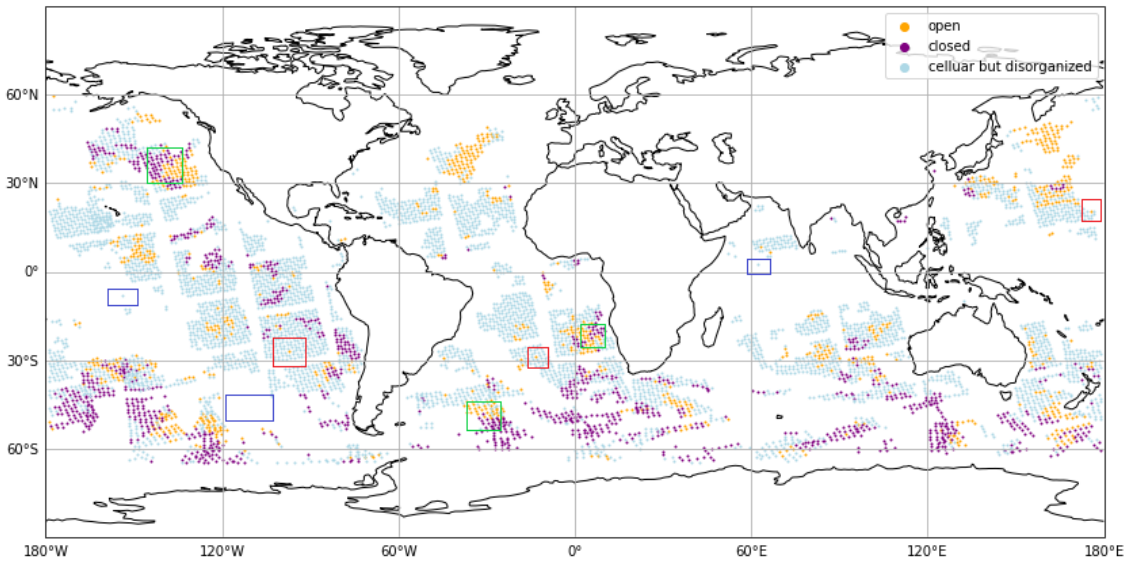


Figure 4.11: Global representation of the MCC data points on 27.08.2008 of the categories open, closed, and cellular but disorganized MCC clouds. Regions outlined in red are regions where a category point is surrounded by many other points of a different category. Regions outlined in green represent transition regions from closed to open MCC. Blue represents regions where no category occurs.

datasets. The possible reasons for this can be enumerated with the help of Figure 4.11. There the data points of the other dataset global for the 27.08.2008 are shown. The high number of no datapoints can be explained by the fact that possibly some datapoints from this work are located in the areas where no or few datapoints from the MCC dataset exist. These are outlined in blue as example areas. The images that did not meet the criterion (No Matches) may be due to two other reasons. For example, a point may have been very close to one of the other dataset points, but surrounded by points of another category, so that it did not count as a match with the necessary criterion. This scenario is illustrated with the red outlines in Figure 4.11. On the other hand, it is of course possible that the points are located in a transition region that can be seen with the green area, and thus cannot be clearly assigned. There it is possible that no majority at the criterion or a majority of another category has been reached, since our scenes have a different size and are therefore also slightly offset compared to the points of the MCC dataset.

Despite this analysis, it cannot be clearly determined whether the datasets match the MCC dataset or not and requires a different method of investigation. The main reasons for this are, first, that the comparison datasets are many times smaller than the MCC dataset and the matches are perhaps so low, simply due to random unfavorable position of the points and, second, the points are slightly offset due to the different

sizes of the scenes and do not lie exactly on the MCC data points. A possible approach here would be to classify scenes of the same size for a selected area to then allow a point-by-point comparison.

5 Discussion

In this chapter the most important results of this thesis will be discussed and compared to recent studies and possible sources of errors are pointed out.

During the scene classification, 19771 scenes are successfully classified as open, closed or no MCC. The disorganized category is removed because of a to low agreement among the classifiers. The statistical analysis shows a high agreement of 98 % among the classifiers under the condition that three or more assigned the same category to a scene. Even with the condition of four or more agreeing, a high level of agreement of 86 % is seen. But this is probably due to the fact that all images that could not be assigned as open or closed are automatically assigned to the category no MCC.

The climatology of MCC clouds shows large differences in the distribution of open and closed MCC clouds between winter and summer in the NA, SEP and SEI regions. In the summer, almost no open MCC clouds are identified and only 16 % of closed MCC clouds are identified in the SEP region in Summer. In the winter months it is noticeable that in the regions NA, SEI, SP which are located at higher latitudes, more open than closed MCC clouds are classified in the areas SEP and NEP (see Table 4.5) which are close to the subtropics. In particular, in the SP region, the fraction of open MCC clouds is six times larger than the fraction of closed MCC clouds.

The low occurrence of open and especially closed MCC clouds in the NA region in the summer months as well as the higher distribution of open compared to closed MCC clouds in the NA, SEI, and SP regions in the winter months can be explained by the observations of recent studies. Lang et al. (2022), Yuan et al. (2020), and Rampal and Davies (2020) show that there are fewer closed MCC clouds distributed at higher latitudes than assumed in previous studies from Muhlbauer et al. (2014) and McCoy et al. (2017). The higher frequency of occurrence of open MCC clouds in the winter months is likely due to MCAO's and cold air advection from continents favoring a closed to open transition (McCoy et al., 2017; Lang et al., 2022).

In the more subtropical regions SEP and NEP the distribution of open and closed MCC clouds can be related to the seasonal cycle of low cloud fraction and SST. The

fact that in winter in these regions about 50 % less open than closed MCC clouds are identified is due to an increased anti-correlation between the occurrence of closed and disorganized MCC clouds (Muhlbauer et al., 2014).

The difference in the observations of these studies may have several factors. For the scene identifaction Muhlbauer et al. (2014) and McCoy et al. (2017) used the ANN of Wood and Hartmann (2006) to identify MCC clouds. There the LWP is used as input for the network, whereas for example the inputs in Lang et al. (2022) and Rampal and Davies (2020) are based on spatial variables. This has the advantage that for the training of the NN no derivative product like the LWP is used, but a direct satellite product like the brightness temperature. Furthermore, LWP results are less reliable for mixed-phase clouds, which are found especially in colder regions and therefore at higher latitudes. Consequently, the differences are may be contributed due to this. Another difference is that in recent studies such as Lang et al. (2022), Rampal and Davies (2020), and Watson-Parris et al. (2021), a CNN was used, which is an extended ANN that is particularly suitable for feature recognition in spatial patterns due to the convolution operation. However, the differences between Muhlbauer et al. (2014) to this work may also be due to the slightly different selected regions, which partly differ in the range by 5 °- 10 °. Especially in the southern ocean only the region around the Southern Pacific is taken in this study. However, the similarities to the more recent studies may be due to the use of reflection as input, which is also a direct satellite product, but the similarities must also be viewed critically as the more recent studies do not cover exactly the same area as in this thesis.

The results of Chapter 4.2 show that the dataset has been successfully multiplied by data augmentation and thus can mimic the spatial and physical variability of MCC clouds. Furthermore, the augmented training datasets in the CNN model Run 1 show improvements over the non-augmented dataset (see Figure 4.3).

In Run 1 the manually augmented dataset shows reduced overfitting compared to the non-augmented dataset. In addition, the distance between the training and validation curves for loss and accuracy improved by 35 % and 75 %, respectively when comparing the non-augmented with the manually augmented model. Adding more layers in Run 2 still does not eliminate the overfitting, and show only an improvement of the validation curves by a value of about 0.014. In Run 3, adding dropout after dense layer shows a decreased gap with a reduced overfitting of 28 % for the loss curves and 37,7 % for the accuracy curves. In the model runs 4 and 5, adding L2 regularizer and batch normalization show only very slight improvements. Only by adding dropout after each max pooling layer in Run 6 did the loss and accuracy curves converge successfully

with difference less than 0.01 between training and validation curves. The observed improvement is probably due to the fact that by removing neurons via dropout, complexity has been taken away from the model and thus could be better converged to the validation dataset.

In the dataset with the augmentation layer, the curves seem to converge directly in model Run 1. However, there is still a slight difference between the training and validation curves. In addition, there is also a high variance of the validation curves between the epochs. This is due to a slightly too high learning rate or the variance generated by the augmentation in the training dataset. Adding additional convolutional layers with an increased number of filters seems to give the best result for the augmentation layer dataset. There, the loss and accuracy curves converge within the first epochs. The loss and accuracy curves stabilize at about 0.53 and 0.79, respectively. By adding the regularizations dropout, batch normalization, and L2 regularizer the curves in the Runs 3 - 6 have barely changed and no improvement can be seen. By adding batch normalization, the validation curves seem to have even a higher variance.

For the prediction of the test dataset with the condition that no strong overfitting is to be recognized, the model Runs 5 and 6 qualify for the manual data augmentation dataset and the model Runs 2 - 6 for the augmentation layer dataset (see Table 4.6). Among all predictions, a prediction accuracy in the range of 76 % - 80 % is observed, and for both augmentation methods the highest accuracy is around 80.6 %.

A closer look at the statistical evaluation of the best models of each augmentation method of a confusion matrix reveals particularly good average accuracies for the categories open and closed with an average of 87 % compared to the no MCC category with a value of 69 % for the augmentation layer and 62 % for the manual augmentation (see Table 4.7). It is also striking that especially many no MCC images are assigned to the closed category, which results in precision values of 79 % for the closed MCC categories in both manual and automatic augmentation and thus results in slightly lower F1 sores compared to the corresponding open MCC categories.

A closer look at the misclassified no MCC images into the categories open and closed also shows that some of these images have open and closed patterns, but were not assigned as such category by the classifiers because they did not meet the classification conditions. Therefore, by sorting out these images from the no MCC dataset, it might be possible to achieve a higher/better accuracy for the no MCC category. Another approach would be to reduce the classification conditions somewhat, so that when a new dataset is created, these are not classified in the no MCC cateogory in the first place. This can also be avoided by implementing new categories in which such images are then classified, or by dividing the images into four zones, in which each zone is

assigned a category and if a category is in the majority, the image is categorized as such.

Comparing the datasets with the MCC dataset, the open and closed images show that only about 18.5 % of the possible images matched the MCC dataset on average (see 4.8). In addition, there is a high number of about 68 points where no point from the MCC dataset is found at all. However, this low agreement with the MCC dataset is not significant enough, since the datasets of this thesis are significantly smaller than those of the MCC dataset. Furthermore, the scene centers are slightly shifted compared to the MCC dataset, so that especially in transition regions a miscomparison takes place. A more precise agreement with the MCC dataset could therefore be achieved by a scene selection in which the scenes of all datasets lie exactly on top of each other.

In the following, the best accuracy rates of this thesis are compared with other studies that have also used a neural network to identify open and closed MCC clouds.

Wood and Hartmann (2006) for example achieved an overall accuracy of 85 %- 90 % by using a three-layer artificial neural network with back propagation, which shows about 5 %- 10 % stronger accuracy than this thesis. As input data, they used 256 x 256 km^2 large scenes with the LWP as input variable to identify open, closed, and disorganized MCC clouds. However, if only the open and closed accuracies are compared, they are very close to those of Wood and Hartmann (2006) with 88 %.

In the study of Yuan et al. (2020), 128 x 128 pixel-sized scenes are successfully used to train a CNN using a reflectance in the range of 0.55 μm. In addition, several thousand images have been classified into six categories instead of three. Furthermore, a pre-trained model (VGG16) based on the ImageNet dataset was used. The advantage of this is that certain patterns are already recognized by the pre-trained model and thus simplified learning can take place. Altogether, this resulted in an overall accuracy of 93 % .

One of the recent MCC cloud classification studies of Lang et al. (2022) is built on training a hybrid CNN model based on the brightness temperature of channel 11 (8.6 μm) as main input and solar and satellite angles of the advanced Himawari Imager of the Himawari-8 geostationary satellite. And here, too, a high overall accuracy of 94 % was achieved with only three categories, which are open MCC, closed MCC, and others. It is particularly noticeable that Lang et al. (2022), in contrast to this thesis, can show a very high accuracy of 98 % for the category "others". This could be due to the fact that the input scenes of the brightness temperature array with a size of 84 x 84 km^2 are much smaller than in this thesis (224 x 224 km^2) and therefore contain less no MCC images with open and closed cell sctructures.

In summary, recent studies from the last few years have shown better results in terms of MCC detection when using a CNN over a regular NN as in the study Wood and Hartmann (2006). This may also be due to other differences like the use of direct satellite variables like reflectance or brightness temperature in contrast to the calculated LWP. Furthermore, much smaller scenes have been chosen. Thus, by reducing the size of the scenes used in this thesis, or by adding more categories, the misinterpretation of the model with respect to no MCC images might be reduced.

6 Summary and conclusion

The goal of this thesis was to create a CNN to classify open, closed, disorganized, and no MCC clouds by their reflectance.

To train the CNN, $224\,\mathrm{x}\,224\,\mathrm{km}^2$ sized AQUA satellite images based on the 2008 and 2009 MODIS data were used. Subsequently, these scenes have been classified by five classifiers into open, closed, disorganized, and no MCC. Due to an insufficient number of disorganized images because of a lack of agreement among the classifiers, the disorganized category has been removed from the dataset. After this, a training and test dataset was then formed from this hand-classified sample.

In addition, two augmented datasets were created based on this training dataset, one using a manual augmentation and the other using an augmentation layer built into the model.

Subsequently, these two augmented training datasets were used to train six different model configurations. It was found that the models based on the augmentation layer dataset had a constant validation accuracy of about 80 % even with much simpler model configurations, which only include an increase of the number of convolutional layer. Further configurations could not improve this result.

The models with the manually augmented dataset only showed a desirable result in the final model configuration, which consisted of a combination of the regularization methods dropout, batch normalization, and L2 regularization.

A prediction of the best models on the test dataset of the respective augmentation method showed a test accuracy of 81 % for both models. By evaluating the two model predictions using a confusion matrix , it was found that both models had good prediction accuracies (88 %) with respect to the open and closed categories, but had difficulties with the classification of the no MCC category. A closer examination of the no MCC scenes revealed that some of the images classified into no MCC had features of open or closed cell structures, which is a possible explanation for the difficulty of the models to classify them correctly.

To avoid this, it would be possible in future work to reclassify the no MCC dataset

to remove all scenes with obvious open and closed MCC cloud structures. Furthermore, it would also be possible to generate smaller scenes from the existing images that are still large enough to recognize the necessary structures, but small enough so that there are no different structures in one image. A different approach could be to introduce additional categories for scenes with features of more than one type of MCC clouds.

In conclusion, the results of this thesis show that, on the one hand, with the help of data augmentation and different model configurations, a CNN was successfully trained, which is able to recognize open and closed MCC clouds. On the other hand, the results also show that by improving the dataset, a possibly better model accuracy can be achieved.

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Anna Possner, for providing me with this interesting topic. I am also very grateful for her advice and guidance during this thesis. Furthermore, my special thanks goes to Jessica Danker. She accompanied and supported me in the development of this thesis from the beginning, no matter if it was technical or scientific questions regarding my topic and always gave me regular feedback. My gratitude also goes to Prof. Dr. Bodo Ahrens, who has taken over the co-supervision for my thesis.

Special thanks goes to Markus Ernst from the Frankfurt Institute for Advanced Studies, who was always available for questions concerning neural networks and without his advice the results would not be present to the extent they are now.

I would also like to thank my friend Kilian Schnelle, who provided a tool to classify the scenes for this thesis and who was always able to find a suitable solution to my problems concerning python.

Furthermore, I would also like to thank all the friends I was able to make during and through my studies. This has created great times both inside and outside of the university. Special thanks goes to my friend and fellow student Kevin Pfannkuch. Through his endless scientific enthusiasm for meteorology and physics, he knew how to motivate me even in bad times. For this I am very grateful and without him I probably would not be where I am now. At last, I would like to thank my family, who have always supported me during my studies and Ricarda Sies for encouraging me during this thesis.

Bibliography

A Blog on Building Machine Learning Solutions (2021). URL: https://programmathically.com/an-introduction-to-neural-network-loss-functions/ (Accessed on: 02 Mar. 2022).

Abadi, M. et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. DOI: 10.48550/arXiv.1603.04467.

Abel, S. J., I. A. Boutle, K. Waite, S. Fox, P. R. A. Brown, R. Cotton, G. Lloyd, T. W. Choularton, and K. N. Bower (2017). "The Role of Precipitation in Controlling the Transition from Stratocumulus to Cumulus Clouds in a Northern Hemisphere Cold-Air Outbreak". In: *Journal of the Atmospheric Sciences* 74.7, pp. 2293–2314. ISSN: 0022-4928. DOI: 10.1175/JAS-D-16-0362.1.

Agee, E. M., T. S. Chen, and K. E. Dowell (1973). "A Review of Mesoscale Cellular Convection". In: *Bulletin of the American Meteorological Society* 54.10, pp. 1004–1012. ISSN: 0003-0007. DOI: 10.1175/1520-0477(1973)054<1004:AROMCC>2.0.CO;2.

Analytics Vidhya (2020a). In: URL: https://medium.com/analytics-vidhya/how-to-choose-the-size-of-the-convolution-filter-or-kernel-size-for-cnn-86a55a1e2d15 (Accessed on: 31 Mar. 2022).

— (2020b). In: *Analytics Vidhya*. URL: https://medium.com/analytics-vidhya/significance-of-kernel-size-200d769aecb1 (Accessed on: 31 Mar. 2022).

— (2021). URL: https://www.analyticsvidhya.com/blog/2021/04/introduction-to-softmax-for-neural-network/ (Accessed on: 01 Mar. 2022).

Arking, A. (1991). "The Radiative Effects of Clouds and their Impact on Climate". In: *Bulletin of the American Meteorological Society* 72.6, pp. 795–813. ISSN: 0003-0007. DOI: 10.1175/1520-0477(1991)072%3C0795:TREOCA%3E2.0.CO;2.

Atkinson, B. W. and J. Wu Zhang (1996). "Mesoscale shallow convection in the atmosphere". In: *Reviews of Geophysics* 34.4, pp. 403–431. ISSN: 87551209. DOI: 10.1029/96RG02623.

Baeldung on Computer Science (2021). In: URL: https://www.baeldung.com/cs/batch-normalization-cnn (Accessed on: 02 Mar. 2022).

Berner, A. H., C. S. Bretherton, R. Wood, and A. Muhlbauer (2013). "Marine boundary layer cloud regimes and POC formation in a CRM coupled to a bulk aerosol scheme". In: *Atmospheric Chemistry and Physics* 13.24, pp. 12549–12572. DOI: 10.5194/acp-13-12549-2013.

Bishop, C. M. (2016). *Pattern Recognition and Machine Learning*. Softcover reprint of the original 1st edition 2006 (corrected at 8th printing 2009). Information Science and Statistics. New York, NY: Springer New York. ISBN: 9781493938438.

Bretherton, C. S. and M. C. Wyant (1997). "Moisture Transport, Lower-Tropospheric Stability, and Decoupling of Cloud-Topped Boundary Layers". In: *Journal of the Atmospheric Sciences* 54.1, pp. 148–167. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1997)054<0148:MTLTSA>2.0.CO;2.

Brownlee, J. (2020). In: *Machine Learning Mastery*. URL: https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/ (Accessed on: 30 Mar. 2022).

— (2021). In: *Machine Learning Mastery*. URL: https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/ (Accessed on: 31 Mar. 2022).

Cahalan, R. F., W. Ridgway, W. J. Wiscombe, T. L. Bell, and J. B. Snider (1994). "The Albedo of Fractal Stratocumulus Clouds". In: *Journal of the Atmospheric Sciences* 51.16, pp. 2434–2455. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1994)051%3C2434:TAOFSC%3E2.0.CO;2.

Caldwell, P., C. S. Bretherton, and R. Wood (2005). "Mixed-Layer Budget Analysis of the Diurnal Cycle of Entrainment in Southeast Pacific Stratocumulus". In: *Journal of the Atmospheric Sciences* 62.10, pp. 3775–3791. ISSN: 0022-4928. DOI: 10.1175/JAS3561.1.

Chen, T., W. B. Rossow, and Y. Zhang (2000). "Radiative Effects of Cloud-Type Variations". In: *Journal of Climate* 13.1, pp. 264–286. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2000)013<0264:REOCTV>2.0.CO;2.

Comstock, K. K., S. E. Yuter, R. Wood, and C. S. Bretherton (2007). "The Three-Dimensional Structure and Kinematics of Drizzling Stratocumulus". In: *Monthly Weather Review* 135.11, pp. 3767–3784. ISSN: 0027-0644. DOI: 10.1175/2007MWR1944.1.

Davim, J. P. (2011). *Machining of Hard Materials*. London: Springer London. ISBN: 978-1-84996-449-4. DOI: 10.1007/978-1-84996-450-0.

Fletcher, J., S. Mason, and C. Jakob (2016). "The Climatology, Meteorology, and Boundary Layer Structure of Marine Cold Air Outbreaks in Both Hemispheres*". In: *Journal of Climate* 29.6, pp. 1999–2014. ISSN: 0894-8755. DOI: 10.1175/JCLI-D-15-0268.1.

Getling, A. V. (1998). *Rayleigh-Bénard Convection*. Vol. 11. WORLD SCIENTIFIC. ISBN: 978-981-02-2657-2. DOI: 10.1142/3097.

Goodfellow, I., Y. Bengio, and Courville A. (2016). *Deep Learning*. MIT Press.

Grandini, M., E. Bagli, and G. Visani (2020). *Metrics for Multi-Class Classification: an Overview*. DOI: 10.48550/arXiv.2008.05756.

Habibi Aghdam, H. (2017). *Guide to Convolutional Neural Networks: A Practical Application to Traffic-Sign Detection and Classification*. Cham: Springer International Publishing. ISBN: 978-3-319-57549-0. DOI: 10.1007/978-3-319-57550-6.

Hahn, C. J., W. B. Rossow, and S. G. Warren (2001). "ISCCP Cloud Properties Associated with Standard Cloud Types Identified in Individual Surface Observations". In: *Journal of Climate* 14.1, pp. 11–28. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2001)014<0011:ICPAWS>2.0.CO;2.

Halevy, A., P. Norvig, and F. Pereira (2009). "The Unreasonable Effectiveness of Data". In: *IEEE Intelligent Systems* 24.2, pp. 8–12. ISSN: 1541-1672. DOI: 10.1109/MIS.2009.36.

Hartmann, D. L., M. E. Ockert-Bell, and M. L. Michelsen (1992). "The Effect of Cloud Type on Earth's Energy Balance: Global Analysis". In: *Journal of Climate* 5.11, pp. 1281–1304. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2.

Hastie, T., R. Tibshirani, and J. H. Friedman (2017). *The elements of statistical learning: Data mining, inference, and prediction*. Second edition. Springer Series in Statistics. New York, NY: Springer. ISBN: 9780387848587.

IBM (2021). In: URL: https://www.ibm.com/cloud/learn/overfitting (Accessed on: 02 Mar. 2022).

KDnuggets (2020). URL: https://www.kdnuggets.com/2020/12/optimization-algorithms-neural-networks.html (Accessed on: 02 Mar. 2022).

Klein, S. A. (1997). "Synoptic Variability of Low-Cloud Properties and Meteorological Parameters in the Subtropical Trade Wind Boundary Layer". In: *Journal of Climate* 10.8, pp. 2018–2039. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1997)010<2018:SVOLCP>2.0.CO;2.

Klein, S. A. and D. L. Hartmann (1993). "The Seasonal Cycle of Low Stratiform Clouds". In: *Journal of Climate* 6.8, pp. 1587–1606. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1993)006%3C1587:TSCOLS%3E2.0.CO;2.

Klein, S. A., D. L. Hartmann, and J. R. Norris (1995). "On the Relationships among Low-Cloud Structure, Sea Surface Temperature, and Atmospheric Circulation in the Summertime Northeast Pacific". In: *Journal of Climate* 8.5, pp. 1140–1155. ISSN: 0894-8755. DOI: 10.1175/1520-0442(1995)008<1140:OTRALC>2.0.CO;2.

Klose, B. (2015). *Meteorologie: Eine interdisziplinäre Einführung in die Physik der Atmosphäre*. 2. Aufl. 2015. Springer eBook Collection. Berlin, Heidelberg: Springer Spektrum. ISBN: 9783662435786. DOI: 10.1007/978-3-662-43578-6.

Korolev, A., G. McFarquhar, P. R. Field, C. Franklin, P. Lawson, Z. Wang, E. Williams, S. J. Abel, D. Axisa, S. Borrmann, J. Crosier, J. Fugal, M. Krämer, U. Lohmann, O. Schlenczek, M. Schnaiter, and M. Wendisch (2017). "Mixed-Phase Clouds: Progress and Challenges". In: *Meteorological Monographs* 58, pp. 5.1–5.50. ISSN: 0065-9401. DOI: 10.1175/AMSMONOGRAPHS-D-17-0001.1.

Kraus, H. (2004). *Die Atmosphäre der Erde: Eine Einführung in die Meteorologie*. Springer eBook Collection Life Science & basic disciplines. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 9783540350170. DOI: 10.1007/3-540-35017-9.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2017). "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6, pp. 84–90. ISSN: 0001-0782. DOI: 10.1145/3065386.

Lang, F., L. Ackermann, Y. Huang, S. C. H. Truong, S. T. Siems, and M. J. Manton (2022). "A climatology of open and closed mesoscale cellular convection over the Southern Ocean derived from Himawari-8 observations". In: *Atmospheric Chemistry and Physics* 22.3, pp. 2135–2152. DOI: 10.5194/acp-22-2135-2022.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner (1998). "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11, pp. 2278–2324. ISSN: 00189219. DOI: 10.1109/5.726791.

Lilly, D. K. (1968). "Models of cloud-topped mixed layers under a strong inversion". In: *Quarterly Journal of the Royal Meteorological Society* 94.401, pp. 292–309. ISSN: 0035-9009. DOI: 10.1002/qj.49709440106.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong (2009). "Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget". In: *Journal of Climate* 22.3, pp. 748–766. ISSN: 0894-8755. DOI: 10.1175/2008JCLI2637.1.

Logan, B. and Miller Denise (3.07.2013). *NASA - The Role of Clouds*. URL: https://www.nasa.gov/audience/forstudents/5-8/features/F_The_Role_of_Clouds.html (Accessed on: 14 Mar. 2022).

Making Developers Awesome at Machine Learning (2020). In: *Machine Learning Mastery*. URL: https://machinelearningmastery.com/pooling-layers-for-convolutional-neural-networks/ (Accessed on: 02 Mar. 2022).

Malberg, H. (2007). *Meteorologie und Klimatologie: Eine Einführung*. 5., erw. u. aktualisierte Aufl. 2007. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN: 9783540372226.

Max-pooling / Pooling - Computer Science Wiki (26.02.2022). URL: https://computersciencewiki.org/index.php/Max-pooling_/_Pooling (Accessed on: 02 Mar. 2022).

McCoy, I. L., R. Wood, and J. K. Fletcher (2017). "Identifying Meteorological Controls on Open and Closed Mesoscale Cellular Convection Associated with Marine Cold Air Outbreaks". In: *Journal of Geophysical Research: Atmospheres* 122.21, pp. 11,678–11,702. ISSN: 2169897X. DOI: 10.1002/2017JD027031.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". In: *The bulletin of mathematical biophysics* 5.4, pp. 115–133. ISSN: 1522-9602. DOI: 10.1007/BF02478259.

Miller, R. and B. McKee (2004). "Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters". In: *Remote Sensing of Environment* 93.1, pp. 259–266. ISSN: 0034-4257. DOI: 10.1016/j.rse.2004.07.012.

Moeng, C.-H., S. Shen, and D. A. Randall (1992). "Physical Processes within the Nocturnal Stratus-topped Boundary Layer". In: *Journal of the Atmospheric Sciences* 49.24, pp. 2384–2401. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1992)049%3C2384:PPWTNS%3E2.0.CO;2.

Mohrmann, J., R. Wood, T. Yuan, H. Song, R. Eastman, and L. Oreopoulos (2021). "Identifying meteorological influences on marine low-cloud mesoscale morphology using satellite classifications". In: *Atmospheric Chemistry and Physics* 21.12, pp. 9629–9642. DOI: 10.5194/acp-21-9629-2021.

Muhlbauer, A., I. L. McCoy, and R. Wood (2014). "Climatology of stratocumulus cloud morphologies: microphysical properties and radiative effects". In: *Atmospheric Chemistry and Physics* 14.13, pp. 6695–6716. DOI: 10.5194/acp-14-6695-2014.

neptuneblog (2021). URL: https://neptune.ai/blog/fighting-overfitting-with-l1-or-l2-regularization (Accessed on: 02 Mar. 2022).

Nicholls, S. (1984). "The dynamics of stratocumulus: Aircraft observations and comparisons with a mixed layer model". In: *Quarterly Journal of the Royal Meteorological Society* 110.466, pp. 783–820. ISSN: 0035-9009. DOI: 10.1002/qj.49711046603.

Norris, J. R. and S. A. Klein (2000). "Low Cloud Type over the Ocean from Surface Observations. Part III: Relationship to Vertical Motion and the Regional Surface Synoptic Environment". In: *Journal of Climate* 13.1, pp. 245–256. ISSN: 0894-8755. DOI: 10.1175/1520-0442(2000)013<0245:LCTOTO>2.0.CO;2.

Park, S. and N. Kwak (2017). "Analysis on the Dropout Effect in Convolutional Neural Networks". In: *Computer Vision – ACCV 2016*. Ed. by S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato. Vol. 10112. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 189–204. ISBN: 978-3-319-54183-9. DOI: 10.1007/978-3-319-54184-6_12.

Parkinson, C. L. and R. Greenstone (2000). In: *NASA's Earth Observing Syste,* (Accessed on: 01 Mar. 2022).

Perez, L. and J. Wang (2017). *The Effectiveness of Data Augmentation in Image Classification using Deep Learning*. DOI: 10.48550/arXiv.1712.04621.

Petters, M. D., J. R. Snider, B. Stevens, G. Vali, I. Faloona, and L. M. Russell (2006). "Accumulation mode aerosol, pockets of open cells, and particle nucleation in the remote subtropical Pacific marine boundary layer". In: *Journal of Geophysical Research* 111.D2. ISSN: 0148-0227. DOI: 10.1029/2004JD005694.

Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey (2003). "The MODIS cloud products: algorithms and examples from terra". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2, pp. 459–473. ISSN: 0196-2892. DOI: 10.1109/TGRS.2002.808301.

Rampal, N. and R. Davies (2020). "On the Factors That Determine Boundary Layer Albedo". In: *Journal of Geophysical Research: Atmospheres* 125.15. ISSN: 2169897X. DOI: 10.1029/2019JD032244.

Roach, W. T., R. Brown, S. J. Caughey, B. A. Crease, and A. Slingo (1982). "A field study of nocturnal stratocumulus: I. Mean structure and budgets". In: *Quarterly Journal of the Royal Meteorological Society* 108.455, pp. 103–123. ISSN: 0035-9009. DOI: 10.1002/qj.49710845507.

Rogers, D. P. and D. Koracin (1992). "Radiative Transfer and Turbulence in the Cloud-topped Marine Atmospheric Boundary Layer". In: *Journal of the Atmospheric Sciences* 49.16, pp. 1473–1486. ISSN: 0022-4928. DOI: 10.1175/1520-0469(1992)049%3C1473:RTATIT%3E2.0.CO;2.

Sandu, I. and B. Stevens (2011). "On the Factors Modulating the Stratocumulus to Cumulus Transitions". In: *Journal of the Atmospheric Sciences* 68.9, pp. 1865–1881. ISSN: 0022-4928. DOI: 10.1175/2011JAS3614.1.

Schnelle, K. (2020). In: *GitHub-Repository* GitHub. URL: https://github.com/KiSchnelle/image_labeling.

Schoeberl, M. R. (2002). "The afternoon constellation: a formation of Earth observing systems for the atmosphere and hydrosphere". In: *IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 354–356. ISBN: 0-7803-7536-X. DOI: 10.1109/IGARSS.2002.1025038.

Schönwiese, C.-D. (2013). *Klimatologie: 31 Tabellen im Text und umfangreicher Tabellenanhang*. 4., überarb. und aktualisierte Aufl. Vol. 1793. UTB Geowissenschaften, Agrarwissenschaften, Biologie, Physik. Stuttgart: Ulmer. ISBN: 9783825239008.

Sharon, T. M., B. A. Albrecht, H. H. Jonsson, P. Minnis, M. M. Khaiyer, T. M. van Reken, J. Seinfeld, and R. Flagan (2006). "Aerosol and Cloud Microphysical Characteristics of Rifts and Gradients in Maritime Stratocumulus Clouds". In: *Journal of the Atmospheric Sciences* 63.3, pp. 983–997. ISSN: 0022-4928. DOI: 10.1175/JAS3667.1.

Simonyan, K. and A. Zisserman (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. DOI: 10.48550/arXiv.1409.1556.

Srivastava, N., Hinton. G, A. Krizhevsky, I. Sutskever, and Salakhutdinov R. (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15.56, pp. 1929–1958. ISSN: 1533-7928.

Stevens, B., S. Bony, H. Brogniez, L. Hentgen, C. Hohenegger, C. Kiemle, T. S. L'Ecuyer, A. K. Naumann, H. Schulz, P. A. Siebesma, J. Vial, D. M. Winker, and P. Zuidema (2020). "Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds". In: *Quarterly Journal of the Royal Meteorological Society* 146.726, pp. 141–152. ISSN: 0035-9009. DOI: 10.1002/qj.3662.

Stevens, B., G. Vali, K. Comstock, R. Wood, M. C. van Zanten, P. H. Austin, C. S. Bretherton, and D. H. Lenschow (2005). "POCKETS OF OPEN CELLS AND DRIZZLE IN MARINE STRATOCUMULUS". In: *Bulletin of the American Meteorological Society* 86.1, pp. 51–58. ISSN: 0003-0007. DOI: 10.1175/BAMS-86-1-51.

Stocker, T. F., G.-K. P. D. Qin, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley (2013). "IPCC, 2013: Climate Change 2013". In:

"Clouds and Aerosols" (2014). In: *Climate change 2013*. Ed. by T. Stocker. Cambridge: Cambridge University Press, pp. 571–658. ISBN: 9781107415324. DOI: 10.1017/CBO9781107415324.016.

Sun, C., A. Shrivastava, S. Singh, and A. Gupta (2017). *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era*. DOI: 10.48550/arXiv.1707.02968.

Towards Data Science (2017). In: URL: https://towardsdatascience.com/l1-and-l2-regularization-methods-ce25e7fc831c (Accessed on: 02 Mar. 2022).

— (2020a). In: URL: https://towardsdatascience.com/overview-of-various-optimizers-in-neural-networks-17c1be2df6d5 (Accessed on: 02 Mar. 2022).

— (2020b). In: URL: https://towardsdatascience.com/introduction-to-convolutional-neural-network-cnn-de73f69c5b83 (Accessed on: 31 Mar. 2022).

— (2021). In: URL: https://towardsdatascience.com/explain-like-im-five-artificial-neurons-b7c475b56189 (Accessed on: 01 Mar. 2022).

v7 (2022). (Accessed on: 09 Apr. 2022).

vanZanten, M. C. and B. Stevens (2005). "Observations of the Structure of Heavily Precipitating Marine Stratocumulus". In: *Journal of the Atmospheric Sciences* 62.12, pp. 4327–4342. ISSN: 0022-4928. DOI: 10.1175/JAS3611.1.

Watson-Parris, D., S. A. Sutherland, M. W. Christensen, R. Eastman, and P. Stier (2021). "A Large–Scale Analysis of Pockets of Open Cells and Their Radiative Impact". In: *Geophysical Research Letters* 48.6. ISSN: 0094-8276. DOI: 10.1029/2020GL092213.

Werbos, P. J. (1994). *The roots of backpropagation: From ordered derivatives to neural networks and political forecasting: Zugl.: Harvard, Univ., Diss., 1974 u.d.T.: Paul J. Werbos: Beyond regression*. A Wiley-Interscience publication. New York, NY and Chichester: Wiley. ISBN: 0471598976.

Wood, R., C. S. Bretherton, D. Leon, A. D. Clarke, P. Zuidema, G. Allen, and H. Coe (2011). "An aircraft case study of the spatial transition from closed to open mesoscale cellular convection over the Southeast Pacific". In: *Atmospheric Chemistry and Physics* 11.5, pp. 2341–2370. DOI: 10.5194/acp-11-2341-2011.

Wood, R. (2012). "Stratocumulus Clouds". In: *Monthly Weather Review* 140.8, pp. 2373–2423. ISSN: 0027-0644. DOI: 10.1175/MWR-D-11-00121.1.

Wood, R. and D. L. Hartmann (2006). "Spatial Variability of Liquid Water Path in Marine Low Cloud: The Importance of Mesoscale Cellular Convection". In: *Journal of Climate* 19.9, pp. 1748–1764. ISSN: 0894-8755. DOI: 10.1175/JCLI3702.1.

Wood, T. (2019). In: *DeepAI*. URL: https://deepai.org/machine-learning-glossary-and-terms/softmax-layer (Accessed on: 01 Mar. 2022).

Wyant, M. C., C. S. Bretherton, H. A. Rand, and D. E. Stevens (1997). "Numerical Simulations and a Conceptual Model of the Stratocumulus to Trade Cumulus Transition". In: *Journal of the Atmospheric Sciences* 54.1, pp. 168–192. ISSN: 0022-4928. DOI: 10.1175/1520-0469.

Xue, H., G. Feingold, and B. Stevens (2008). "Aerosol Effects on Clouds, Precipitation, and the Organization of Shallow Cumulus Convection". In: *Journal of the Atmospheric Sciences* 65.2, pp. 392–406. ISSN: 0022-4928. DOI: 10.1175/2007JAS2428.1.

Yamaguchi, T. and G. Feingold (2015). "On the relationship between open cellular convective cloud patterns and the spatial distribution of precipitation". In: *Atmospheric Chemistry and Physics* 15.3, pp. 1237–1251. DOI: 10.5194/acp-15-1237-2015.

Yuan, T., H. Song, R. Wood, J. Mohrmann, K. Meyer, L. Oreopoulos, and S. Platnick (2020). "Applying deep learning to NASA MODIS data to create a community record of marine low-cloud mesoscale morphology". In: *Atmospheric Measurement Techniques* 13.12, pp. 6989–6997. DOI: 10.5194/amt-13-6989-2020.

Zhang, J., P. Liu, F. Zhang, and Q. Song (2018). "CloudNet: Ground–Based Cloud Classification With Deep Convolutional Neural Network". In: *Geophysical Research Letters* 45.16, pp. 8665–8672. ISSN: 0094-8276. DOI: 10.1029/2018GL077787.