CONVOLUTIONAL NEURAL NETWORKS FOR WIND TURBINE CLUTTER DETECTION

Lappeenranta-Lahti University of Technology LUT

Master's Program in Computational Engineering, Master's Thesis

2025

Pauli Anttonen

Examiners: Professor Lassi Roininen
D.Sc. (Tech.) Annakaisa von Lerber

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Sciences
Computational Engineering

Pauli Anttonen

Convolutional Neural Networks for Wind Turbine Clutter Detection

Master's thesis
2025
50 pages, 22 figures, 5 tables
Examiners: Professor Lassi Roininen and D.Sc. (Tech.) Annakaisa von Lerber

Keywords: machine learning, pattern recognition, signal processing, weather radar

Wind turbine clutter (WTC) presents a significant challenge in weather radar measurements, as the large rotating structures generate strong backscattering that compromises measurement reliability. This study explores the use of convolutional neural networks (CNNs) for the detection of WTC in weather radar in-phase (I) and quadrature (Q) signal data. The research aims to develop a computationally lightweight, efficient model suitable for real-time radar applications. A diverse dataset was created using manually labelled WTC samples and automatically extracted weather and ground clutter samples. A CNN-based classification model was designed and optimized, incorporating modern techniques such as global average pooling and dropout regularization to enhance generalization and robustness. The model achieved an accuracy of 98.24% on test data and demonstrated strong performance in detecting WTC across various radar sites, including one location not included in the training set. Post-processing techniques were implemented to refine detection results and minimize false positives. The findings indicate that CNN-based methods can effectively identify WTC and provide a foundation for further advancements in real-time radar data processing.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Sciences
Laskennallinen tekniikka

Pauli Anttonen

Konvoluutioneuroverkot tuulivoimalahäiriöiden tunnistamisessa

Diplomityö
2025
50 sivua, 22 kuvaa, 5 taulukkoa,
Tarkastajat: Professor Lassi Roininen ja D.Sc. (Tech.) Annakaisa von Lerber

Hakusanat: koneoppiminen, hahmontunnistus, signaalinkäsittely, säätutka

Tuulivoimaloiden aiheuttamat häiriöt ovat merkittävä haaste säätutkamittauksille, sillä suuret pyörivät rakenteet tuottavat voimakkaita häiriökaikuja, jotka heikentävät tutkahavaintojen luottettavuuta. Tässä työssä tutkitaan konvoluutioneuroverkkojen soveltuvuutta häiriöalueiden tunnistamiseen säätutkan signaalidatan vaihe (I) ja kvadratuuri (Q) komponenteista. Tavoitteena on kehittää pohja laskennallisesti kevyelle ja tehokkaalle mallille, joka soveltuisi reaaliaikaiseen käyttöön tutkaprosessoinnissa. Mallin kehitystä varten koostettiin laaja ja monipuolinen aineisto yhdistämällä manuaalisesti valikoituja häiriönäytteitä sekä automaattisesti kerättyjä sade- ja maakaikunäytteitä. Konvoluutioneuroverkkopohjainen luokittelumalli kehitettiin ja optimoitiin hyödyntäen moderneja koneoppimismenetelmiä kuten "global average pooling"-operaattoria ja katoregulaatiota, jotka paransivat mallin yleistettävyyttä ja vakautta. Mallilla saavutettiin 98.24%:n tarkkuus testiaineistolle ja se kykeni tunnistamaan tuulivoimahäiriöitä myös tutkalta, jonka dataa ei oltu käytetty koulutusaineistolle. Jälkikäsittelymenetelmillä hienosäädettiin mallia ja vähennettiin virheellisiä tunnistuksia. Tulokset osoittavat, että konvoluutioneuroverkot tarjoavat tehokkaan ratkaisun häiriöalueiden tunnistamiseen ja muodostavat vankan pohjan jatkokehitykselle reaaliaikaisessa tutkadatan käsittelyssä.

ACKNOWLEDGEMENTS

First, I would like to thank my examiners, Professor Lassi Roininen from Lappeenranta-Lahti University of Technology and Annakaisa von Lerber from the Finnish Meteorological Institute, for sharing their expertise and guiding me through the process of creating this master's thesis as my examiners. I would also like to thank Jenna Ritvanen from the Finnish Meteorological Institute for continued feedback and support as my supervisor during this process.

I would like to thank my family for supporting me on my journey. I want to thank my friends and the computational engineering guild, Lateksii ry, for making my active life during my studies more enjoyable. Lastly, I would like to thank my partner, Matilda, for being supportive and understanding throughout my studies and the thesis process.

This thesis has been spell-checked using Grammarly software to guarantee linguistic quality. Writing tips were utilised from generative AI to improve text flow. The author verified all the suggested corrections.

Lappeenranta, April 2, 2025

Pauli Anttonen

LIST OF ABBREVIATIONS

ADAM	adaptive moment estimation
ASHA	asynchronous successive halving algorithm
CCOR	clutter correction
CNN	convolutional neural network
DFT	discrete Fourier transform
FFT	fast Fourier transform
FIR	finite impulse response
FMI	Finnish Meteorological Institute
GMAP	Gaussian model adaptive processing
HTI	height-time indicator
IDFT	inverse discrete Fourier transform
IF	intermediate frequency
KAN	Kankaanpää Ylisenharju
LUO	Sodankylä Luosto
ML	machine learning
MLP	multilayer perceptron
PPI	plan position indicator
PRF	pulse repetition frequency
PRT	pulse repetition time
ReLU	rectified linear unit
RHI	range-height indicator
STFT	short time Fourier transform
UTA	Utajärvi Korkiakangas
WTC	wind turbine clutter

CONTENTS

1 INTRODUCTION **7**
1.1 Background 7
1.2 Objectives and delimitations 8
1.3 Structure of the thesis 9

2 WIND TURBINE CLUTTER IN WEATHER RADAR MEASUREMENTS **10**
2.1 Weather radar systems 10
2.1.1 Background 10
2.1.2 Pulse radar 11
2.1.3 Doppler radar 12
2.1.4 Dual-polarisation radar 13
2.1.5 FMI radar network 14
2.2 Signal processing 16
2.2.1 In-phase quadrature components 16
2.2.2 Ground clutter filtering 18
2.2.3 Radar reflectivity factor 19
2.2.4 Wind turbine clutter mitigation 21

3 WIND TURBINE CLUTTER DETECTION USING CNN MODEL **23**
3.1 Dataset creation 23
3.2 Model definition 25
3.3 Model optimisation 27
3.4 Post-processing 29

4 MODEL EVALUATION **30**
4.1 Data 30
4.2 Evaluation criteria 32
4.3 Results 33

5 DISCUSSION **43**

6 CONCLUSIONS **46**

REFERENCES **47**

1 INTRODUCTION

1.1 Background

The global shift to renewable energy has introduced various challenges, such as those emerging in electromagnetic systems. One such challenge is the impact of wind turbines on weather radar systems. Research conducted in Sweden demonstrates a significant increase in average weather radar reflectivity in areas surrounding wind turbine construction sites following their installation (Norin, 2015). A similar trend has been observed in the Finnish radar network, where backscattered reflectivity values increased around nearly every wind turbine located within 100 km of a radar installation over a two-year period (Ritvanen et al., 2024). This phenomenon, known as wind turbine clutter (WTC), arises from radar signal interacting with the tower and the rotating blades of large wind turbines. As the adoption of wind energy continues to expand globally, WTC is becoming an increasingly critical issue for radar-based systems (Isom et al., 2009).

WTC presents significant challenges across various electromagnetic communication and detection systems. It affects radio links (Randhawa and Rudd, 2009), broadcast communication (Sengupta and Senior, 1979), air surveillance radars (Lemmon et al., 2008), maritime radars (Grande et al., 2014), and, most notably, weather radars (Gallardo-Hernando et al., 2010; Isom et al., 2009; Norin, 2015; Ritvanen et al., 2024). The interference caused by WTC can compromise the reliability of these systems, leading to potential risks in weather forecasting, aviation safety, and maritime operations.

Doppler weather radars measure the amplitude and phase of the pulses (Rauber and Nesbitt, 2018). To correctly display this information, the signal processor processes the raw received signal into two components: in-phase (I) and quadrature (Q). Analysing these values offers a deep level of access to the radar's signal processing, which can help identify contamination such as WTC (Norin, 2017).

Many methods have been explored to characterise and mitigate WTC in radar data. Techniques such as fuzzy logic schemes (Hood et al., 2010), spectral-spatial interpolation for Doppler moments (Isom et al., 2009), zero-Doppler echo power analysis (Gallardo-Hernando et al., 2010), and range-Doppler domain processing (Nai et al., 2013) have shown varying degrees of success. Despite these advancements, accurately identifying and mitigating WTC, particularly in precipitation spectra, remains a considerable challenge.

As an optimisation-based method, machine learning (ML) methods offer scalable and generalised classification methods for problems like WTC detection (Botsch, 2023). ML offers a wide range of complex methods for feature extraction with minimal prior knowledge of the data structure. In classification tasks like WTC detection, the ML model assigns samples to predefined classes based on their features. One example of an ML technique applied to weather radar data is presented by Husnoo et al. (2021), who develop a ML-based quality-control algorithm for real-time radar data processing.

One widely recognised kernel-based optimisation method for time series classification is a convolutional neural network (CNN) (Goodfellow et al., 2016). A CNN is a modified version of a basic feed-forward neural network where at least one of the layers is replaced with a convolutional layer. The convolutional layer works by employing multiple convolutional kernels. These kernel weights are optimised to extract spatial features that best separate the classes.

CNNs are widely recognised for signal classification in the literature. For example, they are used in detecting cardiovascular diseases in electrocardiography signals (Wu et al., 2021), determining user intentions from brain electricity activity (Bang et al., 2022) and catching bearing faults from vibration signals (Wang et al., 2021). In I/Q processing, Kunze and Saha (2022) successfully performed a drone classification using a CNN-based model on radio signals.

1.2 Objectives and delimitations

The primary objective of this research is to develop an effective and lightweight CNN-based approach for detecting WTC in weather radar I/Q data. To achieve this goal, the following research questions are formulated:

1. How accurately can a CNN-based model differentiate WTC from weather and ground clutter in radar I/Q data?
2. Can a CNN-based model be designed to be computationally lightweight, making it a viable candidate for future integration into operational radar software?
3. Are short time intervals of I/Q data sufficient for creating a diverse dataset that enables the model to generalize effectively across different radar measurements?

4. Is post-processing required for the model's output, and what is the most effective approach to optimize its integration into radar signal processing?

The high volume of I/Q data presents significant challenges in both radar signal processing and data transmission between the radar site and central processing at the institute. The radar signal processor converts the I/Q data into radar variables, and, by default, the I/Q signal is not stored. Consequently, the model must be computationally efficient to operate concurrently in the radar software and capable of processing the I/Q data on-site, as its exceptionally large size prevents external processing.

1.3 Structure of the thesis

The rest of this thesis is organised as follows: Chapter 2 presents the fundamental principles of radar meteorology and describes the signal processing procedures, from I/Q data to the generation of radar products. Chapter 3 details the development of the CNN classification model and the methods used for dataset construction. Chapter 4 explains the dataset preparation process and the model training procedure, followed by an evaluation of the classification results. Finally, Chapter 5 discusses the findings in greater depth and summarises the key conclusions of the study.

2 WIND TURBINE CLUTTER IN WEATHER RADAR MEASUREMENTS

2.1 Weather radar systems

A radar is a device designed for remote sensing and operates as an active sensor by both generating and collecting electromagnetic signals (Rauber and Nesbitt, 2018). Unlike passive sensors, such as the human eye, radar does not rely on solar energy reflection or thermal emissions. Instead, it works by emitting electromagnetic pulses towards a target of interest. After a short delay, the radar detects the backscattered energy, enabling the measurement of the target's distance and relative size (Rauber and Nesbitt, 2018).

2.1.1 Background

Originally developed as a military tool, radar technology was adapted for meteorological sensing after World War II (Fabry, 2015). Initially designed to detect distant military targets, advancements in magnetron research enabled radar systems to operate at much lower wavelengths. These lower wavelengths allowed for the detection of large patches of echoes whose origins were initially unknown. It was soon realized that these echoes were caused by precipitation. Once wartime secrecy was lifted, meteorology personnel were introduced to these findings, unlocking the potential of radar technology for weather observation and forecasting (Fabry, 2015).

Weather radars operate by emitting a microwave pulse towards the clouds to gather information about hydrometeors: water or ice particles suspended in the atmosphere, such as raindrops, snowflakes, hail, or sleet (Rauber and Nesbitt, 2018). Since hydrometeors are spaced unevenly, the wave does not reflect off the cloud like it would from a solid surface. Instead, a portion of the energy is backscattered by the target hydrometeors, while the rest continues to propagate further.

Radars are widely used for weather monitoring as they provide accurate and real-time information over large areas. They can detect precipitation patterns in three dimensions over time, making them unparalleled in tracking the movement and development of weather conditions (Rauber and Nesbitt, 2018).

Weather radar systems operate along two principal directions: the azimuth and elevation angles. The azimuth angle is defined as the horizontal angle between the radar beam and true north, while the elevation angle is the angle between the beam direction and the horizontal plane (Fabry, 2015). Figure 1 illustrates these angles.

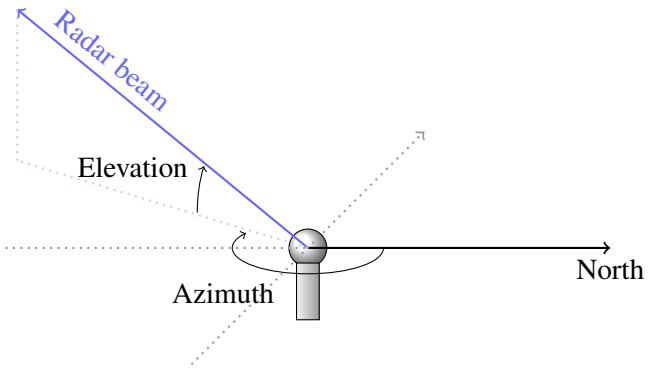

Figure 1. The radar antenna direction is given with two angles shown in the illustration.

Meteorological scanning strategies for weather radar are typically classified into three types (Fabry, 2015). The first type involves the radar rotating around the azimuth axis while maintaining a fixed elevation angle, producing a raw image known as a plan position indicator (PPI). The second type of image, called a range-height indicator (RHI), is generated by keeping the azimuth angle constant while the antenna scans along the elevation axis. The third type of scan is conducted by keeping the azimuth and elevation angles fixed and measuring reflectivity over time. When the radar is oriented straight upwards during this process, the resulting image is referred to as a height-time indicator (HTI).

2.1.2 Pulse radar

The most fundamental radar measurement, slant range, is a distance measured along the radar ray (Rauber and Nesbitt, 2018). Weather radars measure this range by transmitting short pulses, each followed by a much longer listening period. The range r is then calculated by

$$r = \frac{c\Delta t}{2}, \tag{1}$$

where c is the speed of light and Δt is the time between pulse transmission and retrieval.

Selecting the pulse length in pulse radar systems involves a trade-off between resolution and detection capability (Rauber and Nesbitt, 2018). A radar pulse is a finite-length wave packet oscillating at the system's operational frequency. The ability to distinguish weak targets from noise is directly proportional to the amount of returned radiation and, consequently, to the pulse length. However, an extended pulse length adversely affects the radar's capacity to resolve closely spaced targets. Typical pulse lengths for weather radars correspond to a travel distance of a few hundred meters (Rauber and Nesbitt, 2018).

The time between two consecutive pulses defines the maximum range the radar can unambiguously measure (Fabry, 2015). Ambiguous measurements arise from targets located beyond this maximum range. When the radar receives backscatter from these extended targets, it becomes impossible to distinguish whether the signal originates from closer targets illuminated by the current pulse or from farther targets during a previous pulse. The delay between two pulses is known as the pulse repetition time (PRT) and the inverse of this as the pulse repetition frequency (PRF). The maximum measurable range, denoted as r_{max}, can be calculated using the formula

$$r_{\text{max}} = \frac{c}{2(\text{PRF})}. \tag{2}$$

Weather radar antenna patterns comprise two primary components: the main lobe and sidelobes (Rauber and Nesbitt, 2018). The main lobe serves as the principal source of backscattered reflectivity. The pulse width refers to the width of the main lobe, defined by the points at which the power decreases to half of its maximum value. Due to diffraction, some of the transmitted energy spreads from the main beam in all directions around the radar, giving rise to sidelobes.

2.1.3 Doppler radar

Weather radars not only detect precipitation but also measure the motion of atmospheric particles using the Doppler effect. Similar to the frequency shifts in sound waves caused by a fast-moving ambulance siren, slow-moving rain clouds induce frequency shifts in radar microwaves. These shifts, known as Doppler frequency shifts, are a fundamental component of modern weather radar measurements (Rauber and Nesbitt, 2018). By analysing the Doppler shift in the backscattered signal, it is possible to estimate the relative velocity of atmospheric targets with respect to the radar.

Radars capable of measuring Doppler shifts are referred to as Doppler radars. These systems generate Doppler spectra using the Fourier transform, representing the power distribution as a function of Doppler velocity (Fabry, 2015). Given the phase shift $\Delta\phi$ radial velocity relative to radar beam direction v_r can be expressed by

$$v_r = \frac{\lambda(\mathrm{PRF})}{2}\left(\frac{\Delta\phi}{2\pi}\right), \tag{3}$$

where λ is the radar wavelength.

The maximum velocity that can be measured unambiguously by Doppler radar is called Nyquist velocity (Rauber and Nesbitt, 2018). An ambiguity in radial velocity measurements comes from phase shifts with absolute value over π. These measures are folded into the range $[-\pi, \pi]$, which leads to ambiguous velocity measurements. Based on the Equation (3), the Nyquist velocity is defined as

$$V_u = \frac{\lambda(\mathrm{PRF})}{4}. \tag{4}$$

The selection of the PRF involves a trade-off between the Nyquist velocity, as defined in Equation (4), and the maximum measurable range, as given in Equation (2). Determining an appropriate PRF requires balancing these two factors to optimise both the maximum measurable range and velocity (Rauber and Nesbitt, 2018).

2.1.4 Dual-polarisation radar

In addition to velocity estimation, modern weather radars increasingly incorporate dual-polarisation technology, enhancing their ability to characterise precipitation and other atmospheric targets (Rauber and Nesbitt, 2018). Unlike traditional single-polarisation radars, which transmit and receive microwaves in only one polarisation (typically horizontal), dual-polarisation radars transmit and receive signals in both horizontal and vertical polarisations (Fabry, 2015). This capability provides additional information about atmospheric particles' shape, orientation, and phase, thereby improving the classification and differentiation of meteorological and non-meteorological echoes (Rauber and Nesbitt, 2018).

The differing responses of horizontal and vertical polarisations provide valuable insights into target characteristics. The horizontal polarisation is primarily influenced by a target's horizontal dimensions, while the vertical polarisation responds to its vertical extent

(Fabry, 2015). This distinction is particularly useful for differentiating meteorological targets, such as raindrops, which tend to be more spherical, from non-meteorological targets like birds and insects (Rauber and Nesbitt, 2018). These advantages make dual-polarisation radars the current standard in meteorological applications, significantly enhancing detection capabilities over single-polarisation systems. The dual-polarisation principle is illustrated in Figure 2.

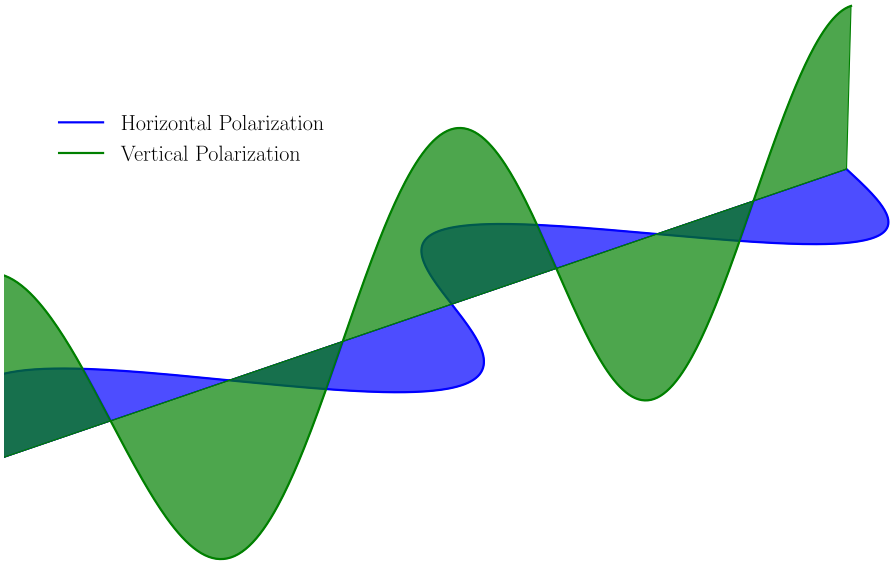

Figure 2. The illustartion shows two microwaves oscillating in two orthogonal directions.

2.1.5 FMI radar network

With an evenly distributed network of weather radars, radar products can be generated by integrating data from multiple radars to achieve comprehensive spatial coverage (Rauber and Nesbitt, 2018). The radar network of the Finnish Meteorological Institute (FMI) includes 11 dual-polarisation Vaisala WRM200 Doppler radars (FMI, 2025). This thesis primarily examines data from three radar sites within the network: Kankaanpää Ylisenharju (KAN), Sodankylä Luosto (LUO), and Utajärvi Korkiakangas (UTA), as depicted in Figure 3. The thesis only focuses on the scans of the lowest elevation angle since the ground objects, such as wind turbines, have the most significant influence at this level. The radar locations and specifications for the radar task are shown in Table 1.

Figure 3. Map shows the three relevant radar sites highlighted with their maximum ranges in the FMI radar network.

Table 1. Technical specifications displayed for each radar site, including the parameters for PPI Task A.

Radar site	**KAN**	**LUO**	**UTA**
Latitude	61.80	67.14	64.77
Longitude	22.50	26.90	26.32
Height (AMSL)	174 m	532 m	118 m
PPI Task A			
Elevation angle	0.3°	0.1°	0.3°
Pulse width	1.0°	1.0°	1.0°
Pulse length	2 μs	2 μs	2 μs
PRF	570	570	570
Sample size	32	32	32
Range resolution	125 m	125 m	125 m
Max range	250 km	250 km	250 km

2.2 Signal processing

The Vaisala WRM200 uses the RVP900 Digital Receiver and Signal processor to interpret the received signal. RVP900 Digital Receiver and Signal Processor consists of two parts: RVP901 IF Digital Receiver and RVP902 Signal Processing Computer (Vaisala, 2021). RVP901 processes the input intermediate frequency (IF) signal into in-phase (I) and quadrature (Q) data pairs. RVP902 then calculates radar variables from the I/Q data. This process is shown in Figure 4.

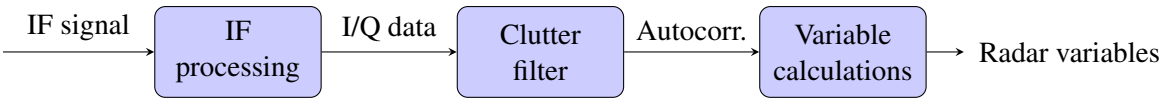

Figure 4. Process diagram shows the RVP900 signal processing structure.

2.2.1 In-phase quadrature components

In RVP900 the raw signal collected by the weather radar is downsampled to 60 MHz to facilitate processing (Vaisala, 2021). Since we can estimate the range from the time it takes for the backscatter to return to the radar, we can determine the distance from which each backscatter originates. Given that the pulse length of 2 microseconds corresponds to 150 metres in travel distance, making the estimate for range uncertain, a generalised formula is required to estimate the backscatter for each range interval.

From the 60 MHz intermediate frequency (IF) signal, I/Q data pairs can be calculated using a low-pass finite impulse response (FIR) filter with N points equally distributed around the reference point corresponding to the desired range. The formulas presented by Vaisala (2021) for calculating the I and Q components are as follows:

$$\begin{aligned} I &= \sum_{n=0}^{N-1} f_n^i \times p_n \\ Q &= \sum_{n=0}^{N-1} f_n^q \times p_n, \end{aligned} \tag{5}$$

where N is the number of points and

$$f_n^i = l_n \times \sin\left[\frac{\pi}{4} + 2\pi \frac{f_{\text{IF}}}{f_{\text{SAMP}}}\left(((N-1)-n) - \frac{N-1}{2}\right)\right], \tag{6}$$

where l_n are the coefficients of the symmetric low-pass FIR filter, f_{IF} is the IF frequency and f_{SAMP} is the sampling frequency of the radar. These coefficients are chosen to reduce noise and average the power in N points into a single I/Q pair (Vaisala, 2021). The complex representation of the I/Q data pair also holds the signal's phase information, which is why f_n^i is represented with a periodic function. The formula for coefficients f_n^q is identical to Equation (6) with the sine function replaced with the cosine function.

In each range interval, the I and Q components are combined to form a complex representation of the backscattered pulse, representing both its amplitude and phase. In PPI scans, the radar's rotation speed along the azimuthal axis is configured such that a one-degree rotation yields a signal comprising 32 consecutive pulses. Our primary interest lies in analysing successive pulses originating from the same geographical region, later referred to as a radar bin. The time series of consequent pulses s_m in a radar bin can be denoted as

$$s_m = I_m + jQ_m, \tag{7}$$

where m denotes the pulse index in the time series. An example of an I/Q signal is shown in Figure 5.

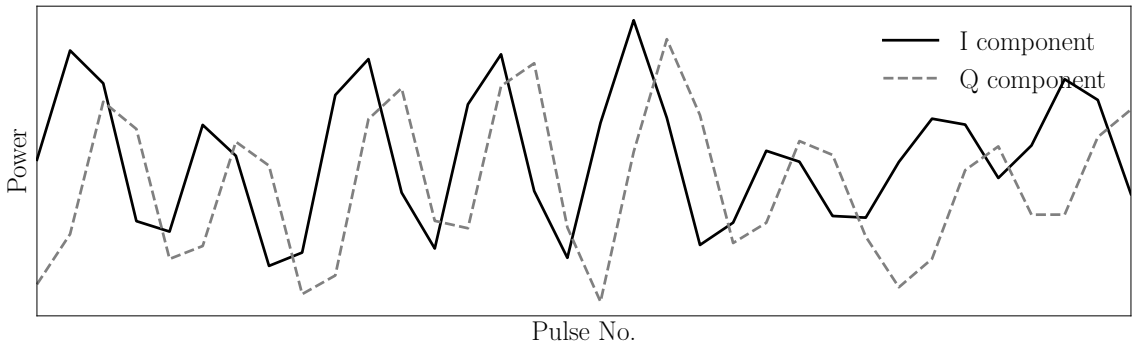

Figure 5. Example I/Q signal for a single radar bin with weather measurement. The signal is constructed from consequent pulses in a single bin.

From the complex I/Q time series, Doppler spectra X_k can be obtained using discrete Fourier transform (DFT):

$$X_k = \sum_{m=0}^{M-1} s_m \exp\left(-j2\pi \frac{k}{N} m\right), \tag{8}$$

where M denotes the number of I/Q pairs in each range interval. Using the inverse discrete Fourier transform (IDFT) I/Q signal can be composed back from Doppler spectra with the

following expression:

$$s_m = \frac{1}{M} \sum_{k=0}^{M-1} X_k \exp\left(j2\pi \frac{k}{M} m\right). \tag{9}$$

In practice DFT is calculated using fast Fourier transform (FFT) algorithm which reduces the time complexity from $O(m^2)$ to $O(m \log m)$ by pairing odd and even base functions during the computation (Cooley and Tukey, 1965).

2.2.2 Ground clutter filtering

Large, stationary objects on the ground, as well as the ground surface itself, can generate echoes that interfere with weather measurements (Rauber and Nesbitt, 2018). This interference, known as ground clutter, can be mitigated by utilising the Doppler radar's ability to form a Doppler spectrum from the observed data. Basic signal processing techniques can then be applied to reduce the impact of these effects.

The RVP900 provides various signal processing options for ground clutter filtering (Vaisala, 2021). The FMI weather radars use Gaussian model adaptive processing (GMAP) introduced by Siggia and Passarelli (2004). Algorithm 1 shows a simplified version of the method.

Algorithm 1 GMAP Algorithm Steps

1. Apply windowed DFT to time series s_n to obtain Doppler power spectra.
2. Fit a Gaussian model representing the ground clutter with a predetermined width to three central points. These points are highlighted in Figure 6.
3. Remove points within the intersection of the Gaussian model and noise level.
4. Fit the Gaussian model representing the weather to the remaining points and fill in removed points from this Gaussian model.
5. Analyse the result and return to Step 1 using a more appropriate window if needed.

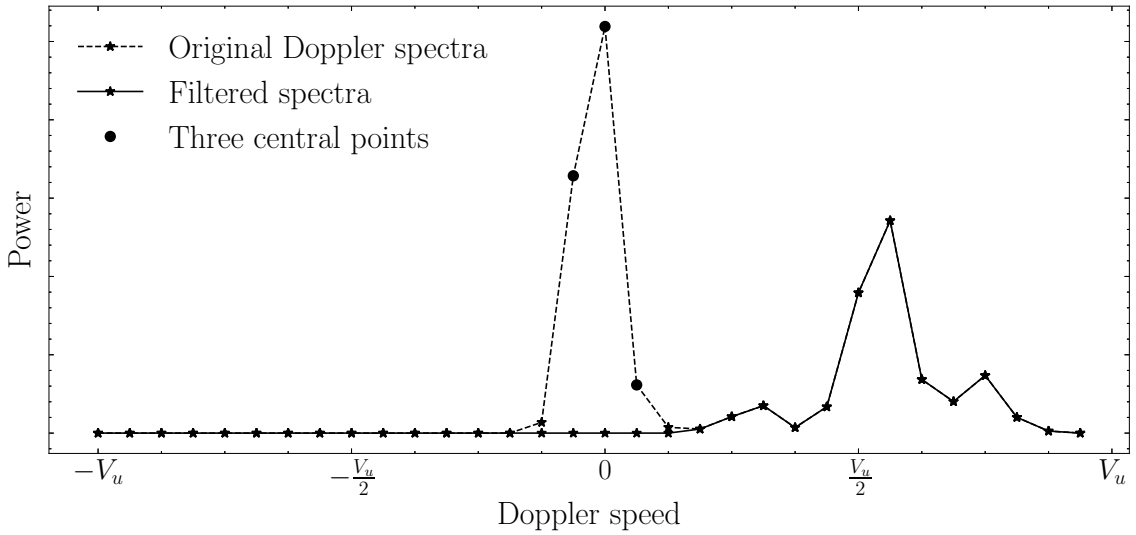

Figure 6. Example Doppler spectrum of a signal filtered using GMAP. The Doppler spectrum is computed from the I/Q signal using the DFT.

The GMAP is employed due to its robustness and accuracy that offers advantages over many previously used techniques (Siggia and Passarelli, 2004). A key feature of this method is its ability to apply minimal filtering in the absence of ground clutter, ensuring that all radar bins are processed without unnecessary modification.

The IDFT is applied to the filtered Doppler spectra, which results in the filtered time series s'_n. Mean power for original and filtered time series are then calculated with the following expressions

$$\begin{aligned} T_0 &= \frac{1}{M}\sum_{n=1}^{m} s_n \times s_n \\ R_0 &= \frac{1}{M}\sum_{n=1}^{m} s'_n \times s'_n, \end{aligned} \tag{10}$$

where M is the number of pulses per range interval.

2.2.3 Radar reflectivity factor

The reflectivity factor is one of the most important quantities in radar meteorology (Rauber and Nesbitt, 2018). It corresponds to the sum of the sixth power of the diameter of all raindrops within a given volume

$$Z = \frac{\sum_i D_i^6}{V}, \tag{11}$$

where D_j is the diameter of the hydrometeor, and V is the contributing volume. In addition, the reflectivity factor Z provides an estimate of the rainfall rate R. Numerous Z-R relationships have been published, often featuring different exponential relationships between the two variables (Fabry, 2015). For example, Marshall and Palmer (1948) introduced the widely-used relationship $Z = 200R^{1.6}$ while Leinonen et al. (2012) proposed $Z = 223R^{1.53}$ specifically for the Finnish climatology.

The RVP900 uses the signal power at the receiver P_r to calculate the reflectivity factor

$$Z = Z_0 \left(\frac{P_r - N}{N} \right) r^2, \tag{12}$$

where N is the calibrated noise level, and r is the distance from the radar. The constant Z_0 is called the calibration reflectivity factor, and it equals the minimum detectable reflectivity at the reference range (Vaisala, 2021). The calibration reflectivity factor is a product of many factors dependent on the radar equipment and radar pulse settings.

The reflectivity factor Z is usually expressed in logarithmic units $\mathrm{dBZ} = 10 \log_{10} Z$ since it can vary by many orders of magnitude (Rauber and Nesbitt, 2018). Vaisala (2021) introduces the reflectivity factor dBZ as

$$\mathrm{dBZ} = 10 \log \left(\frac{T_0 - N}{N} \right) + \mathrm{dBZ}_0 + 20 \log r + ar + \mathrm{CCOR}, \tag{13}$$

where P_r is replaced with the calculated power T_0 from Equation (10). The negative term ar accounts for the attenuation caused by atmospheric gases (Rauber and Nesbitt, 2018). The ground clutter mitigation is taken into account by adding a negative clutter correction (CCOR) term. This is estimated by

$$\mathrm{CCOR}_{\mathrm{est}} = 10 \log_{10} \left(\frac{R_0}{T_0} \right). \tag{14}$$

This simply reduces the reflectivity factor by the amount of power reduced in the GMAP. The total reflectivity calculated without the CCOR term is often referenced as dBT.

2.2.4 Wind turbine clutter mitigation

The wind turbine clutter (WTC) can be thought to consist of two sources: the tower and the blades (Isom et al., 2009). The clutter signal produced by stationary towers can be filtered with ground clutter filtering techniques such as GMAP, whereas the fast-moving blades produce a much more complicated signal.

Unlike typical ground clutter, wind turbine blades generate complex signals that cause them to be detected as non-stationary targets (Isom et al., 2009). GMAP is designed for stationary objects and is therefore ineffective in mitigating this type of clutter. In I/Q data, WTC introduces unpredictable amplitude variations between consecutive pulses (Norin, 2017). These abrupt fluctuations create intricate patterns in the Doppler spectra, making them difficult to identify. Examples of weather and WTC signals are presented in Figure 7.

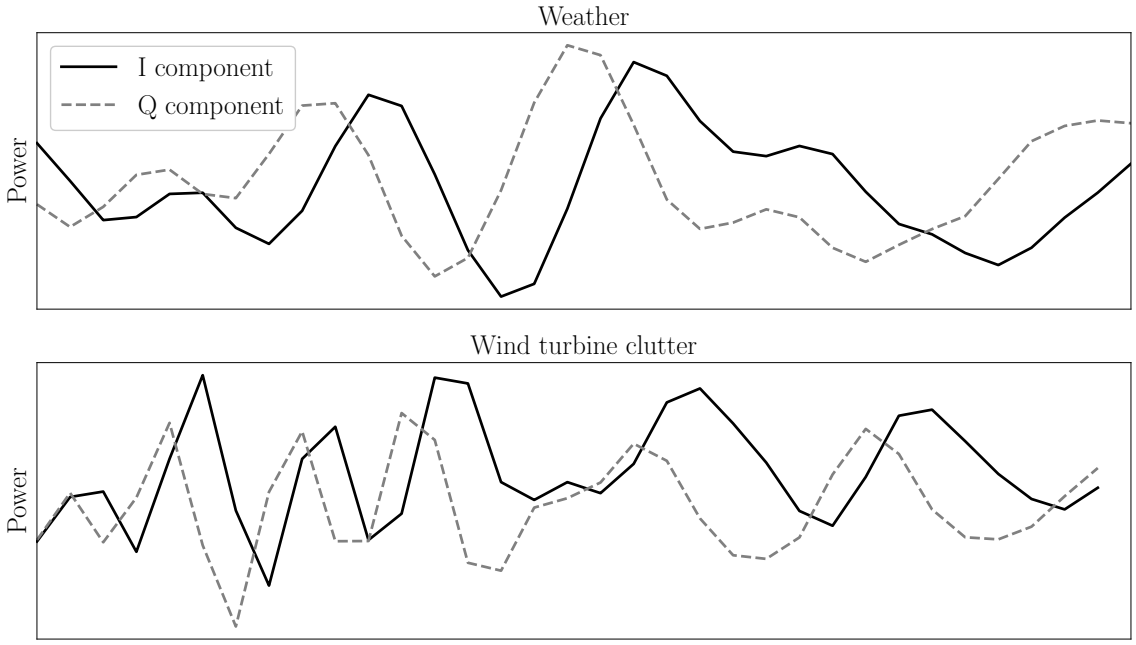

Figure 7. Two I/Q signals plotted from different measurements. Each figure shows the I and Q signals in a single bin constructed from consecutive pulses.

Various methods have demonstrated differing levels of success in detecting and mitigating WTC. Isom et al. (2009) proposes an effective multiquadric interpolation method for masking the impact of WTC. Interpolation can be beneficial when the contaminated bins are identified properly. One of the most promising approaches for detecting WTC involves applying the short time Fourier transform (STFT) with a sliding window (Nai and Torres, 2024). This method leverages the observation that wind turbines induce greater

contamination in radar pulses that are more closely aligned with their location, whereas steady backscatter caused by weather remains relatively uniform.

With the restrictions imposed by the RVP900, I/Q processing methods must be applied after IF processing but before the GMAP filter. In the intended design, WTC should be identified and corrected at the I/Q level before variable calculations. The final implementation is expected to detect bins contaminated by WTC and apply appropriate substitution or filtering techniques to mitigate the impact. The integration of WTC detection into the RVP900 framework is illustrated in Figure 8.

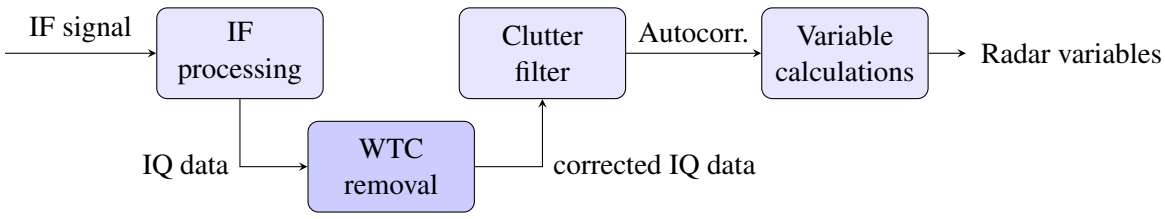

Figure 8. Ideal signal processing structure shown with implemented WTC removal.

3 WIND TURBINE CLUTTER DETECTION USING CNN MODEL

A convolutional neural network (CNN) is a type of neural network specifically designed for processing data with a grid-like structure (Goodfellow et al., 2016). In a manner similar to how the STFT extracts spatial features from signals through spectrograms (Kumawat and Raman, 2020), CNNs capture spatial dependencies via convolutional layers. This capability makes them particularly effective in pattern recognition tasks. In radar signal processing, the I and Q signals of a radar bin can be represented as one-dimensional data grids to be processed by these convolutional operations.

CNN models have already been successfully applied in other domains of electromagnetic research, including signal classification tasks. In Kunze and Saha (2022), a CNN-based approach was shown to achieve over 95% classification accuracy in identifying drones using I/Q data derived from radio frequency signals. These findings suggest that CNNs could be a promising tool for the detection of WTC in weather radar I/Q data, as they can effectively learn and differentiate relevant signal features.

3.1 Dataset creation

The main goal of dataset creation is to create a diverse dataset that contains a variety of different measurements from rain, ground clutter, and WTC to be used for training and testing the developed model. Large datasets expose the model to a variety of different scenarios and allow it to learn complex patterns (Emam et al., 2021). A diverse dataset with a broad spectrum of variation also makes the model more robust, improving its capability to deal with rare edge cases (Gong et al., 2019).

Due to the limitations of automated WTC detection in I/Q data, manual identification of contaminated samples is required to build a high-quality training set for machine learning. This labeling process utilises wind turbine locations from aviation obstacle data (Fintraffic, 2025) and radar reflectivity values derived from I/Q data. The labeling was conducted by inspecting the reflectivity values near known wind turbine locations. A radar bin was classified as contaminated if its reflectivity value exceeded the surrounding reflectivity levels, which were attributed to weather phenomena or pure receiver noise. WTC regions typically exhibited sharper boundaries compared to weather-related reflectivity, which displayed more gradual variations. The labelling process was deliberately focused on

marking only clearly contaminated cells, as ensuring high-quality data was deemed more critical than maximising the quantity of labelled samples.

Rain measurements were sampled from all the radar bins with reflectivity values over 8 dBZ (light precipitation) and distance to the nearest wind turbine over 5 km. To ensure evenly distributed rain measurement across different precipitation classes, valid radar bins were sampled with three different intensities detailed in Table 2.

Table 2. Three precipitation classes shown with their reflectivity ranges.

Light precipitation	8 - 15 dBZ
Medium precipitation	15 - 33 dBZ
High precipitation	Over 33 dBZ

Due to the architecture of the RVP900, the classification module must be located before the ground clutter filtering. Therefore, samples with ground clutter interference must be distinguished from the WTC samples. Bins with ground clutter are sampled using a technique similar to that of weather samples. The bins where $\mathrm{CCOR} > 3$ dB, as expressed in Equation (14), are considered as ground clutter. Since ground clutter can appear in bins with weather, these bins are sampled with and without weather. To achieve this, half of the ground clutter samples are equally split between the precipitation categories in Table 2, and half are sampled from the bins where reflectivity after clutter filtering dBZ < -6 (applied detection threshold for light snowfall).

Since bins containing weather or ground clutter can be automatically sampled, the primary limiting factor for the dataset size is the availability of WTC-contaminated bins. The dataset size can be expanded by generating new samples through the augmentation of existing WTC samples, thereby enhancing data diversity (Kim and Chung, 2024).

Synthetic WTC data samples are created by adding white noise to the existing WTC data input samples to mimic receiver noise. This process is applied to input signal s by

$$\tilde{s}_n = s_n + e_i \quad \forall\, s_n \in s\,, \tag{15}$$

where $\tilde{s}$ denotes the augmented signal and additive noise is expressed with $e_i \sim \mathcal{N}(0, \sigma^2)$. The parameter σ corresponding to the noise level is chosen to be close to the calibration noise level of the radar. As a side effect of increasing the number of training samples this way, the resulting model is less volatile towards noisy measurements (Kim and Chung, 2024).

3.2 Model definition

The classifier model consists of two convolutional layers, a global average pooling layer, and a single fully connected layer. The architecture is built on the PyTorch framework (Paszke et al., 2019) for efficient training and evaluation. The structure of the model is shown in Figure 9.

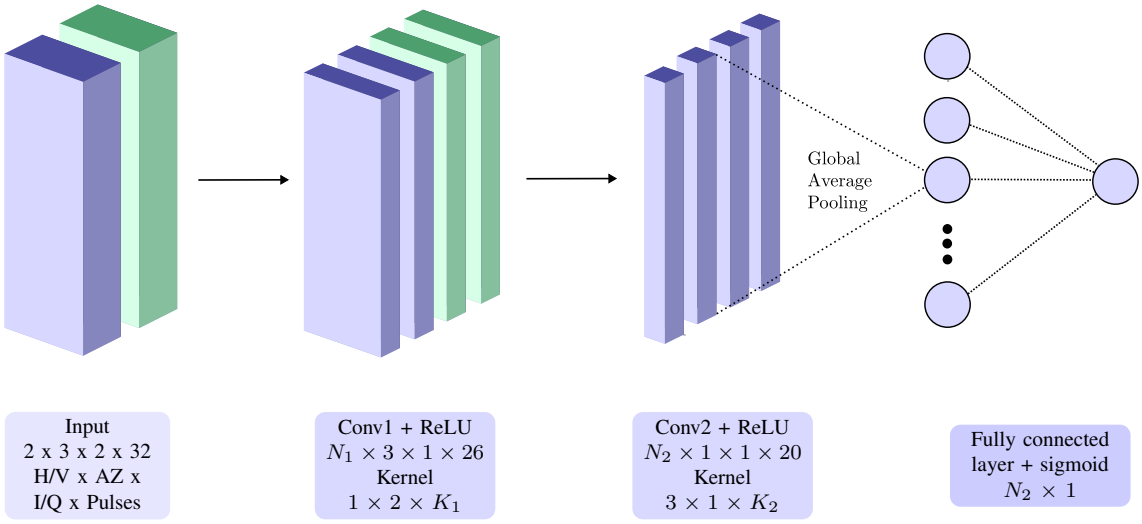

Figure 9. The schematic diagram shows the elements constructing the CNN model. The two input colors correspond to different polarisations and groups in the first convolutional layer.

The model input is formed by combining 32 pulse signals in a grid-like formation to form a single sample. It consists of one pair of I and Q signals for three adjacent bins in azimuthal direction for each polarization. During model experimentation, the number of adjacent bins was tested with five, but this led to overfitting. Thus, only three were used in the final model input. In the case of missing pulses, the I and Q signals are padded with zeros to a length of 32. The input is constructed into a three-dimensional block to capture adjacency relevance between the data points.

Backscattered reflectivity decreases with range, which makes I/Q values near the radar considerably higher (Rauber and Nesbitt, 2018). A standard dataset-wide normalisation approach, such as min-max scaling, may introduce distortions in the data. The model employs per-sample normalisation to mitigate this issue, preserving the shape and relative amplitude within each input sample. The normalisation scheme for input sample X can be given as

$$\bar{x} = \frac{x}{\max\limits_{y \in X} |y|} \quad \forall\, x \in X, \tag{16}$$

where $\bar{x}$ denotes the normalised data point and x is the unnormalised data point. This normalisation helps the model learn patterns that are not dependent on the range or the total amount of backscattered reflectivity.

The proposed model employs two convolutional layers with two-dimensional kernels to extract spatial features from the data. The first convolutional layer processes the two polarisations as separate channels by utilising a kernel that spans both I and Q values and adjacent points in the time dimension. The resultant grid of neurons is referred to as a feature map. In this layer, two distinct groups of trainable kernels are employed, with each group dedicated exclusively to one polarisation. This design facilitates the independent extraction of features specific to each polarisation.

The second convolutional layer operates on the feature maps generated by the first layer. It is designed to capture spatial relationships among geographically neighbouring radar bins, assuming that weather samples exhibit similar patterns across adjacent bins, making distinguishing them easier. The spatial information is summarised by employing a kernel that contains three adjacent bins in the azimuthal direction while also incorporating neighbouring neurons in the time dimension.

Some of the hyperparameters for the convolutional layers are intentionally left undefined to allow for subsequent optimisation. Specifically, the kernel sizes, K_1 and K_2, as well as the number of filters, N_1 and N_2, are treated as model hyperparameters to be fine-tuned in later stages of development. This approach ensures that the model can be adapted to achieve optimal performance based on empirical results. Both convolutional layers use the rectified linear unit (ReLU) function, enhancing the network's ability to learn complex patterns.

Traditional convolutional neural networks typically feed into the multilayer perceptron (MLP) structure with a single hidden layer to classify the samples. That way, all positional information from the feature maps is preserved (Goodfellow et al., 2016). The proposed model employs an averaging operator on the feature maps in this step instead. The procedure of averaging all feature map information into a single value is called global average pooling, which enforces the correspondence between feature maps and classes (Lin, 2013). This operator is implemented to replace the traditional MLP structure with a simpler averaging operator.

Global average pooling has multiple features beneficial for the WTC detection. Since each time step in the I/Q signal corresponds to a different point in the radar scan volume, WTC can appear in any part of the input signal. This procedure ensures that instead of

overfitting to positions of the wind turbines in the training set, the model learns general features for WTC detection. Since this application is not reliant on spatial information, this procedure should also reduce the complexity of the model, resulting in faster training convergence and better generalisation.

Regularisation is an essential part of many machine learning (ML) techniques, which helps models become more robust and improves their performance on unseen data (Moradi et al., 2020). It also prevents models from memorising specific inputs, which can reduce their ability to generalise. Among the most commonly used regularisation methods are L1 and L2 regularisation. However, in this study, these techniques were found to be overly restrictive, as penalising large weights led to underfitting. The dropout method, on the other hand, showed promising results during experimentation and is therefore chosen.

Dropout regularisation works by zeroing a part of the neurons to prevent memorisation and overfitting (Hinton et al., 2012). In the proposed model, dropout is applied to averaged feature maps, effectively zeroing whole maps of neurons. Using the dropout layer forces the model to learn to use multiple feature maps for each sample instead of being able to rely only on a few. This process is only applied during the training phase, and all feature maps are used during the evaluation phase.

The output is constructed from the averaged feature maps, weighting each averaged value by a trainable weight in a single fully connected layer. The output value is normalised to the range $[0, 1]$ using the sigmoid function to form a final model output. This value can be thresholded later for a binarised output. The sigmoid function is incorporated into the loss function during training.

3.3 Model optimisation

The model is trained using the PyTorch framework (Helmus and Collis, 2016), with binary cross-entropy as the loss function. Binary cross-entropy loss l_n is defined as

$$l_n = -[y_n \log(p_n) + (1 - y_n) \log(1 - p_n)], \tag{17}$$

where n is the sample index, p_n is the predicted probability, and y_n is the true label, which is either 0 or 1 in the binary case. The loss function is modified to incorporate a sigmoid activation to enhance numerical stability in gradient calculations. Given that the dataset contains more neutral samples than WTC samples, a positive class weighting was applied

to the loss function to account for this imbalance. The final loss $\tilde{l}_n$ with sigmoid function and positive class weighting is defined as

$$\tilde{l}_n = -[p_c y_n \log \sigma(x_n) + (1 - y_n) \log(1 - \sigma(x_n)], \tag{18}$$

where p_c is the positive class weight, $\sigma(x)$ denotes the sigmoid function and x_n is the unnormalised model output.

Minimizing the cross-entropy between the model's output and the target facilitates convergence towards accurate predictions. For model optimisation, adaptive moment estimation (ADAM) was employed to enhance gradient-based learning (Kingma and Ba, 2017). ADAM provides a robust and computationally efficient approach to managing gradient updates during backpropagation.

To maximise the model's accuracy, the hyperparameters are optimized using Ray Tune library (Liaw et al., 2018). Ray Tune offers a wide range of tools for parallel hyperparameter tuning, such as asynchronous successive halving algorithm (ASHA) (Li et al., 2020), which is used to dynamically allocate resources during the training process.

The ASHA scheduler works asynchronously by executing multiple models and adaptively adjusting the thresholds for each bracket (Li et al., 2020). A bracket is defined by a start epoch and an end epoch, after which half of the trial models are terminated based on their validation loss. An epoch refers to one complete pass of the training dataset through the model, during which its parameters are updated to improve performance. Since this process is asynchronous, the thresholds for validation losses are adaptive and change throughout the training. The first bracket endpoint, denoted as t_0, is set to 4, and subsequent endpoints t_n are defined recursively as

$$t_n = 2 \cdot t_{n-1}. \tag{19}$$

In addition to the ASHA scheduler, a custom early stopping rule is implemented to prevent the allocation of computational resources to models that do not demonstrate improvement. Specifically, if the validation loss does not decrease over ten consecutive epochs, the training process for that model is terminated. If early stopping is not triggered, model training concludes at epoch 128.

3.4 Post-processing

In a real-life operation, the model would process whole radar images at once to detect all contaminated bins from the images. To improve efficiency, the model processes only radar bins with total reflectivity values exceeding -6 dBT. Since pure noise samples are not present in the training set and processing them would be a waste of resources, they are left out of the model's input samples. The detection mask is constructed from the processed radar bins, and the model's output is represented in each processed radar bin, forming the detection field.

Since WTC typically affects multiple radar bins (Isom et al., 2009), single-bin false positives can be removed by filtering out connected areas from the detection field that fall below the given area threshold. The resulting mask is modified to fill in all holes inside connected areas to catch any undetected WTC bins.

The detection threshold for the sigmoid function output and area threshold can be tuned to better meet the specific requirements of the application. Higher thresholding values result in fewer false positives but can filter out some WTC areas. In contrast, smaller thresholding values can make the detection mask noisier. The whole process of detecting WTC from a PPI radar image is presented in Figure 10.

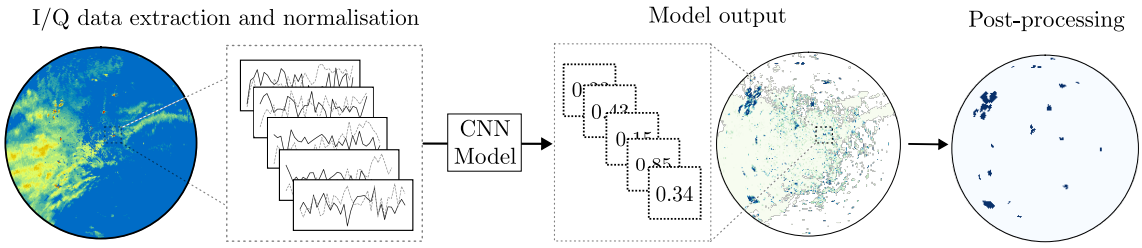

Figure 10. The flow diagram shows the whole detection procedure from I/Q extraction to the post-processing.

4 MODEL EVALUATION

4.1 Data

As described before, the dataset is constructed using manually labelled WTC data in combination with automatically sampled weather and ground clutter data. It is employed for training, hyperparameter optimisation, and testing the model. For constructing the dataset, I/Q data from two radars, KAN and UTA, is utilised. This data is acquired at fixed intervals from both radars, and thirty randomly selected radar images from each interval are used to form the dataset. The images are subsequently divided between the two radars, as detailed in Table 3.

Table 3. The data descriptions are shown for both radars in the training set.

Radar site	Radar task	Time interval	Number of images
KAN	PPI, Task A	25.11.2024 - 27.11.2024	30
UTA	PPI, Task A	31.5.2023 - 3.6.2023	30

The large amount of data in each radar image allows for efficient sampling on fewer radar images. Given the radar system's 1° resolution in the azimuthal direction and a 125-metre range resolution over a maximum range of 250 kilometres, each radar image comprises 360 azimuth intervals and 2000 range intervals, resulting in 720,000 signals per image. Although the number of radar images available for processing is limited, this is compensated for by the substantial volume of data contained within each image.

The radar images are manually labelled by marking each WTC contaminated bin, thereby generating a set of contaminated coordinates referencing the WTC samples. As wind turbines are stationary structures, it is reasonable to assume that the backscattered reflectivity exhibits similar patterns at different temporal instances. Taking this into account, each contaminated radar bin coordinate is selected from only five distinct time points. This approach is intended to assign greater weight to the rarer instances of WTC, preventing the dataset from being dominated by the most frequently occurring samples from the same areas.

Thirty images were randomly sampled from the dataset, and preliminary analysis revealed that not all of these images contained weather measurements. Consequently, radar bins with weather were exclusively extracted from the twenty images with the highest number

of radar bins with reflectivity values exceeding 8 dBZ. The weather measurements used for the dataset construction were sampled with different reflectivity values as outlined in Table 2.

Bins containing ground clutter were sampled in a similar manner. Specifically, half of the ground clutter samples were selected from radar bins that also contained weather signals, while the remaining half were obtained from ten radar images that did not exhibit any weather. These ten images were identified from the original set based on having the fewest radar bins exceeding the 8 dBZ threshold.

Manually labelled samples are augmented to increase the number of samples in the WTC class. This is accomplished by adding noise equivalent to the receiver's calibrated noise level, as expressed in Equation (15). This augmentation process is carried out twice, tripling the sample size of the WTC samples. The final dataset class partition is shown in Figure 11.

WTC	WTC (augmented)	Weather	Ground clutter
20 000	40 000	60 000	60 000

Figure 11. The data distribution shown on bar plot where each color represent one data type.

The complete input samples are generated by extracting the I/Q signals from sampled radar bins. Each signal is stored alongside the adjacent radar bin on either side in the azimuthal direction to preserve information about neighbouring bins. This process is conducted separately for each polarisation, resulting in a total of 384 data points per dataset sample.

The constructed dataset is randomly split between training, validation, and test training partitions. This is done to ensure a model is trained effectively, fine-tuned properly, and evaluated fairly (Goodfellow et al., 2016). The training partition is used to train the model. The validation partition is used to prevent overfitting during the hyperparameter optimisation, and the model convergence during training is determined from the loss calculated from the validation partition. The test data is used last to verify the training process and give final accuracy for the optimised model. This partition is shown in Figure 12.

Train data	Validation data	Test data
144 000	18 000	18 000

Figure 12. The dataset partition shown with a bar plot where each bar represents one data partition.

4.2 Evaluation criteria

The model evaluation was conducted using test partition accuracy, which measures the proportion of correctly classified instances within the test dataset. This metric is defined as

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}. \tag{20}$$

The performance of the model across different classes is commonly assessed using a confusion matrix, a tabular representation that illustrates the relationship between predicted and actual classifications. Each cell within the matrix corresponds to a specific combination of predicted and true class labels, with the numerical value indicating the count of instances that fall into that category. To facilitate comparative analysis, these raw counts can be normalised by dividing each value by the total number of predictions, yielding proportionate accuracy metrics that provide a clearer insight into the model's classification performance.

The model's performance in the intended operational scenario is evaluated using complete PPI radar scan images. This assessment ensures that the model generalises to radar images not included in the training set while also validating the effectiveness of post-processing. To achieve this, two radar images are selected from each of the three radars. These images were not used for dataset construction. One image shows a significant presence of weather, while the other has minimal or no weather. The effectiveness of WTC detection is assessed visually by comparing the detected clutter with actual wind turbine locations and reflectivity values.

4.3 Results

The model training is started with hyperparameter optimisation. A set of trial models is generated and trained to find the optimal hyperparameters for the CNN model. The final model is selected based on the lowest validation loss, which is determined by the average loss function, as defined in Equation (18), computed over the validation dataset. This procedure was done using the previously discussed combination of the Ray Tune library and the ASHA scheduler.

The hyperparameters for each trial model were selected from a predefined hyperparameter space presented in Table 4. Each convolutional kernel had a configurable number of filters and a kernel size in the time dimension. Additionally, a dropout layer was applied with a specific probability to mitigate overfitting. The learning rate and batch size were optimized to help the model's loss rapidly converge towards its minimum value. During the hyperparameter optimisation, 512 trial models were sampled from the predefined parameter space and trained. The results for the hyperparameter optimisation are shown in Figure 13.

Table 4. The optimised hyperparameters are shown with their possible values. The parameters in the optimal model are bolded.

Parameter type		**Choices**			
Conv1	Kernel size K_1	3	5	**7**	9
	Number of filters N_1	24	**32**	48	64
Conv2	Kernel size K_2	3	5	**7**	9
	Number of filters N_2	24	32	48	**64**
Dropout	Dropout probability	20%	**30%**	40%	50%
Training	Learning rate	0.1	0.01	**0.001**	0.0001
	Batch size	64	**128**	256	512

Figure 13 illustrates a linear decline in validation loss under the current scaling. Given that the x-axis is presented in a logarithmic scale, the actual loss followed an exponential decay curve, initially decreasing rapidly before gradually slowing down as training progresses. This behaviour is characteristic of a ML model loss during training, indicating that the training process proceeded as expected without significant issues.

Many trial lines terminated at the end of ASHA brackets, with distinct thresholds observable at the first three bracket endpoints: 4, 8, and 16. These clear thresholds indicate

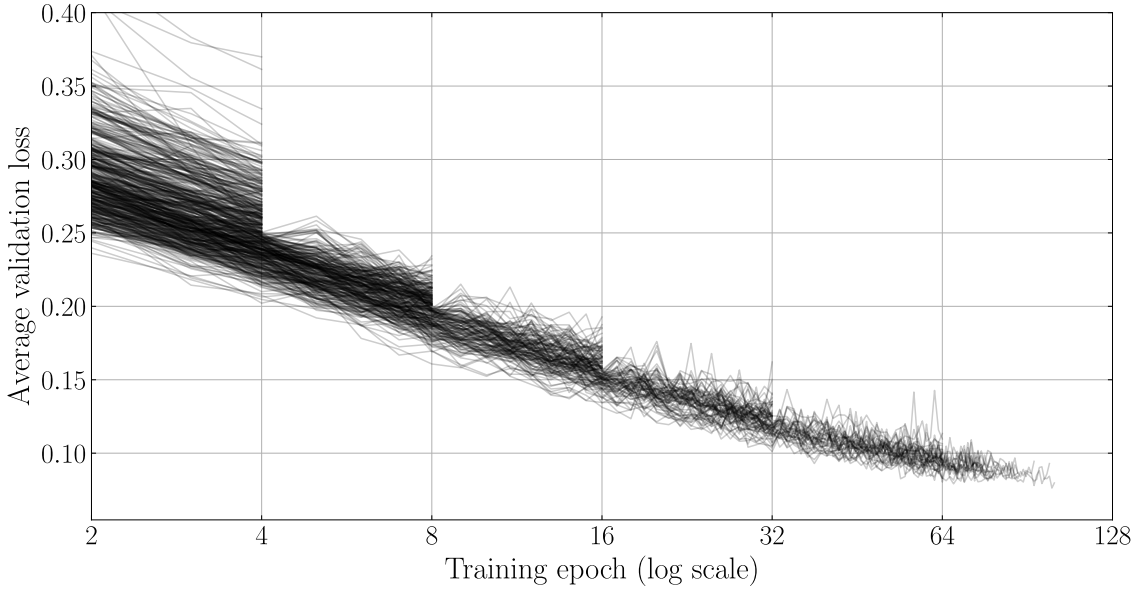

Figure 13. Validation loss represented as a function of training epoch. Each line represents a single trial during the training.

that, over the course of training, adaptive thresholding successfully identified stable loss thresholds. After epoch 64, minimal improvement is observed in the models, and the trial lines appear to vanish midway before reaching the maximum epoch. This effect is due to the early stopping rule implemented during model optimisation.

The final model is obtained from the hyperparameter tuning process, and its performance is evaluated on the test dataset to ensure that no overfitting has occurred with the validation data. The average loss and overall accuracy for each data partition are presented in Table 5.

Table 5. Average loss function output and accuracy for the final model are shown for each data partition.

Data partition	**Loss**	**Accuracy**
Training	0.049	98.56%
Validation	0.061	98.18%
Test	0.059	98.24%

Table 5 shows similar losses and accuracy scores between the different data partitions. Notably, the validation loss and test loss were almost identical. If overfitting to the validation data had occurred, the validation loss would be significantly lower than the test loss. However, in this case, the test loss was actually lower, eliminating this concern.

The accuracy scores across all data partitions were consistent. The training data achieved only approximately 0.3% higher classification accuracy than the validation and test partitions. While overfitting to the training data cannot be entirely avoided, these results suggest it was minimal. Furthermore, the findings indicate that distinct patterns were present in all data partitions, allowing the model to generalise effectively.

To further explore the model accuracy between each class, a confusion matrix was formed using the class information in the test dataset. The confusion matrix for the test data is shown in Figure 14.

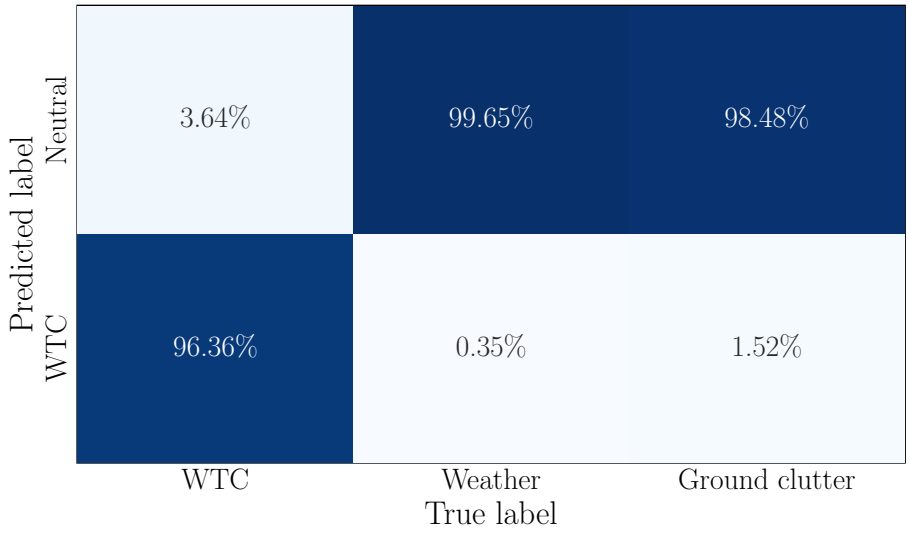

Figure 14. Classification probabilities for each class are shown in a confusion matrix. The values are normalised based on the true label.

The confusion matrix presented in Figure 14 demonstrates that weather and ground clutter samples were classified with greater consistency compared to WTC samples. Notably, the classification accuracy for WTC samples was approximately 2.5% lower than that of the neutral class. This suggests that the model more effectively differentiates between features indicative of weather or ground clutter than those characteristic of WTC.

The model exhibited exceptional performance in classifying weather samples, achieving an accuracy exceeding 99.5% on the test partition. This high accuracy indicates that the model successfully learned periodic weather variation patterns from the I/Q data. In contrast, the number of misclassified ground clutter samples was higher than that of weather samples. This suggests a greater degree of similarity between the patterns of WTC and ground clutter, which may have contributed to the increased misclassification rate.

The final model was applied to full radar PPI images to verify the results and test the model in real-life situations. The CNN output is post-processed by filtering out areas under 1.5 km^2 and filling in holes in the detected areas. The detection threshold for the sigmoid function was set to 0.75 to filter out clearly classified WTC areas. These post-processing parameters are set to preserve the detected areas in the example images.

Figures 15 - 20 present a series of polar-coordinate plots that visualise radar reflectivity data and the results of the WTC detection. The radar is positioned at the origin of each plot. Panel (a) illustrates the reflectivity values corresponding to each radar bin. In panel (b), bins identified as contaminated are replaced with white. Panel (c) applies the same procedure after the model output has been post-processed. The wind turbine locations, derived from aviation obstacle data, are depicted in panel (d) as black circles with a radius of 0.5 km. Panels (e) and (f) display the CNN model output and the post-processed output, respectively. Areas where WTC has been correctly detected are highlighted with red circles in each panel. These areas are identified visually based on reflectivity values and wind turbine locations.

The detection procedure is applied to two PPI scans for each radar: one containing weather measurements and one without. The results for the radars included in the training set, KAN and UTA, are presented in Figures 15–18. None of these images are among the thirty images used for dataset construction, ensuring that the evaluation is conducted on previously unseen data. The results for the LUO radar, which is exclusively used for testing the model, are shown in Figures 19 and 20.

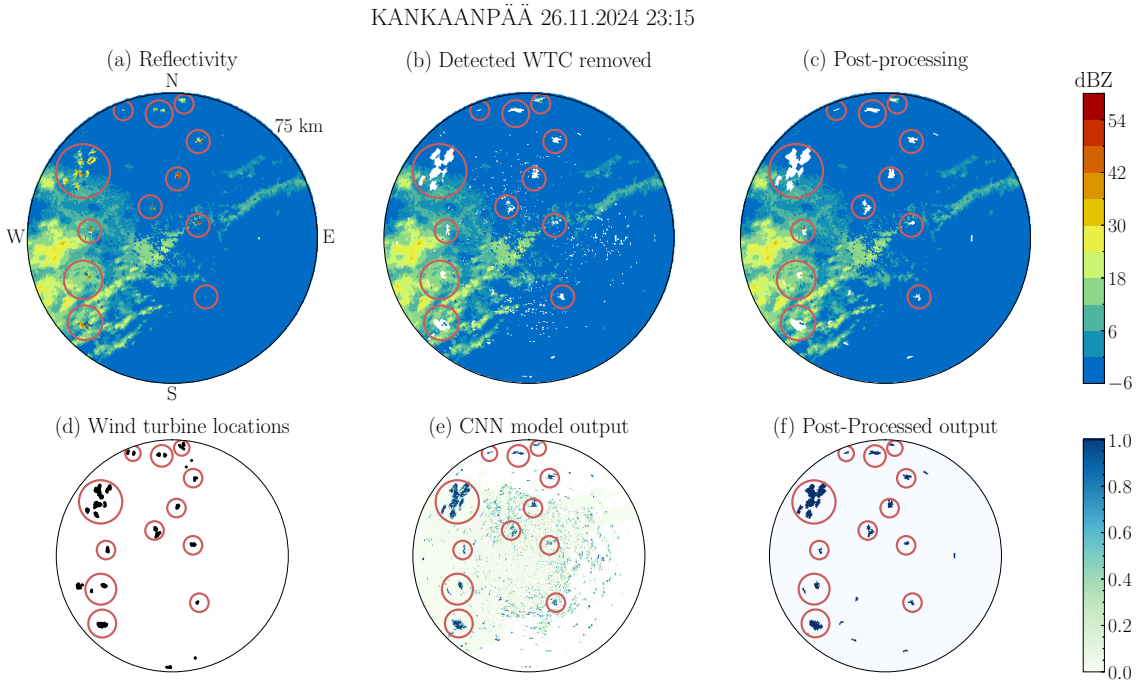

Figure 15. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

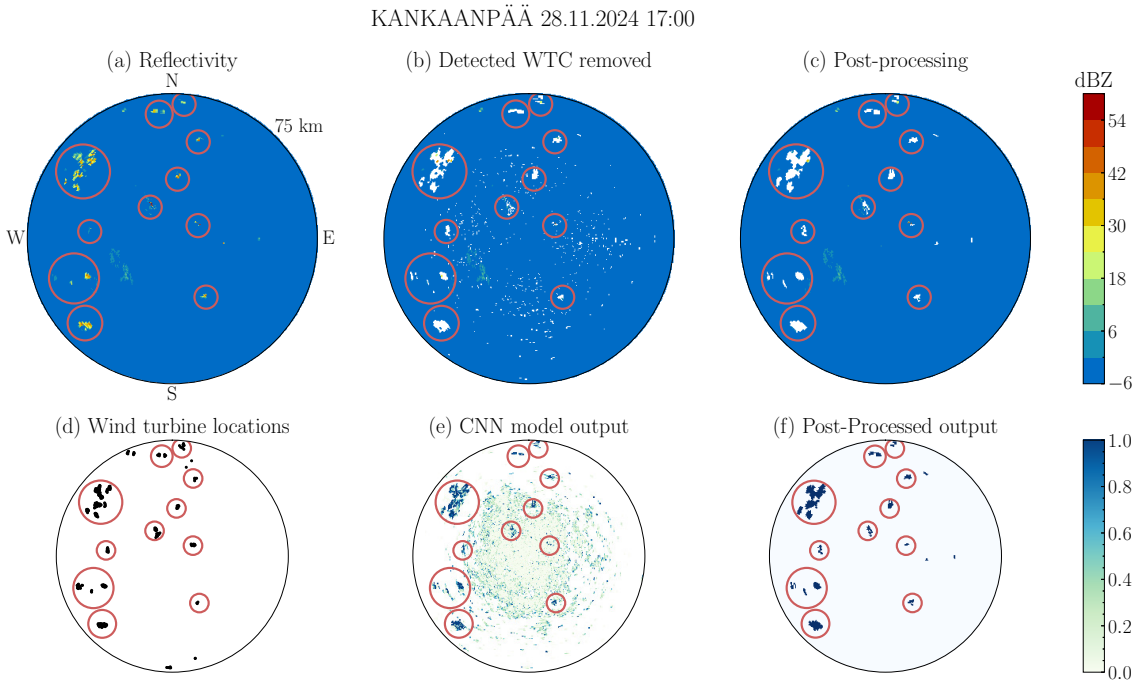

Figure 16. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

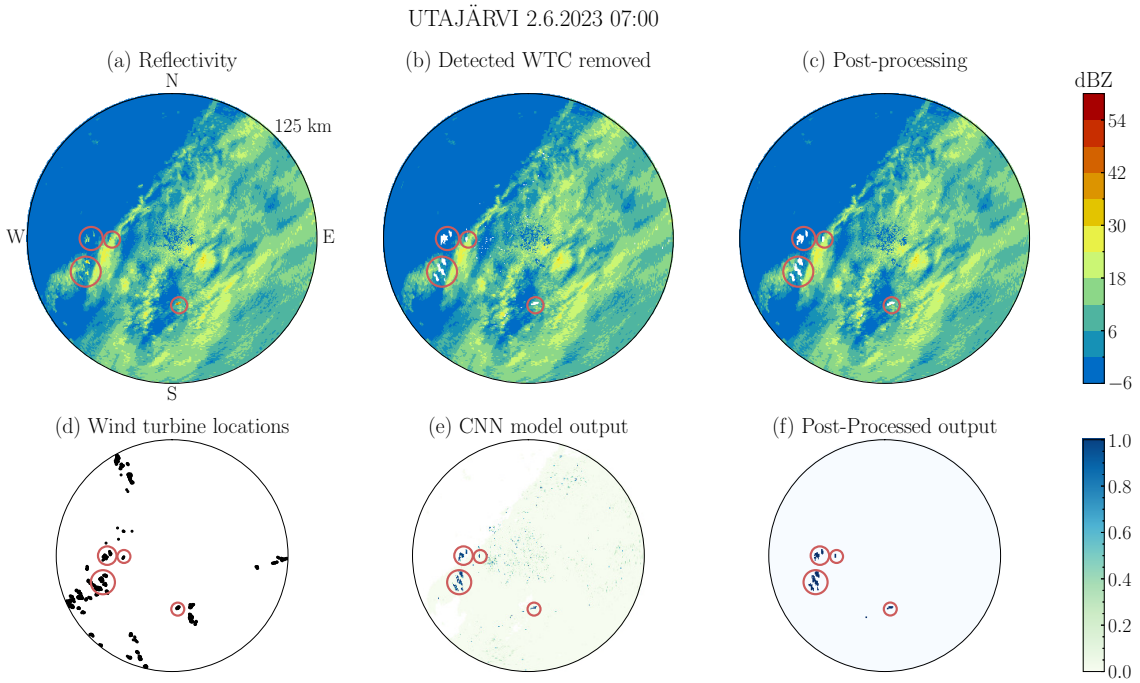

Figure 17. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

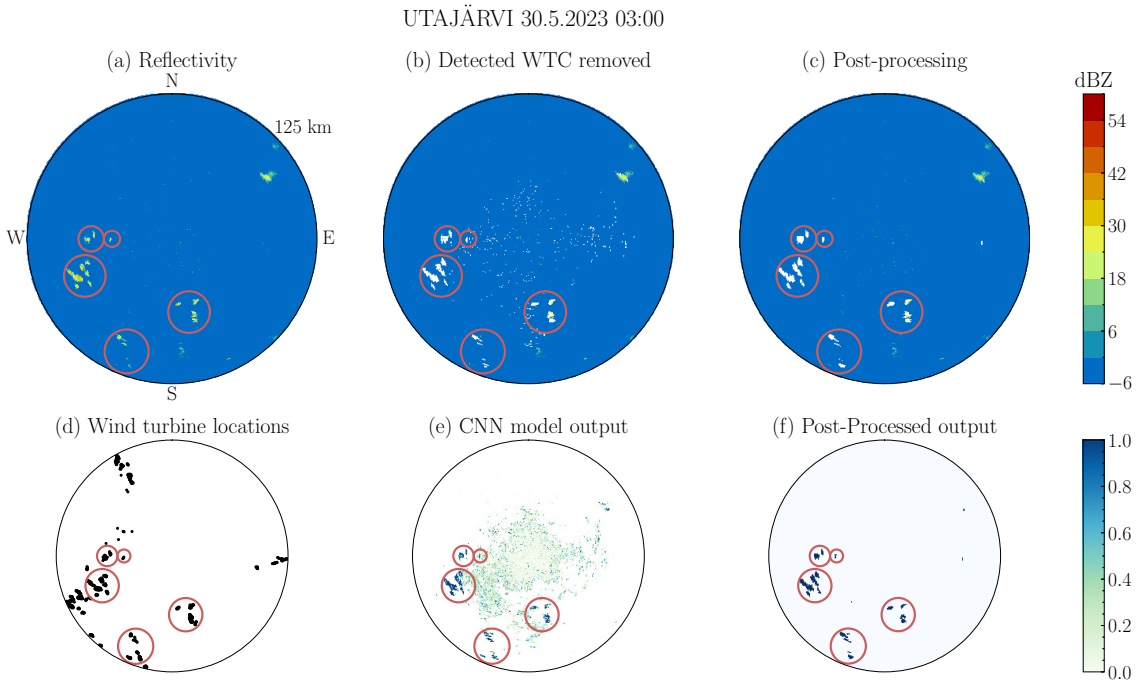

Figure 18. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

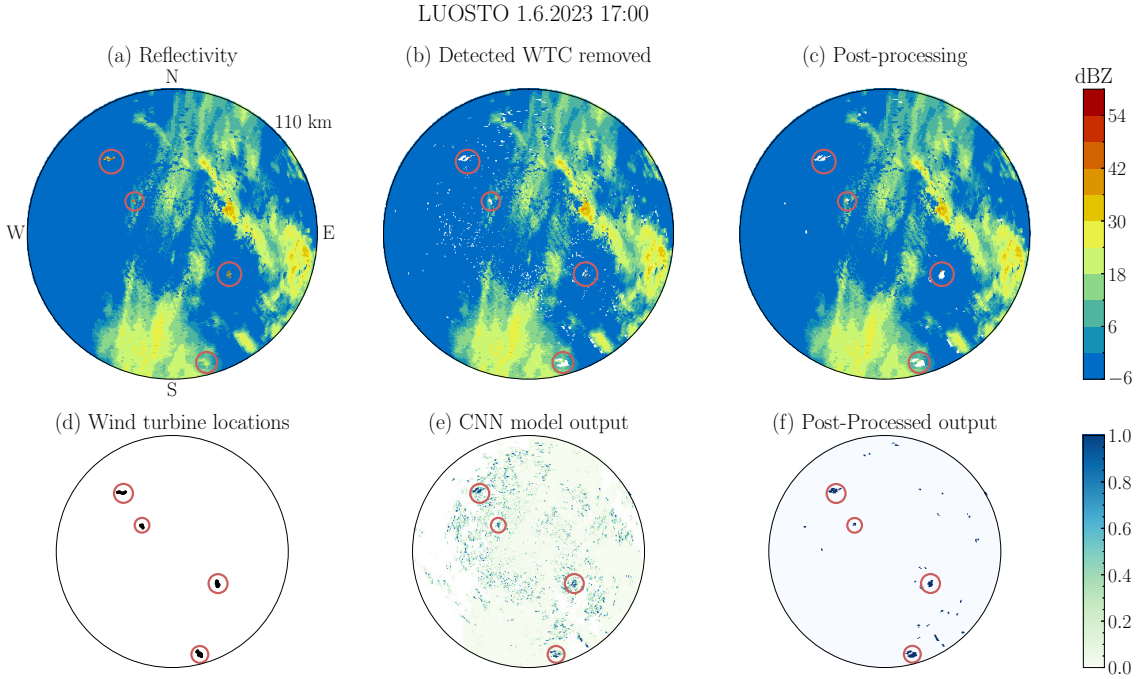

Figure 19. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

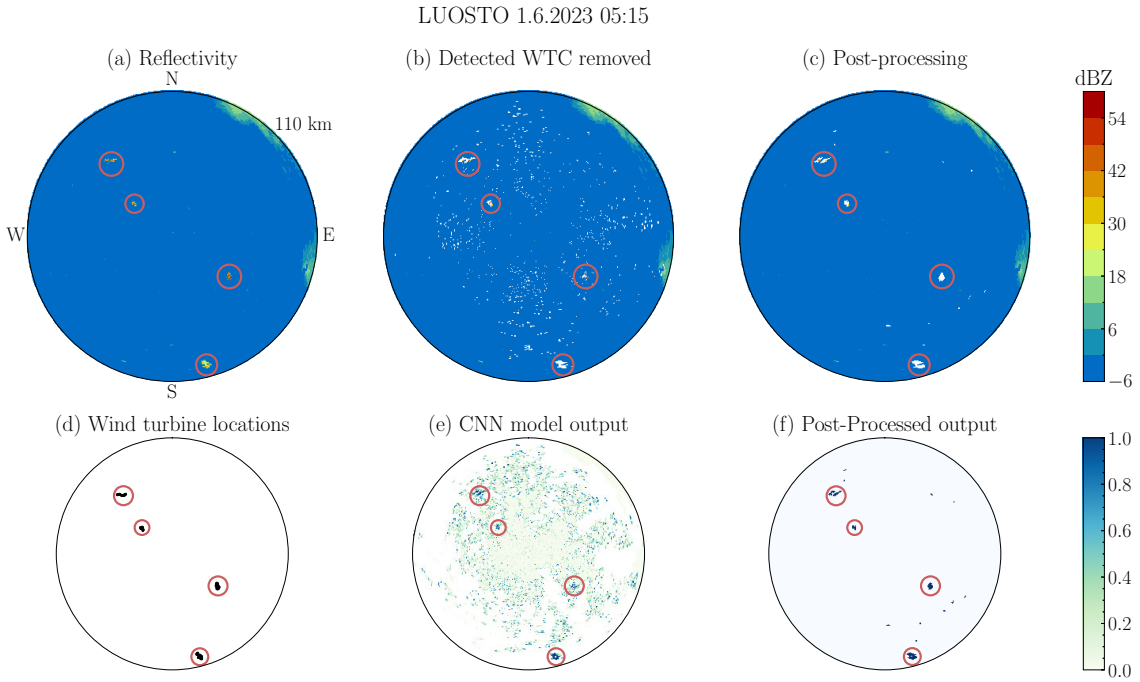

Figure 20. Polar-coordinate plots visualising radar reflectivity and WTC detection. Correctly detected WTC areas are highlighted with red circles.

In Figures 15 and 16, the detection procedure was applied to two PPI images from the KAN radar site. Figure 15(e) shows that the model successfully detects most visible WTC areas, although some are only partially captured. Similarly, in Figure 16(e), the model produces consistent results, demonstrating strong repeatability in detecting WTC areas. However, a single undetected area to the north of the radar that was correctly identified in Figure 15 suggests a potential limitation in the model's ability to distinguish WTC under certain conditions.

Examining the detection fields, Figure 15(e) shows that the model effectively separates WTC areas from background weather. In Figures 15(e) and 16(e), the model's output highlights processed ground clutter observations. Although these do not contribute to the reflectivity values, having been filtered by GMAP during processing, the edges of these cluttered regions still appear to be the main source of false detections, as seen in Figures 15(b) and 16(b). This aligns with previous observations of the model's difficulty in distinguishing ground clutter, as demonstrated in Figure 14.

Post-processing played a key role in refining the detection results. In Figure 16(c), it successfully removed most false detections and filled gaps within larger WTC areas. However, in Figure 15(c), while many faulty detections were eliminated, some correctly identified WTC regions—particularly towards the west—were also removed. This suggests a trade-off in the model's post-processing step, where reducing false detections comes at the risk of mistakenly removing true WTC areas.

At the UTA radar site (Figures 17 and 18), detection performance remains comparable to that of the other radar. However, differences in the visibility of WTC between the figures highlight the unpredictability of WTC occurrences—sometimes appearing and sometimes not. This variability arises from differences in wind turbine characteristics, which affect how they interact with the radar beam and influence their detectability at different times. Despite this, the model adapts well, successfully detecting WTC areas in Figure 18 that were not present in Figure 17.

The number of visible ground clutter targets is significantly lower in Figures 17 and 18 compared to those in Figures 15 and 16. As noted earlier, the model tends to misclassify isolated ground targets, which explains why the number of misclassified bins in Figure 18(c) is much lower than in Figure 16(c).

To further evaluate the model, Figures 19 and 20 present verification PPI images from LUO. Here, the model successfully detects all four WTC areas despite being applied to a new radar and dataset. This suggests that the model generalises well across different radar

sources, effectively recognising WTC patterns learned from previous data. However, Figures 19(e) and 20(e) still show some falsely classified ground clutter, an issue exacerbated by LUO's lower elevation angle compared to KAN and UTA. This lower angle increases the visibility of ground clutter, with some targets extending beyond 100 km.

To delve deeper into the workings of the CNN model, feature maps between the layers can be extracted and visualised. The feature maps for two selected examples are shown in Figures 21 and 22. The six complex signals constructing the input are stacked on the left side, each block containing signals for one polarisation. The two columns visualise the feature maps after the ReLU operation. The first column feature maps are organised in groups related to the model's structure. The second column of feature maps is sorted, so the feature map on top corresponds most to a sample being classified as WTC, and the feature map on the bottom corresponds most to neutral classification.

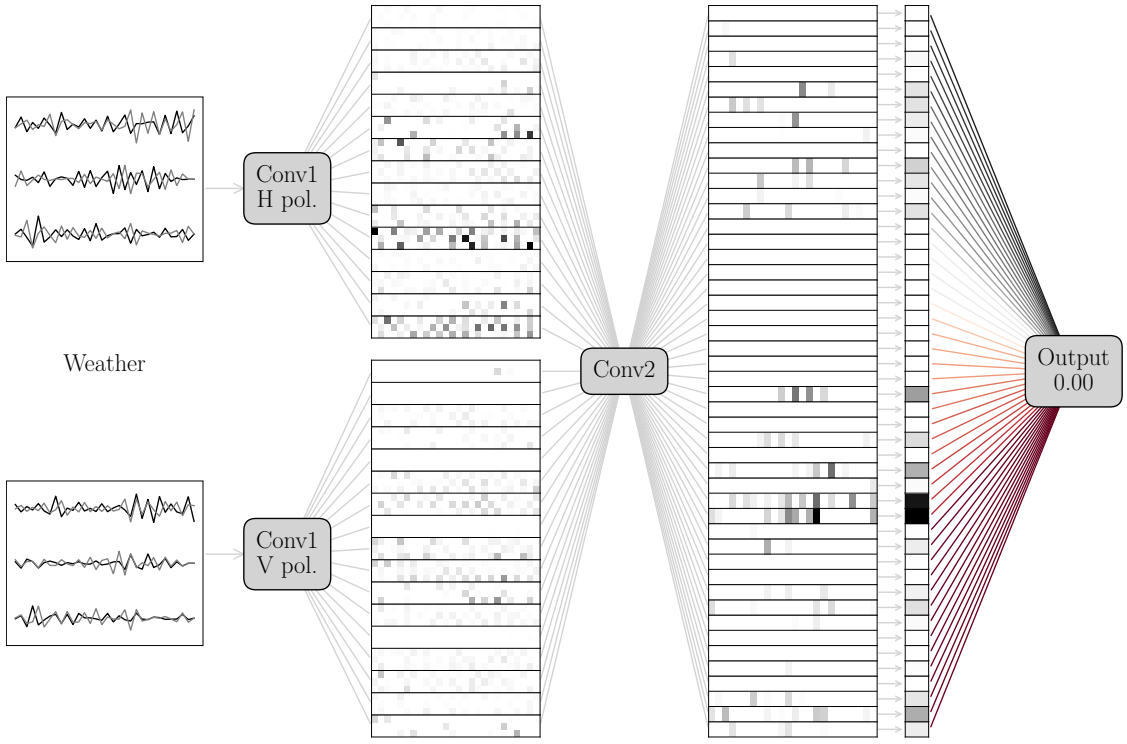

Figure 21. The first column shows input signals, the second displays feature maps after the first convolution, and the last shows feature maps after the second convolution with their averages. Black tones indicate positive weights, red tones indicate negative weights.

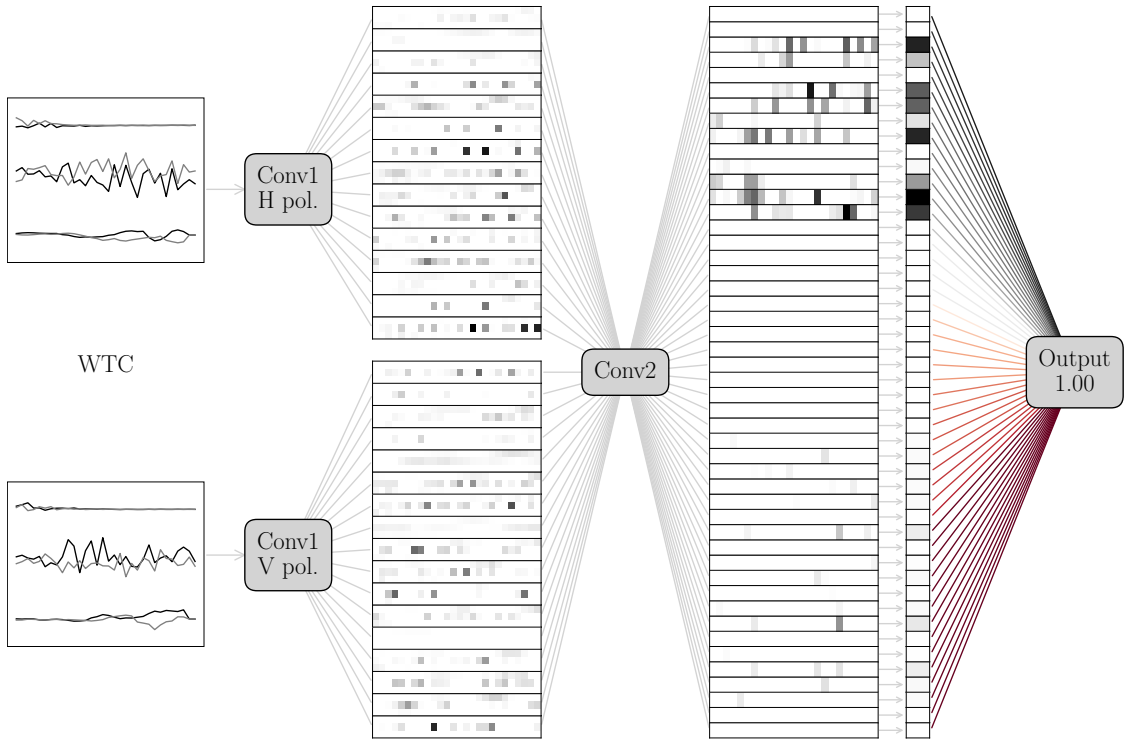

Figure 22. The first column shows input signals, the second displays feature maps after the first convolution, and the last shows feature maps after the second convolution with their averages. Black tones indicate positive weights, red tones indicate negative weights.

The global average pooling makes the features correspond to a single class, which makes interpreting feature maps more intuitive. The darker points in the second column correspond to patterns more likely to belong to WTC class, whereas the points below correspond to the neutral class. Figure 21 shows sharp variations across all three adjacent bins in azimuthal direction corresponding to rainfall with great radial velocity. Figure 22 shows variation only in the middle radar bin caused by WTC.

Approximately one-third of the feature maps appear to be dedicated to the WTC class, while the remaining feature maps correspond to the neutral class. This distribution likely arises because the patterns exhibited by the neutral class are more readily recognisable compared to those associated with WTC. However, it is crucial that the model effectively learns patterns for both classes to ensure accurate classification and robust generalisation.

5 DISCUSSION

The main objective of this thesis was to create a lightweight CNN model that could reliably detect WTC areas from weather radar I/Q data. The research was guided by key questions regarding the model's ability to differentiate WTC from weather and ground clutter, its computational feasibility for real-time radar processing, and the adequacy of short time intervals of I/Q data for training a generalizable model. Additionally, we explored whether post-processing techniques were necessary to optimize the model's integration into radar signal processing. Given the constraints of radar systems, particularly the high volume of I/Q data and its typical lack of storage, the approach was designed to operate directly within radar software while minimizing computational overhead.

In the first question, we considered the CNN model's capabilities of distinguishing WTC samples from weather and ground clutter samples. Considering the problems and challenges established in the prior research, it was challenging to propose any concrete expectations on the models' accuracy. In the end, the model achieved an overall accuracy of 98.24% on the test partition, which showed the model's impressive ability to distinguish WTC samples from the weather and ground clutter samples within the created dataset.

The introduction of the global average pooling operator marked a significant advancement towards a more generalised model with higher accuracy. By averaging over the temporal dimension, the model became capable of detecting WTC at any position within the signal, rather than merely identifying locations where clutter most frequently occurred. Furthermore, the integration of the dropout method further improved generalisation by preventing the model from memorising specific patterns. The combination of these modern ML methods facilitated the detection of more intricate patterns while maintaining robustness through regularisation.

Dropout also played a key role in enabling the model to utilise multiple features for classification by preventing over-reliance on specific features. By randomly deactivating a proportion of feature maps during training, dropout encouraged the model to develop redundant representations across different features. This, in turn, enhanced the model's ability to learn and distinguish patterns from both the WTC and neutral classes rather than becoming biased towards the dominant class. Consequently, the model became more resilient to variations in input data and generalised better when exposed to unseen data.

Another crucial factor contributing to the model's success was the chosen normalisation scheme. Normalising the data on a per-sample basis enabled the model to focus on pattern

recognition rather than absolute signal amplitude. This approach also allowed the model to process data with significantly lower signal strength than in the training set's. The processing threshold for the reflectivity value in example images was reduced from 8 dBZ to -6 dBZ, meaning that the processed signals were more than 20 times weaker than those in the training data. Given that the training set exclusively contained signals with a total power of at least 8 dBZ, this result further confirms that the model effectively generalises patterns across varying signal amplitude levels.

Since WTC comprises both contamination caused by the turbine blades and a ground clutter-like effect induced by the stationary tower, the model tends to misclassify large, isolated ground objects as WTC. During dataset construction, the WTC samples were sampled using dBT values, which do not incorporate GMAP. Consequently, numerous data points corresponding to stationary wind turbines were included in the positive class. Distinguishing these instances from similar targets, such as tall buildings, was challenging, leading the model to classify certain strong ground clutter inaccurately as WTC.

In the second research question, we considered whether it is possible to create a computationally lightweight model that is a viable candidate for future integrations. During development, we started with a simple model and gradually increased its complexity. Due to the limited dataset, we observed that increasing the model's complexity further caused overfitting, and without modifications to the dataset, it would be challenging to increase the complexity further.

The final model contained only approximately 2000 parameters, making it many magnitudes smaller than many modern CNN models. Since the exact limitations of radar hardware were unresolved, there was no reason to set a hard cap on the parameter count. If the model is too computationally exhausting, a new model can be trained by restricting the complexity based on the resolved limitations.

The third question pondered the limitations proposed by data availability. The data used for the dataset construction was only gathered from two short time periods. The limited number of different weather phenomena and WTC scenarios could lead to a repeating pattern, making learning general patterns harder.

To use the available data efficiently, the dataset construction went through many iterations. As a single factor, dataset quality seemed to be the most important factor in producing a robust model that could generalise patterns in the data. Small improvements in the dataset construction, such as removing falsely labeled data points, improved the performance of the final model on the example images.

The primary challenge affecting the accuracy of the model for potential operational use arose from its inability to detect weak WTC regions. In certain instances, the model failed to identify WTC near known wind turbine locations, leading to false negative classifications. The dataset was initially constructed using only bins where reflectivity exceeded dBZ > 8, which was the intended processing threshold. However, during experimentation, this threshold was lowered to dBZ < -6, resulting in improved detection performance in some areas. These variations in threshold selection may account for the model's weaker detection performance at lower reflectivity values.

In the fourth and last question, we considered the relevance of post-processing and the most effective approach to radar signal processing. The relevance of post-processing was brought up early during the development when the model accuracy was lower and the number of false detections was notably higher. The post-processing scheme was developed to reduce the number of falsely detected radar bins and reduce the number of processed radar bins later in the potential WTC removal step.

The post-processing scheme underwent many iterations during the model development. Initially, the detection mask was expanded using binary dilation, but this was later replaced with an operation that fills holes surrounded by positive detection. Depending on the WTC removal and processing, the whole procedure of removing false detection, mainly caused by ground clutter, might be unnecessary. In some cases, the post-processing removed smaller WTC areas under the area threshold, making the results arguably worse.

6 CONCLUSIONS

This thesis explored the use of a lightweight CNN for WTC detection in weather radar I/Q data. The model successfully achieved 98.24% accuracy, demonstrating its promising ability to distinguish WTC from weather and ground clutter. While dataset limitations posed challenges, the results indicate that machine learning can be a viable approach for WTC classification in radar I/Q data.

The model's strong performance was driven by key techniques such as global average pooling, dropout, and per-sample normalization. These methods enhanced generalization, reduced overfitting, and allowed the model to detect WTC across varying signal amplitudes. With only 2000 parameters, the model remains computationally efficient, making it a feasible candidate for real-time radar integration.

Future improvements should focus on expanding dataset diversity to improve generalization, particularly in detecting weaker WTC regions. Additionally, refining normalization strategies and adapting post-processing techniques could enhance detection reliability. Further optimization will be essential for potential deployment in operational radar systems.

REFERENCES

Ji-Seon Bang, Min-Ho Lee, Siamac Fazli, Cuntai Guan, and Seong-Whan Lee. Spatio-spectral feature representation for motor imagery classification using convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33 (7):3038–3049, 2022.

Michael Botsch. *Machine learning techniques for time series classification*. Künstliche Intelligenz Digitalisierung. Cuvillier Verlag, second edition, 2023.

James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

Zeyad Emam, Andrew Kondrich, Sasha Harrison, Felix Lau, Yushi Wang, Aerin Kim, and Elliot Branson. On the state of data in computer vision: Human annotations remain indispensable for developing deep learning models, 2021. arXiv:2108.00114 [cs.CV].

Frédéric Fabry. *Radar Meteorology: Principles and Practice*. Cambridge University Press, 2015.

Fintraffic. Obstacle data. ais.fi/en/obstacle-data-csv-files, 2025. Online, Accessed: 2025-02-04.

FMI. FMI radar network. en.ilmatieteenlaitos.fi/fmi-radar-network, 2025. Online, Accessed: 2025-01-21.

Beatriz Gallardo-Hernando, Felix Perez-Martinez, and Fernando Aguado-Encabo. Wind turbine clutter detection in scanning weather radar tasks. In *Proceedings of the ERAD*, volume 610, 2010.

Zhiqiang Gong, Ping Zhong, and Weidong Hu. Diversity in machine learning. *IEEE Access*, 7:64323–64350, 2019.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

Olatz Grande, Josune Cañizo, Itziar Angulo, David Jenn, Laith R Danoon, David Guerra, and David de La Vega. Simplified formulae for the estimation of offshore wind turbines clutter on marine radars. *The Scientific World Journal*, 2014(1):982508, 2014.

Jonathan J Helmus and Scott M Collis. The Python ARM Radar Toolkit (Py-ART), a library for working with weather radar data in the Python programming language. *Journal of open research software*, 4, 2016.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors, 2012. arXiv:1207.0580 [cs.NE].

Kenta Hood, Sebastián Torres, and Robert Palmer. Automatic detection of wind turbine clutter for weather radars. *Journal of Atmospheric and Oceanic Technology*, 27:1868–1880, 2010.

Nawal Husnoo, Timothy Darlington, Sebastián Torres, and David Warde. A neural network quality-control scheme for improved quantitative precipitation estimation accuracy on the U.K. weather radar network. *Journal of Atmospheric and Oceanic Technology*, 38(6):1157 – 1172, 2021.

Bradley Isom, Robert Palmer, G. Secrest, R. Rhoton, D. Saxion, T. Allmon, J. Reed, T. Crum, and R. Vogt. Detailed observations of wind turbine clutter with scanning weather radars. *Journal of Atmospheric and Oceanic Technology*, 26, 2009.

Gyu Il Kim and Kyungyong Chung. Extraction of features for time series classification using noise injection. *Sensors*, 24(19), 2024.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. arXiv:1412.6980 [cs.LG].

Sudhakar Kumawat and Shanmuganathan Raman. Depthwise-stft based separable convolutional neural networks. In *Proceedings of the ICASSP*, pages 3337–3341. IEEE, 2020.

S Kunze and B Saha. Drone classification with a convolutional neural network applied to raw IQ data. In *Proceedings of the AT-AP-RASC*, pages 1–4. IEEE, 2022.

Jussi Leinonen, Dmitri Moisseev, Matti Leskinen, and Walter A Petersen. A climatology of disdrometer measurements of rainfall in Finland over five years with implications for global radar observations. *Journal of Applied Meteorology and Climatology*, 51(2): 392–404, 2012.

John J Lemmon, John E Carroll, Frank H Sanders, and D Turner. Assessment of the effects of wind turbines on air traffic control radars. *US Department of Commerce*, 2008.

Liam Li, Kevin Jamieson, Afshin Rostamizadeh, Ekaterina Gonina, Moritz Hardt, Benjamin Recht, and Ameet Talwalkar. A system for massively parallel hyperparameter tuning, 2020. arXiv:1810.05934 [cs.LG].

Richard Liaw, Eric Liang, Robert Nishihara, Philipp Moritz, Joseph E. Gonzalez, and Ion Stoica. Tune: A research platform for distributed model selection and training, 2018. arXiv:1807.05118 [cs.LG].

M. Lin. Network in network, 2013. arXiv:1312.4400 [cs.NE].

John S Marshall and W Mc K Palmer. The distribution of raindrops with size. *Journal of Atmospheric Sciences*, 5(4):165–166, 1948.

Reza Moradi, Reza Berangi, and Behrouz Minaei. A survey of regularization strategies for deep models. *Artificial Intelligence Review*, 53(6):3947–3986, 2020.

F. Nai and S. Torres. A wind turbine clutter mitigation solution for the nexrad network. In *Proceedings of the ERAD*, 2024.

Feng Nai, Sebastián Torres, and Robert Palmer. On the mitigation of wind turbine clutter for weather radars using range-doppler spectral processing. *IET Radar, Sonar Navigation*, 7:178–190, 2013.

L. Norin. A quantitative analysis of the impact of wind turbines on operational doppler weather radar data. *Atmospheric Measurement Techniques*, 8(2):593–609, 2015.

L. Norin. Wind turbine impact on operational weather radar I/Q data: characterisation and filtering. *Atmospheric Measurement Techniques*, 10(5):1739–1753, 2017.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library, 2019. arXiv:1912.01703 [cs.LG].

BS Randhawa and R Rudd. Rf measurement assessment of potential wind farm interference to fixed links and scanning telemetry devices. *ERA Report*, 2009.

R.M. Rauber and S.W. Nesbitt. *Radar Meteorology: A First Course*. Advancing Weather and Climate Science. Wiley, 2018.

Jenna Ritvanen, Pauli Anttonen, Harri Hohti, Mikko Kurri, and Annakaisa von Lerber. Operational wind turbine clutter removal in the Finnish weather radar network: Methodology and impact on data quality. Poster presented at ERAD, 2024.

Dipak L Sengupta and Thomas BA Senior. Electromagnetic interference to television reception caused by horizontal axis windmills. *Proceedings of the IEEE*, 67(8):1133–1142, 1979.

AD Siggia and RE Passarelli. Gaussian model adaptive processing (GMAP) for improved ground clutter cancellation and moment calculation. In *Proceedings of the ERAD*, volume 2, pages 421–424, 2004.

Vaisala. *RVP900 Digital Receiver and Signal Processor*. Vaisala, 2021.

Jiaxing Wang, Dazhi Wang, Sihan Wang, Wenhui Li, and Keling Song. Fault diagnosis of bearings based on multi-sensor information fusion and 2D convolutional neural network. *IEEE Access*, 9:23717–23725, 2021.

Mengze Wu, Yongdi Lu, Wenli Yang, and Shen Yuong Wong. A study on arrhythmia via ecg signal classification using the convolutional neural network. *Frontiers in computational neuroscience*, 14:564015, 2021.