Detection and Measurement of Inhibition Zones in AST Using Deep Learning

Bachelor's Thesis

Computer Applications

Spring 2025

Veronika Jadrná

HAMK
Häme University
of Applied Sciences

Abstract

DP	Computer Applications
Author	Veronika Jadrná Year 2025
Subject	Detection and Measurement of Inhibition Zones in AST Using Deep Learning
Supervisors	Mazhar Mohsin

The antimicrobial resistance is one of the current global health treats in the 21st century. One of the common methods for classifying bacteria as resistant or susceptible is antimicrobial susceptibility testing. This method involves measuring the inhibition zones created by the reaction of bacteria to different antimicrobials and measuring them based on the interpretation guidelines and breakpoints. One of the disadvantages of this method is a manual and time-consuming measurement process.

This study presents an alternative solution for the detection and measurement of inhibition zones using deep learning and aims to contribute to more efficient and faster laboratory diagnostics. During this study several object detection models were implemented for the identification of inhibition zones from agar plate images. After the evaluation of different models including YOLO, SSD and Faster R-CNN, it was found out that the YOLO models achieved the best performance results. Specifically, YOLOv8m achieved the best recall of 100% and mean average precision of 85.3% at IoU threshold between 0.5 and 0.95.

Subsequently, three best models were used for bounding box measurement analysis. The measurement task included width calculation of the bounding boxes in millimetres and their comparison to ground truth annotations. The results suggest that YOLOv8m model achieved best localization accuracy with 88.8% predictions falling within 1 mm deviation and 98.9% predictions falling within 2 mm deviation. Furthermore, the results showed that the maximum difference between predicted bounding boxes and their annotations was 3 mm.

These findings support a conclusion that the deep learning models can achieve high detection and measurement results and have a potential for practical use in laboratory environment. However, it is important to point out, that while these results are promising, further improvements in terms of localization accuracy of bounding boxes are inevitable to ensure more precise and accurate laboratory diagnostics.

Keywords	Antimicrobial Resistance, Antimicrobial Susceptibility Testing, inhibition zones, object detection, measurement.
Pages	32 pages and appendices 3 pages

Glossary

Antimicrobials	Common medical substances such as antibiotics, antifungals, antivirals and antiparasitic drugs.
AMR	Antimicrobial resistance occurs when microorganisms develop an ability to resist antimicrobials that were previously effective.
AST	Antimicrobial susceptibility testing is a method used for classifying bacteria as resistant or susceptible against different antimicrobials.
Inhibition Zone	A circular area around antibiotic disk where the bacterial growth has been prevented.
Agar Plate	A shallow container called petri dish filled with solid nutrient medium used to grow bacteria.
Disk Diffusion Method	A standard AST method where antibiotic disks are placed on agar plates previously filled with bacteria.
Breakpoint	A defined concentration of antibiotic used to differentiate susceptible, intermediate and resistant bacteria.
EUCAST	An organization that provides guidelines and breakpoints for AST.
CLSI	An organization that sets guidelines for AST procedures and interpretations.
Klebsiella	Increasingly resistant bacteria causing pneumonia, bloodstream infections, meningitis and wound infections.
Annotation	A process of labeling specific features in data to train a machine learning model.
Bounding box	A rectangular shape commonly used to capture objects of interest during annotation process.
Deep Learning	A subset of machine learning which learns from data by using artificial neural networks.

Abstract

Object Detection	A computer vision task used to localize and classify objects in images.
YOLO	A popular real-time and single-stage object detection model.
SSD	Another real-time and single stage object detection model.
Faster R-CNN	A two-stage object detection model.
Recall	An evaluation metric used to monitor a model´s ability to detect relevant objects.
mAP	A mean average precision is a metric which measures the model´s ability to accurately localize and detect objects across different confidence thresholds.
IoU	Intersection over union compares the ground truth annotation and a model´s prediction by calculating the overlap between them.

Table of Contents

Figures

Tables

Equations

Appendices

1 Introduction

Antimicrobials, such as antibiotics, antifungals, antivirals and antiparasitic drugs are common medical substances used for prevention and treatment of various infections (World Health Organization, 2023). They are used to prevent and eliminate the growth of microorganisms, including bacteria, viruses, fungi and parasites (World Health Organization, 2023; Antimicrobial Resistance | EFSA, 2025) .

The introduction of antimicrobials has greatly influenced the medical practises and has enabled a control over major infectious diseases which used to lead to numerous deaths in the past. As a result of the first antimicrobial drug, introduced in 1910, the average life expectancy was increased by 23 years (Hutchings et al., 2019). Another, and perhaps the most notable example is the discovery of penicillin in 1928. This discovery is considered one of the most important advances in therapeutic medicine (American Chemical Society, n.d.).

The antimicrobials are undoubtedly an inevitable part of modern medicine. However, their misuse and overuse can have detrimental effects. According to the World Health Organisation (WHO), antimicrobial resistance belongs to main global health threats in the 21st century. Antimicrobial resistance (AMR) occurs when the antimicrobial medicines are no longer effective. This means the microorganisms develop the ability to resist the antimicrobials designed to eliminate them. (World Health Organization, 2023; Tang et al., 2023)

In 2019 alone, the antimicrobial resistance led to 1.27 million fatal cases and is expected to result in 10 million annual deaths by 2050 (Tang et al., 2023). With the lack of effective antimicrobials, not only the treatments of infections, but also many medical procedures such as chemotherapy, diabetes treatment and organ transplantation are put under a serious risk (World Health Organization, 2023).

One of the current solutions for identifying effective antimicrobials for the treatment of individual patients is Antimicrobial Susceptibility Testing (AST). Widely used method in AST is disk diffusion, also known as Kirby-Bauer test. This method includes the placement of antibiotic-impregnated disks on agar plates previously infected with bacteria (Rivera et al., 2023). As the drug diffuses it creates inhibition zone around each disk. After the incubation, which is usually from 16 to 18 hours, the diameter of each inhibition zone is measured.

Depending on the diameter size of the inhibition zone the antimicrobials are classified as susceptible, intermediate or resistant. (Vazquez-Pertejo & Bush, 2025)

However, since the measurement of the inhibition zones is done manually by experts, the process is very slow and limits the efficiency needed in microbiology lab, especially now, with the current antimicrobial resistance crisis we are facing. Therefore, automating this process with the use of computer vision could have a significant impact on the current methods of AST.

This study aims to explore and evaluate different deep learning models for the automation of antimicrobial susceptibility testing. The goal of the study is to detect the inhibition zones in agar plates and measure their diameter. By creating an automated solution, the study aims to contribute to ongoing efforts to combat antimicrobial resistance, more efficient and faster antimicrobial susceptibility testing and reliable laboratory diagnostics.

To meet the objectives outlined above, the study aims to address the following questions:

- How can deep learning be used for automation of AST?
- How well do deep learning models detect relevant objects in images ?
- How can an object detection model be optimized to measure inhibition zones and how accurate are these measurements?

2 Antimicrobial Resistance (AMR)

Antimicrobial Resistance is becoming a global threat that makes previously treatable infections more difficult to manage. This chapter outlines the main concepts behind antimicrobial resistance, its global impact and the most relevant resistance mechanisms.

2.1 Introduction to AMR

Antimicrobial resistance occurs when the microorganisms develop the ability to tolerate and resist the antimicrobial medications such as antibiotics, antifungals, antivirals and antiparasitic drugs that were designed to kill them (Salam et al., 2023).

The antimicrobial medication has been used for many years now and led to significant improvements in treating many infectious diseases. Since the discovery of penicillin in 1928 the course of medicine was enormously changed. The infections that were previously severe and often resulted in fatal cases, were easily treated with this new and powerful antibiotic. (Kalvaitis, 2023)

Interestingly, already during that era researchers noticed, that certain bacteria like B. Coli are resistant to penicillin by producing a specific enzyme, also called penicillinase, that destroys this antibiotic (Abraham & Chain, 1940). This proves the early signs of resistance mechanism in bacteria and indicates that the antimicrobial resistance is natural and ancient occurrence (Endale et al., 2023).

However, the current stage of antimicrobial resistance is far more severe. Even though the AMR occurs naturally over time due to genetic change in pathogens, the human activity such as incorrect use of antimicrobials has led to rapid increase of antimicrobial resistance (Hutchings et al., 2019; World Health Organization, 2023). Furthermore, the incorrect prescription also results in increased infection severity, increased disease length and cost of treatment (Moyo et al., 2023).

The misuse and overuse of antimicrobials remain among the main factors for developing AMR (World Health Organization, 2023). Since we were relying on antimicrobials to treat most of the infectious diseases for a long time now, it is natural that the bacteria started developing its resistance to these substances. Simply put, the more antibiotics we use, the faster resistance develops in bacteria (Ventola, 2015).

2.2 Global Impact

According to World Health Organisation (WHO) the antimicrobial resistance has become one of the top global public health threats (World Health Organization, 2023). With the rapid rate of AMR, the infections and death rates are being closely monitored. In the European region, there are already over 700 000 cases of infections that are resistant to antibiotics. These infections have contributed to over 33 000 deaths on an annual basis. (Ahmed et al., 2024)

People living in low-resource settings are especially vulnerable to both the causes and consequences of AMR. Limited access to clean water, insufficient sanitation and hygiene, poor preventions, unaffordable vaccines and medicines and lack of awareness are all contributing factors. The resistance to dangerous diseases such as malaria, tuberculosis and HIV have serious impact in these settings. Over 480 000 cases of drug-resistant TB were reported in 2013, where the majority was not treated. Furthermore, it has been observed that in India, once in every nine minutes a child death occurs from bacterial infections that are resistant to antibiotics. (Ahmed et al., 2024)

Unfortunately, the future does not look optimistic either. It has been reported, that there were 1.27 million fatal cases that were directly caused by AMR in 2019, and the numbers are expected to raise to 10 million by 2050 (World Health Organization, 2023; Breijyeh et al., 2020). As a result of AMR, minor injuries and infections could become deadly once again. Life-saving medical procedures including organ transplantation, chemotherapy, caesarean section could become truly dangerous (Ahmed et al., 2024). This fast decline in effective antimicrobial medicine could set us decades of medical advancements behind and the bacterial infections could become deadly again.

2.3 Resistance in Bacteria

The bacterial resistance differs based on the structure and characteristics of the specific bacteria. The main distinction is between Gram-positive and Gram-negative bacteria which react differently to antibiotics. This chapter explains the difference between these groups and introduces several clinically important gram-negative organisms.

2.3.1 Gram-positive vs Gram-negative Bacteria

Gram-positive and Gram-negative bacteria are two different types of bacteria. One of the common techniques to classify bacteria is Gram staining. The gram refers to a staining method, introduced by Hans Christian Gram in 1884. This method includes applying a coloured dye, usually of crystal violet colour, to the bacteria which is afterwards viewed under a microscope to see its reaction to the stain. (Breijyeh et al., 2020)

The gram-positive bacteria have a thick cell wall causing them to keep the dye in gram testing. This makes them appear blue or purple when viewed under a microscope. In contrast, Gram-negative bacteria have a thinner cell wall. Due to this, they do not hold the dye and appear red or pink under a microscope. (Breijyeh et al., 2020)

Both bacteria can be treated with antibiotics. However, the treatment for gram-negative infections can be more challenging and limited. Gram-negative bacteria can cause serious infections such as pneumonia, bloodstream infections, wound infections and many more. The infections from these types of bacteria are becoming more and more difficult to treat due to antimicrobial resistance. (Breijyeh et al., 2020)

The main reason for antibiotic resistance in gram-negative bacteria is their outer membrane. This outer membrane is a unique and protective layer that distinguishes the Gram-negative bacteria from the Gram-positive. The antibiotics need to pass the outer membrane to reach their target. Some small antibiotics can pass through easily, but others need to enter through special channels called porins. The problem is that if the bacteria block these porins or changes the structure of the outer membrane, the antibiotics cannot get inside anymore. The gram-positive bacteria do not have this extra layer, which is why the gram-negative bacteria are generally more resistant. (Breijyeh et al., 2020)

2.3.2 Clinically Relevant Gram-negative Bacteria

There are different types of gram-negative bacteria. Among the ones which are becoming increasingly resistant is Klebsiella. This bacterium is usually found in human stool and can cause various infections. Klebsiella can cause pneumonia, bloodstream infections, meningitis and wound infections. This infection can spread easily from person-to-person. Other factors contributing to the spread of Klebsiella include, contact with contaminated water or soil, contact with contaminated equipment and injury or surgery wounds. (CDC, 2024b)

Another resistant gram-negative bacterium is Acinetobacter. This group of germs can be found in water and soil. The infections caused by Acinetobacter usually occur inside healthcare settings. Among the common infections from these bacteria are blood infections, urinary tract infections and lung and wound infections. People who are on breathing machines, have wounds from surgery, weakened immune system and chronic lung disease or diabetes are at higher risk of getting this infection. (CDC, 2024a)

Pseudomonas aeruginosa infections also belong to a group of increasingly resistant bacteria. These infections also often occur in healthcare settings. P. Aeruginosa causes infections in the blood, lung, urinary tract and other parts of the body after surgery. To reduce the risk of this bacterial infection it is important to keep good hand hygiene and good infection control inside the medical facilities. (CDC, 2024c)

One of the most common Gram-negative bacteria responsible for many diseases is Escherichia coli. This bacterium causes many types of serious infections including bloodstream infections. The problem of this bacteria is that it is becoming resistant by the production of enzymes that break down antibiotics. People that are at increased risk are children younger than 5 years, adults over 64, people with weak immune system and foreign travellers (CDC, 2024d). Among the various types of E. coli, the ExPEC is being increasingly reported as a cause of serious illness, especially in newborns (Poolman & Wacker, 2016).

3 Antimicrobial Susceptibility Testing (AST)

One of the main techniques for monitoring antimicrobial resistance and choosing the correct treatment for patients with infectious diseases is antimicrobial susceptibility testing. This chapter provides the procedure description and explores the most common methods for practising AST.

3.1 The Role and Importance of AST

Antimicrobial susceptibility testing is a laboratory method used to determine which antibiotics a certain microorganism is susceptible to (Wenzler et al., 2023). In this context, the term susceptible refers to the microorganism´s vulnerability to a specific antibiotic drug.

The AST method has been used for many decades now and is the main technique to find the best therapeutic option for treatment of various infections. This procedure provides insights into resistance patterns and supports infection control. (Gajic et al., 2022)

The role of antimicrobial susceptibility testing has become increasingly important, especially with the current AMR crisis. Programs such as WHO, Global Antimicrobial Resistance and Use Surveillance System (Glass) and European Antimicrobial Resistance Surveillance Network (ESRS-Net) rely on results from AST to monitor resistance patterns and potential outbreaks (European Antimicrobial Resistance Surveillance Network (EARS-Net), 2017; Global Antimicrobial Resistance and Use Surveillance System (GLASS), n.d.).

Several different methods are used to monitor the activity of antibiotics against different bacteria and provide the data needed for clinical decisions. Over time, many different techniques were developed. The methods can be divided into different groups such dilution and diffusion methods, qualitative and quantitative methods and overnight and rapid methods. (Murray, 2015)

3.2 Methods of AST

The methods used for antimicrobial susceptibility testing differ in terms of speed, accuracy, complexity and cost. The choice of method used usually depends on the available resources and the type of microorganism tested. The most common methods used are disk

diffusion and broth dilution. Both methods are phenotypic which means they examine the effect of antibiotic on bacterial growth.

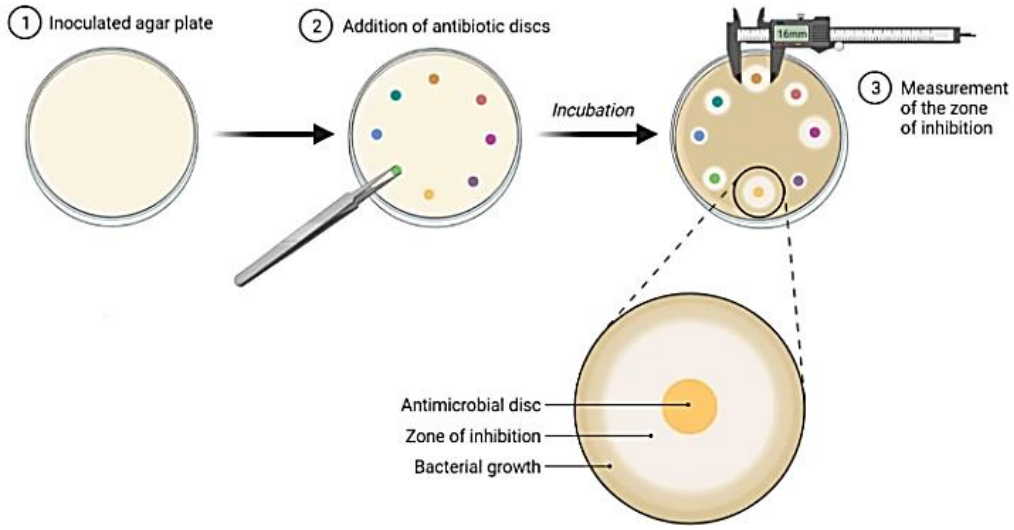
3.2.1 Disk Diffusion Method (Kirby-Bauer)

The disk diffusion method is one of the widely used methods for antimicrobial susceptibility testing in clinical microbiology laboratories. This method provides reliable qualitative results for the treatment of numerous infections.

This method is based on placing a paper disk containing predefined amount of antibiotics onto an agar plate, shallow container filled with solid nutrient medium, that has been covered with bacteria. The antibiotic spreads though the agar plate and affects the bacterial growth. As the drug spreads, it creates so called inhibition zones which are circular areas around the antibiotic disk. These zones represent the area where the bacterial growth has been prevented. (Gajic et al., 2022; Rivera et al., 2023)

These inhibition zones are measured in millimetres after the incubation time, which is usually from 16 to 18 hours (Vazquez-Pertejo & Bush, 2025). The step-by-step process of this method can be visually seen in Figure 1 (Sharma, 2022).

Figure 1. Kirby-Bauer Method process



After the measurement process, the inhibition-zone diameters are classified as susceptible (S), intermediate (I) or resistant (R) according to regularly updated clinical breakpoints.

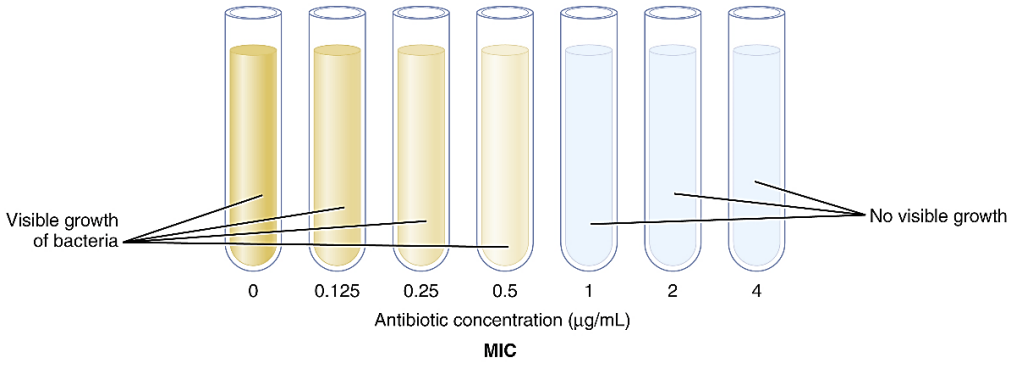
These breakpoints are the values used to determine the group a specific bacterium falls into, based on the measured inhibition zone diameter. (Vazquez-Pertejo & Bush, 2025)

3.2.2 Dilution Methods (Broth, Agar)

Other common methods in AST are broth and agar dilution. These methods provide quantitative results (MIC) as well as qualitative results (category interpretation). The minimum inhibitory concentration (MIC) represents the minimal concentration of an antibiotic at which the microorganism is prevented from growing. The MIC is highly important for interpreting bacterial susceptibility, especially when disk diffusion cannot be applied. (Tankeshwar, 2013)

The broth dilution method is used to find the minimal amount of antibiotic that can stop bacteria from growing. Firstly, the antibiotic needs to be diluted in a liquid called broth, which provides nutrients to make the bacteria grow. This is done in a series of containers, usually test tubes, where each one has less antibiotic than the one before. A small amount of bacteria is added to each of the test tubes and afterward incubated for around 16 to 24 hours. If the broth looks cloudy, the bacteria is growing. On the other hand, if the broth stays clear, the bacterial growth has been stopped by the antibiotic (Gajic et al., 2022; Tankeshwar, 2013). Figure 2 depicts the test tubes with different results after the incubation period (Themes, 2017).

Figure 2. Broth dilution



Agar dilution is another method that uses the minimum inhibitory concentration (MIC). In comparison to the broth dilution, the agar dilution uses a solid medium instead of liquid one. Agar is a substance used to grow bacteria. This substance has a liquid form when its warm and becomes solid after cooling down. In the agar dilution method, different

concentrations of antibiotic are added to the liquid agar. Afterwards the agar is poured into the petri dishes, glass or plastic containers used in microbiology laboratories, where it becomes solid. Subsequently small drops of bacteria are added on the top of the agar and left to incubate. After incubation, the agar plates are examined, and MIC is recorded. (Gajic et al., 2022)

3.3 Interpretation Standards

The interpretation of the results from Antimicrobial Susceptibility Testing involves classifying a microorganism as susceptible or resistant to various antimicrobial drugs. The interpretation is done according to established guidelines and breakpoints. These guidelines are provided by different organizations including the Clinical and Laboratory Standard Institute (CLSI) or the European Committee on Antimicrobial Susceptibility Testing (EUCAST). The main categories for interpreting the results are susceptible (S), intermediate (I) or resistant (R). (Medical Lab Notes, 2023)

When the microorganism is inhibited by the antimicrobial drug at a concentration that is achievable in the body, it suggests a high likelihood of effective treatment and the microorganism is classified as susceptible. The microorganism is classified as intermediate when the response to antimicrobial drug is uncertain and indicates limited effectiveness. Lastly, when the microorganism is not inhibited at the concentrations that are possible in the body, the treatment is likely to fail. (Medical Lab Notes, 2023)

There are some cases where additional categories such as non-susceptible (N) or full resistance (F) are used. However, these categories are only used in specific situations, require expert consultation and are not universally applicable. It is also important to point out that the interpretation categories depend on the specific organism and antimicrobial. In some cases, they may vary and require consultation. (Medical Lab Notes, 2023)

The results of AST must be reported in a clear and concise form. Information regarding the microorganism´s name, the used antimicrobial and the interpretation category must be included in the report. This information is important for healthcare professionals and helps them select the correct antimicrobial therapy for the treatment of different infections. (Medical Lab Notes, 2023)

4 Materials and Methods

This chapter aims to provide description of materials, tools and methods used during this study to detect and measure the inhibition zones for antimicrobial susceptibility testing. The chapter introduces the research design and its objectives along with the detailed sections covering the data preparation, data processing and methodology implementation.

4.1 Research Methodology

This study follows an applied, quantitative and experimental research methodology. The main purpose of this research it to test and evaluate the effectiveness of object detection models in identifying and measuring of inhibition zones.

The approach involves preparation and annotation of the dataset depicting antimicrobial susceptibility data, which is later used for training of different deep learning models.

The YOLO model was selected for the object detection task as it is one of the most competitive object detection models, famous for its speed and accuracy results. To provide further performance analysis the YOLO model is compared to other models such as SSD and Faster R-CNN which are also popular and commonly used models.

The object detection task is followed by the measurement of inhibition zones which are captured by the bounding boxes. The goal is to detect and measure the inhibition zones as accurately as possible to provide an automated solution for addressing challenges connected with antimicrobial susceptibility testing.

The overall workflow consists of data preparation, data annotation, model training, measurement of inhibition zones and evaluation. The implementation of these tasks was guided by a review of relevant and trusted literature in the fields of computer vision. The official documentation was used for the implementation of the selected models, frameworks and libraries.

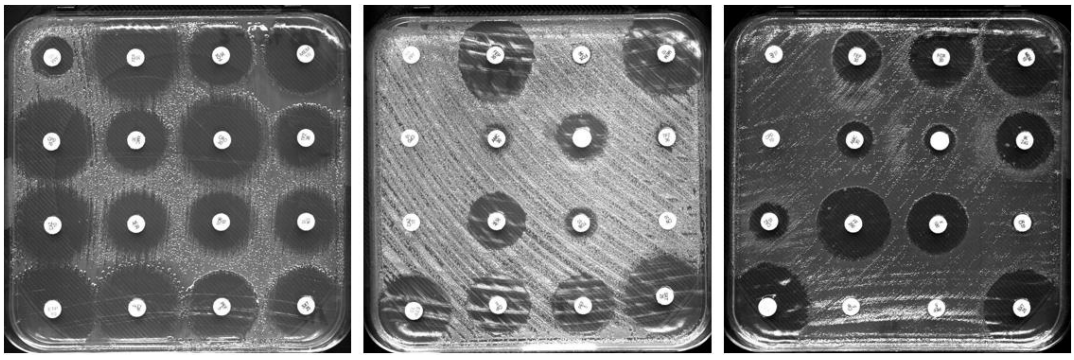
4.2 Data Collection

The dataset used in this study consists of 100 grayscale images obtained from a publicly available Dryad Digital Repository (Giske et al., 2024), which includes antimicrobial susceptibility data. This repository was created to support and enhance the accuracy in identifying antimicrobial resistance in Gram-negative bacteria, particularly focusing on Klebsiella pneumoniae strains.

The dataset consists of images and measurements from routine microbiological diagnostics using disk diffusion method in accordance with EUCAST guidelines to access the antimicrobial resistance in bacterial isolates. The bacterial strains were obtained from patient samples which were used at the Institute of Medical Microbiology, University of Zurich for diagnostic purposes. The data collection process took place from January 2020 to December 2022 and included the total of 225 Gram-negative bacterial isolates, representing 862 unique phenotypic categories.

The images of disk diffusion plates were taken after the incubation period, which took from 16 to 20 hours, and saved in JPEG format with a resolution of 1024x1024 pixels. Each image depicts 16 antibiotic disks placed in a 4x4 layout. Depending on the bacterial response, the inhibition zones of different sizes are visible around most of the disks. The images in the dataset are of different quality, lightning and contrast. Example images with different lightning conditions and quality can be seen in Figure 3.

Figure 3. Example images depicting white antibiotic disks with dark inhibition zones



In addition to the image folder, the dataset includes CSV files for each image. All CSV files have a clear and consistent filenames which match the corresponding images. These CSV files provide the inhibition measurement data in millimetres for each antibiotic tested against each bacterial strain.

4.3 Data Annotation

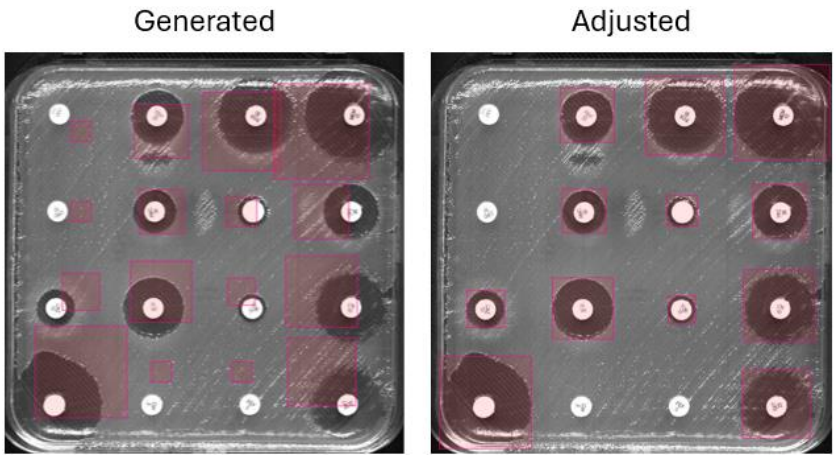
Accurate data annotation is one of the most important aspects in object detection tasks. To achieve this, the process involved a combination of disk diffusion images and the corresponding antimicrobial susceptibility data included in the CSV files. Each image was paired with a word table containing the inhibition zone diameters measured in millimetres. The values from the tables were extracted using a custom Python script to automate the annotation process.

After the extraction process the inhibition diameters were converted from millimetres to pixels using a scale factor which was calculated based on the antibiotic disks in the images, which are known to be 6 millimetres in actual size (Hudzicki, 2009). To correctly match the order of antibiotics in the tables and their actual visual arrangement in the images, multiple predefined mappings had to be created within the script.

Afterwards, the bounding boxes were generated in YOLO format for a consistent 4x4 disk layout across 1024x1024 pixel images. As a result of an accurate mapping technique, the generated bounding boxes were placed at the correct inhibition zone location.

These auto-generated annotations were saved in text files and subsequently imported in Computer Vision Annotation Tool (CVAT). To ensure a perfect bounding box position and alignment with the inhibition zones, the generated annotations were manually adjusted as shown in Figure 4. The bounding boxes were removed for resistant antibiotics which produced no inhibition zone. The total of 1295 labels were used.

Figure 4. Generated vs manually adjusted annotations for inhibition zones



This automated approach for image annotation ensured accurate creation of bounding boxes for inhibition zones based on the actual real-world measurements. By generating the annotations, the process was significantly accelerated and by manually adjusting the boxes, a correct alignment with the inhibition zones was achieved.

4.4 Technical Setup

The implementation process was conducted using the local machine and a cloud-based environment to optimize the resource usage. Preprocessing tasks such as annotation generation and dataset split scripts were run locally using PyCharm 2024.2.1 as the integrated development environment (IDE). These tasks were easily performed using solely the CPU resources.

The deep learning model training had to be conducted using Google Collab environment due to the computational demands. This cloud-based platform offers an access to Tesla T4 GPU with 16 GB of VRAM, which is sufficient for training and testing of different deep learning models. The use of the GPU allowed for faster and more efficient model training and helped avoid the limitations of CPU-based training.

The model implementation was done using PyTorch framework (PyTorch, 2024) which is compatible for many deep learning tasks. Libraries such as OpenCV, Matplotlib and Pillow were used for image handling and visual inspection, as well as NumPy for numerical operations.

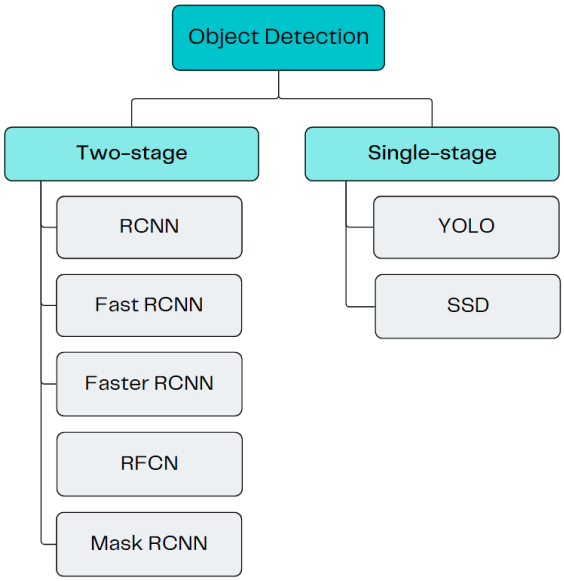
To visually inspect the training and testing results, TensorBoard was employed. Specifically, TensorBoard was used to monitor and analyse details of the training process including a training and validation loss. During testing, the Tensorboard was used to display evaluation metrics such as recall, precision and mAP along with the output images with predictions.

4.5 Object Detection

Object detection can be described as a computer vision task that locates objects, within an image or a video. In comparison to image classification, in object detection the model does both, the classification and object localization by drawing bounding boxes around the objects. The bounding boxes are typically defined by four coordinates that specify the

position and size of the bounding box (Park & Park, 2023). Generally, the object detection algorithms can be divided into two groups as seen from Figure 5 (Kundu, 2023) . This division is based on the number of times the input image is processed by the network.

Figure 5. One and two stage detectors



During this study, different object detection models were tested and evaluated. To make the dataset ready for the model implementation, it was split into training, validation and testing sets. For each of the models, 70% of the data was used for training, 15% for validation and 15% for testing. The following chapters provide a brief description of the tested models and their implementation process.

4.5.1 YOLO

YOLO (You Only Look Once) is a real-time and single-stage detection model. As a single-stage detection model, YOLO processes the input image only once, which is known as a single pass, and simultaneously predicts bounding boxes and class labels (Jani et al., 2023). This feature makes YOLO models significantly faster than two-stage detectors, while producing high accuracy results. Factors such as speed, accuracy, good generalization and it´s open-source feature make this model highly competitive. (Du, 2018)

The architecture of YOLO consists of three foundation parts. These are backbone, neck and head. The backbone is typically a Convolutional Neural Network (CNN), a specialized

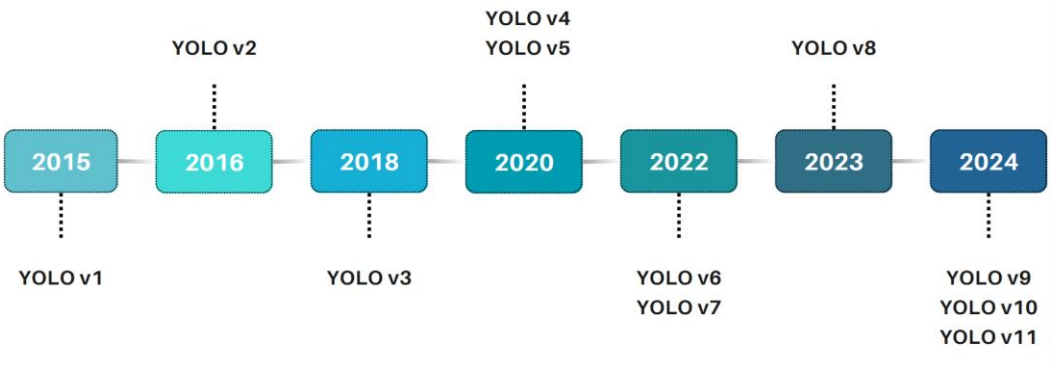
type of deep learning algorithm for object detection and classification tasks, which deals with the feature extraction from input data. (Ali & Zhang, 2024)

The extraction of features consists of applying filters called kernels that analyse the image and look for patterns (Alzubaidi et al., 2021). These learned features such as edges, textures and patterns are called feature maps. The neck component is responsible for further processing of the generated feature maps (Ali & Zhang, 2024).

Lastly, the head of the YOLO is focusing on the prediction of bounding boxes. For effective detection the head typically uses multi-scale anchor boxes, which are used in the early stage of object detection to define possible regions where objects may be localized. (Ali & Zhang, 2024)

The architecture of YOLO experienced continuous evolution since the first version release in 2015 (Ali & Zhang, 2024). Figure 6 shows a timeline of YOLOvX models since 2015 till 2024 (Boesch, 2024). The models starting from the YOLOv5 series offer different size variants such as nano (n), small (s), medium (m), large (l), extra-large (x) (Jani et al., 2023).

Figure 6. Timeline of YOLOvX models



During this study, different model versions such as YOLOv8, YOLOv9, YOLOv10 and YOLOv11 were implemented. The annotations were exported from CVAT in YOLO format and saved with the images in a zip file. A custom data.yaml file was created which specified the paths to the training, validation and testing sets. A single label inhibition_zone was specified in this file as well.

The YOLO models were trained using Ultralytics YOLO library (Ultralytics, 2024) which had to be installed first. The model was trained using the AdamW optimizer with an initial

learning rate of 0.002. The optimizer controls and adjusts the model´s weights and the learning rate is an important hyperparameter that determines the rate at which the model is learning (Agarwal, 2023). The number of samples a model learns from at a time, was specified with a batch size of 16.

Different augmentation techniques such as mosaic, random erasing and horizontal flip were used. The augmentation is used to form new and different data by slightly modifying the input images to further improve the model´s learning and accuracy (Mumuni & Mumuni, 2022). During the training, different model sizes such as nano/tiny, small and medium were used.

4.5.2 Single Shot Detector (SSD)

Another popular object detection model is Single Shot Detector referred to as SSD. The features of this model share similar characteristics as YOLO models, since both use single pass to process images or videos. The speed and efficiency of this model make it suitable for real-time object detection tasks. (Helmy, 2023)

The architecture typically consists of a base network such as ResNet or VGG which extracts the feature maps. After the feature map extraction, the base network is followed by multiple layers called extra layers which are designed to manage the detection of objects at different scales. (Helmy, 2023)

Generally, the SSD models can struggle with detection of objects at different scales (Helmy, 2023). This study aimed to implement and evaluate the performance of SSD models on a custom dataset consisting of objects with different sizes.

During this study SSD300 with VGG-16 backbone was implemented using PyTorch framework. The annotations were exported in COCO format since this model does not accept the previously used YOLO annotations.

The classification head was modified to match the number of classes in the dataset and a custom CocoDataset class was created to handle the image and annotation loading. The images were resized to 300x300 pixels and the bounding box coordinates were transformed accordingly.

During the training, a combination of classification loss, which reports an object class prediction, and bounding box regression loss, which reports an object localization, was used (Bounding Box Regression Loss, n.d.). The model was trained using a batch size of 8 and initial learning rate of 0.001. The optimizer that was used is Stochastic Gradient Descent (SGD). The SGD is an optimization algorithm that is responsible for updating weights using a smaller batch of data instead of the whole dataset. This makes the training much faster and memory efficient (ML | Stochastic Gradient Descent (SGD), 2025).

To help the model learn better, a MultiStepLR scheduler, which reduces the learning rate at specific epochs, was used (MultiStepLR — PyTorch 2.6 Documentation, n.d.). After the training, the best model was saved and subsequently tested on the test set. The results were printed and logged using TensorBoard for better visualization.

4.5.3 Faster R-CNN

Faster R-CNN is a two-stage object detection model with an improved R-CNN architecture. The architecture of R-CNN family consists of region proposal networks to find locations within an image where the objects possibly exist. Another phase of this architecture is convolutional neural network (CNN) to extract the features, and classification to predict the object class. Furthermore, regression is included in this architecture to fine-tune the bounding box coordinates. (Faster R-CNN | ML, 2023)

The main purpose of the Faster R-CNN is to detect objects within an image with precise localization. The architecture of Faster R-CNN consists of region proposal network, which is responsible for predicting regions of interest where the objects might be possibly localized, and Faster R-CNN detector which is responsible for the object detection within the suggested region proposal. (Faster R-CNN | ML, 2023)

To train the Faster R-CNN model, the annotations were exported in COCO format and uploaded in Google Collab. Afterwards, the Faster R-CNN model was implemented with a ResNet-50 backbone for feature extraction and Feature Pyramid Network to further improve the backbone. The Feature Pyramid Network is a widely used component of object detection models which enhances the ability of the model to detect objects of different sizes. (Lin et al., 2017)

During the model implementation the torchivsion library was used. To meet the number of classes in our dataset, the classification head was modified before the training. The model was trained with a learning rate of 0.005 and Stochastic Gradient Descent optimizer.

In addition to the ResNet-50 backbone, a ResNet-101 backbone with combination of Feature Pyramid Network was tested. This backbone provides deeper architecture with 101 layers instead of 50 that are used in ResNet-50 (Ma'rifah et al., 2023).

4.6 Measurement of Inhibition Zones

After training the models the measurement logic was added within a testing phase. The models were evaluated at the original image resolution of 1024x1024 pixels. The predicted bounding boxes from the models served as a base for the measurement of inhibition zones. After the bounding boxes were predicted by the model, their width was used for the measurement of inhibition zones. The width of the bounding box was converted into millimetres using a known scale factor of 6 mm per 52 pixels as shown in Equation 1.

Equation 1: Conversion of pixel width into millimetres

$$Diameter_{mm} = Width_{px} \times \frac{6}{52}$$

The converted values in millimetres were afterwards displayed as a label on top of each bounding box. To make sure the results were correctly displayed, TensorBoard was used for visualization of the images with predicted and measured bounding boxes. The results of the testing were subsequently saved in text files. These files included the predicted bounding box coordinates which were used for further analysis and evaluation of the model.

4.7 Evaluation Metrics

To analyse the models´ performance in detecting inhibition zones from images of disk diffusion plates, several performance metrics were taken into consideration. The most important metrics were precision, recall and mean Average Precision (mAP). In object detection, the calculation of these metrics is done using the values of true positives (TP), false positives (FP) and false negatives (FN) which are explained in the Table 1 below.

Table 1. Confusion matrix terminology (Koech, 2020)

Term	Definition
True Positive (TP)	correct prediction by the model that matches the annotation (ground truth)
False Positive (FP)	incorrect prediction by the model that does not match the ground truth
False Negative (FN)	missed ground truth bounding box

The precision represents the amount of true positive predictions among all positive predictions from the model and can be calculated using following Equation 2 (Koech, 2020):

Equation 2. Precision calculation

$$P = \frac{TP}{(TP + FP)}$$

Another metric that was used during the model evaluation is recall. Recall can be described as the model´s ability to detect all ground truths and is calculated using the following Equation 3 (Koech, 2020):

Equation 3. Recall calculation

$$R = \frac{TP}{(TP + FN)}$$

Lastly, mean average precision was used to determine the average precision over all classes (Koech, 2020). Even though our dataset consists of a single class, the mAP remains a valid performance metric which measures the ability to accurately localize and detect objects across different confidence thresholds.

The mAP can be calculated at different Intersection over Union (IoU). The IoU is a metric which calculates the degree of overlap between the ground truth and prediction (Koech, 2020). A common IoU used for mAP is 0.5, which means the bounding box is considered a true positive if the overlap with the ground truth is at least 50%.

5 Experiments and Results

This chapter provides an overview of the experiments and results that were conducted for this study to address the research questions. The primary focus was to find the best model for detection of inhibition zones from images depicting a disk diffusion plate and measure the diameter of the predicted bounding boxes in millimetres.

5.1 Object Detection Performance

During this study, three different object detection models, YOLO, SSD and Faster R-CNN, were tested and evaluated on our dataset. Experiments with different versions and architectures of these models were done for better comparison and evaluation purposes.

The YOLO models demonstrated impressive results in detecting the inhibition zones with minimal error. During the experimentation, different model versions and size variants with the default 640x640 pixels image resolution were compared. Table 2 provides an overview of YOLO models evaluated during this study and their testing results.

Table 2. YOLO object detection results in %

Model	Input size	Precision	Recall	mAP(0.5)	mAP(0.5:0.95)
YOLOv8n	640	99.2	98.9	99.5	82.6
YOLOv8s	640	100	98.8	99.5	85.6
YOLOv8m	640	98.3	**100**	99.5	**85.3**
YOLOv9t	640	98.8	98.3	98.5	83.2
YOLOv9s	640	100	**99.4**	99.5	**85.2**
YOLOv9m	640	99.4	99.4	99.5	83
YOLOv10n	640	96	93.3	98.8	81.6
YOLOv10s	640	99.4	98	99.4	83.7
YOLOv10m	640	98.7	97.8	99.3	86
YOLOv11n	640	100	98.3	99.5	84.1
YOLOv11s	640	98.7	99.4	99.5	83.9
YOLOv11m	**640**	100	**99.4**	99.5	**85.2**

The tested YOLO models achieved strong performance in terms of precision, recall and mAP. In this study, the most important metric is recall, to ensure all relevant objects are detected, and mAP(0.5:0.95) to strictly evaluate the localization accuracy. The best models were selected based on achieving at least 99% recall and minimum mAP(0.5:0.95) of 84%. The YOLOv8m, YOLOv9s and YOLOv11m were the best performing models meting this condition. The highest recall of 100% as well as the highest mAP(0.5:0.95) was achieved with the YOLOv8m.

Even though the YOLO models demonstrated very strong performance the study aimed to test and evaluate additional models such as SSD and Faster R-CNN for a comparison analysis.

The SSD model was trained with an input image resolution of 300x300 pixels. The model successfully detected most of the large inhibition zones, but it struggled with the medium and small-sized objects. Since the low input resolution can make it challenging for the model to capture enough details, the study experimented with higher resolution of 512x512 pixels.

While SSD300 is not originally designed for 512x512 resolution, this experiment served as an alternative of SSD515 which is unavailable in torchvision. The results of the SSD VGG-16 models can be seen in Table 3, where AP stands for average precision and AR (100) for average recall at 100 detections per image.

Table 3. Results of SSD300-VGG16 model in %

Metric	**SSD300 (300x300)**	**SSD300 (512x512)**
mAP(0.5:0.95)	73.1	74.4
mAP(0.5)	94	94
AP (large objects)	76	77.4
AP (medium objects)	5.1	5
AR(100)	77.3	78.2

The higher resolution slightly improved the average recall and mean average precision but did not improve the detection of medium objects.

The Faster R-CNN model achieved better results in comparison to SSD. During the study Faster R-CNN was evaluated using both, ResNet50 and ResNet101 backbone. These models performed better in detecting medium-sized objects but still resulted in low scores. The final testing results of the Faster R-CNN models can be seen in Table 4.

Table 4. Performance results of Faster R-CNN models

Metric	Faster R-CNN ResNet50	Faster R-CNN ResNet101
mAP(0.5:0.95)	80.3	80
mAP(0.5)	99.9	98.9
AP (large objects)	82	81.2
AP (medium objects)	68	69.1
AR(100)	84.8	84.7

From the table we can conclude that the difference between the results of these models was minimal. The Faster R-CNN Resnet101 slightly performed better in detecting medium objects and the Faster R-CNN ResNet50 in detecting large objects. Both models achieved high mean average precision of approximately 99% at the IoU of 50%. The average recall was around 85% which suggest the model successfully detected most of the relevant objects.

After comparison of all three models tested, we can undoubtedly conclude, that the YOLO models achieved the highest results in both mean average precision and recall. The second best- performing model was Faster R-CNN and the worst results were seen in SSD300 model.

5.2 Results of Inhibition Zone Measurement

The selected models for measurement comparison were YOLOv8m, YOLOv9s and YOLOv11m since they achieved the best recall and mean average precision. To evaluate the accuracy of inhibition zone measurement, the predicted bounding boxes were compared to the ground truth annotations. The millimetre values for the ground truth annotations and predicted boxes were rounded. The logic for matching the ground truth boxes with the predicted ones can be seen in Appendix 1.

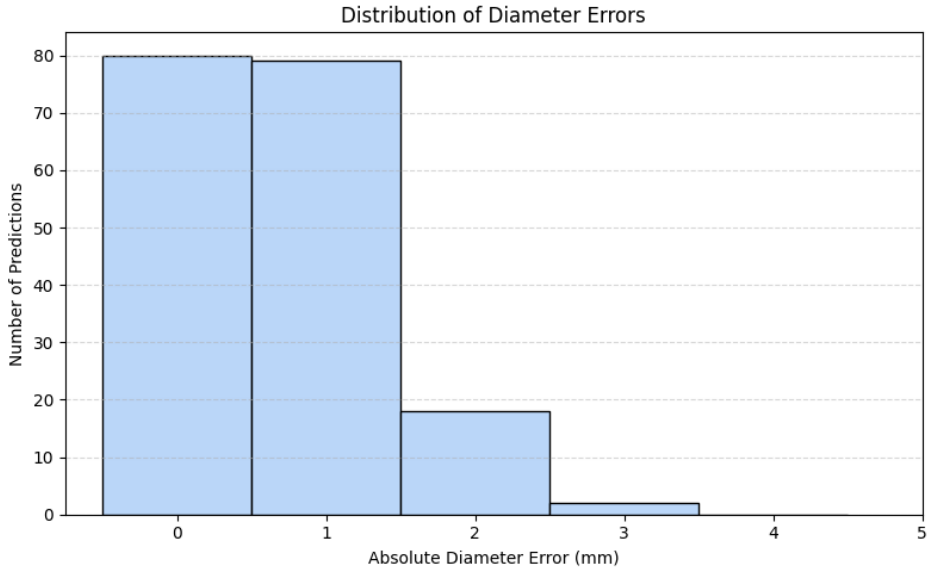
The deviation from the ground truth annotations was measured in millimetres. The evaluation was based on fixed error thresholds of 1 mm, 2 mm and 3 mm and is further explained in Appendix 2. The percentage of the predictions that fall within each range was calculated and is presented in Table 5.

Table 5. Percentage of predictions within defined millimetre error ranges

Model	1 mm error	2 mm error	3 mm error
YOLOv8m	88.8	98.9	100
YOLOv9s	86.4	97.2	100
YOLOv11m	87	95.5	99.4

Among the evaluated models, YOLOv8m achieved lowest mean absolute error of 0.68 millimetres and outperformed YOLOv9s and YOLOv11m which reached 0.72 and 0.73 error. From the Table 5 we can see that a high proportion of predictions fell within 1 mm and 2 mm ranges. The maximum difference between predictions and ground truth annotations was 3-millimetre error, with no prediction exceeding this value. The distribution of the YOLOv8m predictions falling within the defined ranges can be seen in Figure 7.

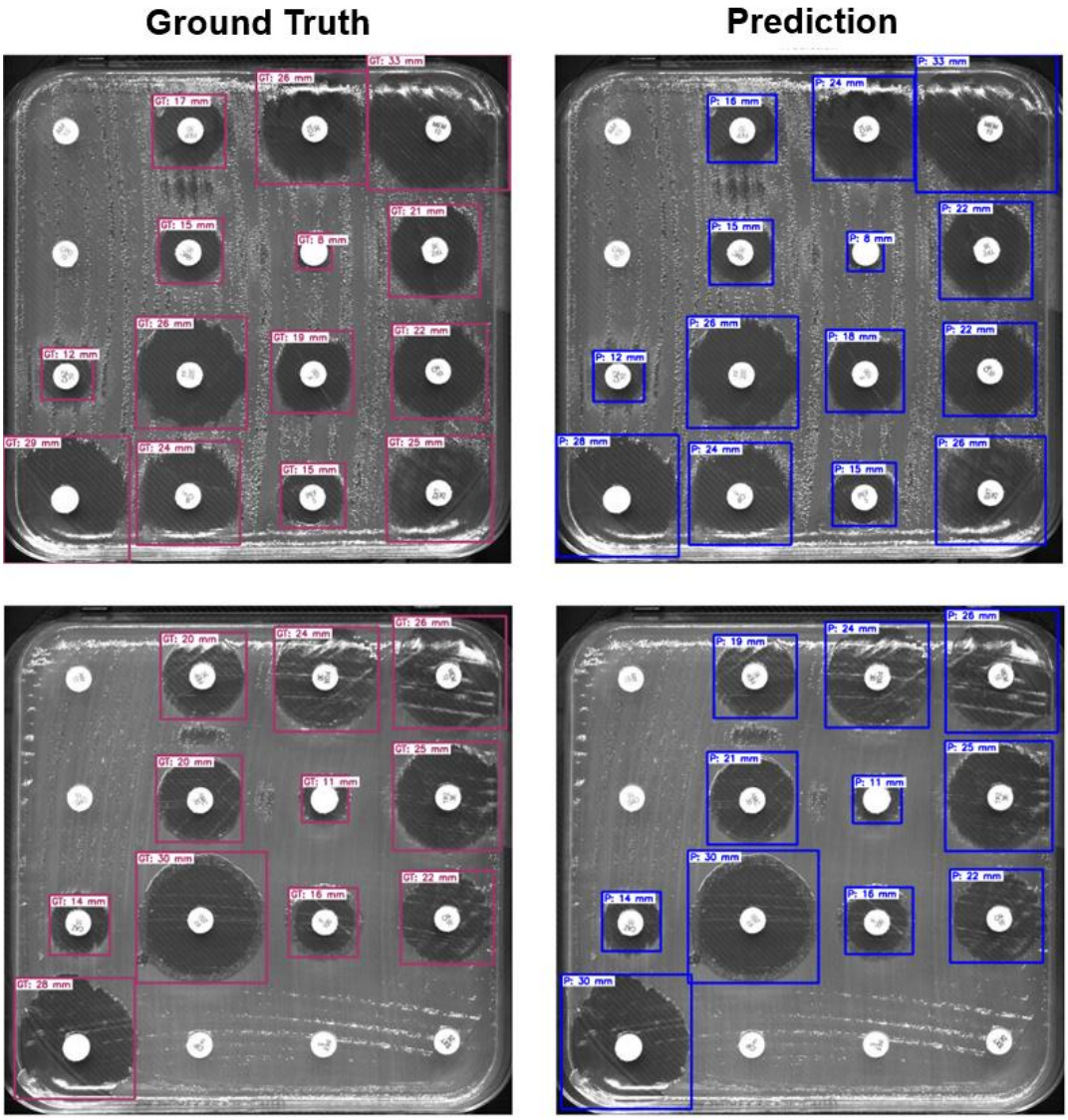
Figure 7. Distribution of YOLOv8m Diameter Errors



From the provided histogram it is seen that most of the predictions had 0- or 1-mm error. While some predictions had 3 mm difference from the ground truth annotations, the amount of such cases was minimal.

The visual comparison of ground truth annotations and the predictions of the YOLOv8m model on example images can be seen in Figure 8. The ground truth boxes are displayed in dark pink colour and the predicted bounding boxes are in blue colour. Each bounding box has a corresponding label displaying the width of the box in millimetres.

Figure 8. Visual comparison of ground truth and predicted bounding boxes



The provided figure proves that model successfully localized the relevant objects. As a result of effective training, the model learned to detect inhibition zones and ignore the antibiotic disks. From the labels displayed within the bounding boxes it is seen that the millimetre difference was minimal.

6 Conclusion

Antimicrobial resistance is a growing global health threat which requires accurate diagnostic methods such as antimicrobial susceptibility testing (AST). This study points to the importance of efficiency in AST and presents an automated solution for detection and measurement of inhibition zones by incorporating deep learning methods.

The detection of the inhibition zones was performed using different deep learning models such as YOLO, SSD and Faster R-CNN. These models were tested using different image resolutions and model versions. Among the tested models, YOLOv8m, YOLOv9s and YOLOv11m achieved highest scores in both, recall and mean average precision at IoU threshold from 0.5 to 0.95. The recall was one of the most important metrics the study took into consideration to make sure all relevant objects are detected. The mean average precision is another metric that was monitored to ensure as accurate predictions as possible. The highest recall of 100% was achieved with YOLOv8m as well as highest mAP(0.5:0.95) of 85.3%.

The measurement analysis was done on three best performing models for more detailed insights into their accuracy performance. The predicted bounding boxes were compared to the ground truth annotations and their difference in millimetres was calculated. The predictions were separated into 1 mm error, 2 mm error and 3 mm error deviation. The proportion of predictions falling within each of the groups was calculated for the selected models and subsequently compared. The best results were achieved with YOLOv8m which achieved the lowest mean absolute error of 0.68 mm. The proportion of predictions falling within 1 mm error was 88.8%, 98.9% for 2 mm error and 100% for 3 mm error. All the predictions falling within 3 mm error indicate that there were no predictions with more than 3 mm difference from the ground truth annotations.

These results demonstrate the model´s ability in detecting and measuring inhibition zones of different antibiotics against bacterial strains. The YOLOv8m proved its efficiency and achieved high recall, precision and accuracy scores. As a result of the precise annotation process, the model accurately localized the relevant objects with minimal difference from the ground truth boxes. These findings support a conclusion that deep learning models can achieve high detection and measurement results and have a potential for practical use in laboratory procedures such as antimicrobial susceptibility testing.

7 Future development

While the presented model demonstrated a strong performance in detection and measurement of inhibition zones, the future work is still necessary. Even though the model achieved high results in the object detection, it can serve as a base model for further experiments and improvements.

Firstly, to further evaluate the detection performance the dataset could be enlarged. While the detection results were high with our testing set, it is important to compare its performance with large variety of different images. This experiment would provide more valuable insights into the model performance and help the model generalize across different scenarios.

The measurement results provided a valuable information about the common millimetre deviation from the ground truth bounding boxes. The models achieved solid performance with a maximum of 3 mm error. While the model showed its ability to accurately localize the relevant objects, the results could be still improved. In the medical context, 1 mm and 2 mm error could be acceptable, but the 3 mm error might be a concern.

This type of error could easily lead to false classification of the bacteria tested against a specific antibiotic. Therefore, it is necessary to further improve the localization accuracy. This could be done by several methods including refining bounding box predictions, advanced loss functions such as CIoU, data quality and quantity or different augmentation techniques. These improvements could lead to more accurate and reliable results which are necessary in medical settings.

References

Abraham, E. P., & Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, *146*(3713), 837–837. https://doi.org/10.1038/146837a0

Agarwal, R. (2023, September 13). *Complete Guide to the Adam Optimization Algorithm*. Built In. https://builtin.com/machine-learning/adam-optimization

Ahmed, S. K., Hussein, S., Qurbani, K., Ibrahim, R. H., Fareeq, A., Mahmood, K. A., & Mohamed, M. G. (2024). Antimicrobial resistance: Impacts, challenges, and future prospects. *Journal of Medicine, Surgery, and Public Health*, *2*, 100081. https://doi.org/10.1016/j.glmedi.2024.100081

Ali, M. L., & Zhang, Z. (2024). The YOLO Framework: A Comprehensive Review of Evolution, Applications, and Benchmarks in Object Detection. *Computers*, *13*(12), Article 12. https://doi.org/10.3390/computers13120336

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. https://doi.org/10.1186/s40537-021-00444-8

American Chemical Society. (n.d.). *Alexander Fleming Discovery and Development of Penicillin—Landmark*. American Chemical Society. https://www.acs.org/education/whatischemistry/landmarks/flemingpenicillin.html

Antimicrobial resistance | EFSA. (2025, March 5). https://www.efsa.europa.eu/en/topics/topic/antimicrobial-resistance

Boesch, G. (2024, December 6). *YOLO Explained: From v1 to v11*. Viso.Ai. https://viso.ai/computer-vision/yolo-explained/

Bounding Box Regression Loss. (n.d.). CloudFactory Computer Vision Wiki. https://wiki.cloudfactory.com/docs/mp-wiki/loss/bounding-box-regression-loss

Breijyeh, Z., Jubeh, B., & Karaman, R. (2020). Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It. *Molecules*, *25*(6), 1340. https://doi.org/10.3390/molecules25061340

CDC. (2024a, April 12). *About Acinetobacter*. Acinetobacter</Em>. https://www.cdc.gov/acinetobacter/about/index.html

CDC. (2024b, May 16). *About Klebsiella*. Klebsiella</Em>. https://www.cdc.gov/klebsiella/about/index.html

CDC. (2024c, June 28). *About Pseudomonas aeruginosa*. Pseudomonas Aeruginosa</Em>. https://www.cdc.gov/pseudomonas-aeruginosa/about/index.html

CDC. (2024d, July 5). *About Escherichia coli Infection*. E. Coli</Em> Infection (Escherichia Coli</Em>). https://www.cdc.gov/ecoli/about/index.html

Du, J. (2018). Understanding of Object Detection Based on CNN Family and YOLO. *Journal of Physics: Conference Series*, *1004*(1), 012029. https://doi.org/10.1088/1742-6596/1004/1/012029

Endale, H., Mathewos, M., & Abdeta, D. (2023). Potential Causes of Spread of Antimicrobial Resistance and Preventive Measures in One Health Perspective-A Review. *Infection and Drug Resistance*, *16*, 7515–7545. https://doi.org/10.2147/IDR.S428837

European Antimicrobial Resistance Surveillance Network (EARS-Net). (2017, May 19). https://www.ecdc.europa.eu/en/about-us/networks/disease-networks-and-laboratory-networks/ears-net-data

Faster R-CNN | ML. (2023). GeeksforGeeks. https://www.geeksforgeeks.org/faster-r-cnn-ml/

Gajic, I., Kabic, J., Kekic, D., Jovicevic, M., Milenkovic, M., Mitic Culafic, D., Trudic, A., Ranin, L., & Opavski, N. (2022). Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods. *Antibiotics*, *11*(4), 427. https://doi.org/10.3390/antibiotics11040427

Giske, C. G., Bressan, M., Fiechter, F., Hinic, V., Mancini, S., Nolte, O., & Egli, A. (2024). *Image dataset of disk diffusion assay scanned with the SIRscan system* (Version 4, p. 357269120 bytes) [Dataset]. Dryad. https://doi.org/10.5061/DRYAD.5DV41NSFJ

Global Antimicrobial Resistance and Use Surveillance System (GLASS). (n.d.). https://www.who.int/initiatives/glass

Helmy, B. E.-D. (2023, February 9). *Single Shot Detectors (SSDs) | Baeldung on Computer Science*. https://www.baeldung.com/cs/ssd

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: Past, present and future. *Current Opinion in Microbiology*, *51*, 72–80. https://doi.org/10.1016/j.mib.2019.10.008

Hudzicki, J. (2009). *Kirby-Bauer disk diffusion susceptibility test protocol*. American Society for Microbiology. https://asm.org/getattachment/2594ce26-bd44-47f6-8287-0657aa9185ad/kirby-bauer-disk-diffusion-susceptibility-test-protocol-pdf.pdf

Jani, M., Fayyad, J., Al-Younes, Y., & Najjaran, H. (2023). *Model Compression Methods for YOLOv5: A Review* (No. arXiv:2307.11904). arXiv. https://doi.org/10.48550/arXiv.2307.11904

Kalvaitis, K. (2023). *Penicillin: An accidental discovery changed the course of medicine*. https://www.healio.com/news/endocrinology/20120325/penicillin-an-accidental-discovery-changed-the-course-of-medicine

Koech, K. E. (2020, August 26). *On Object Detection Metrics With Worked Example*. Towards Data Science. https://towardsdatascience.com/on-object-detection-metrics-with-worked-example-216f173ed31e/

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature Pyramid Networks for Object Detection* (No. arXiv:1612.03144). arXiv. https://doi.org/10.48550/arXiv.1612.03144

Medical Lab Notes. (2023, June 27). Antimicrobial Susceptibility Testing (AST): Introduction, Principle. *Medical Notes*. https://medicallabnotes.com/antimicrobial-susceptibility-testing-ast-introduction-principle-test-methods-test-requirements-procedure-and-result-interpretation-application-and-keynotes/

ML | Stochastic Gradient Descent (SGD). (2025). GeeksforGeeks. https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/

Moyo, P., Moyo, E., Mangoya, D., Mhango, M., Mashe, T., Imran, M., & Dzinamarira, T. (2023). Prevention of antimicrobial resistance in sub-Saharan Africa: What has worked? What still needs to be done? *Journal of Infection and Public Health*, *16*(4), 632–639. https://doi.org/10.1016/j.jiph.2023.02.020

MultiStepLR — PyTorch 2.6 documentation. (n.d.). https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.MultiStepLR.html

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Murray, P. R. (2015). 16—The Clinician and the Microbiology Laboratory. In J. E. Bennett, R. Dolin, & M. J. Blaser (Eds.), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition)* (pp. 191–223). W.B. Saunders. https://doi.org/10.1016/B978-1-4557-4801-3.00016-3

Park, H.-M., & Park, J.-H. (2023). YOLO Network with a Circular Bounding Box to Classify the Flowering Degree of Chrysanthemum. *AgriEngineering*, *5*(3), Article 3. https://doi.org/10.3390/agriengineering5030094

Poolman, J. T., & Wacker, M. (2016). Extraintestinal Pathogenic Escherichia coli, a Common Human Pathogen: Challenges for Vaccine Development and Progress in the Field. *The Journal of Infectious Diseases*, *213*(1), 6–13. https://doi.org/10.1093/infdis/jiv429

PyTorch. (2024). *PyTorch Documentation*. https://docs.pytorch.org/docs/stable/

Rivera, A., Viñado, B., Benito, N., Docobo-Pérez, F., Fernández-Cuenca, F., Fernández-Domínguez, J., Guinea, J., López-Navas, A., Moreno, M. Á., Larrosa, M. N., Oliver, A., & Navarro, F. (2023). Recommendations of the Spanish Antibiogram Committee (COESANT) for *in vitro* susceptibility testing of antimicrobial agents by disk diffusion. *Enfermedades Infecciosas y Microbiologia Clinica (English Ed.)*, *41*(9), 571–576. https://doi.org/10.1016/j.eimce.2022.12.009

Salam, M. A., Al-Amin, M. Y., Salam, M. T., Pawar, J. S., Akhter, N., Rabaan, A. A., & Alqumber, M. A. A. (2023). Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. *Healthcare*, *11*(13), Article 13. https://doi.org/10.3390/healthcare11131946

Sharma, R. (2022, May 18). *Kirby Bauer Disc Diffusion Method For Antibiotic Susceptibility Testing*. https://microbenotes.com/kirby-bauer-disc-diffusion/

Susceptibility Testing—Infectious Diseases. (n.d.). MSD Manual Professional Edition. https://www.msdmanuals.com/professional/infectious-diseases/laboratory-diagnosis-of-infectious-disease/susceptibility-testing

Tang, K. W. K., Millar, B. C., & Moore, J. E. (2023). Antimicrobial Resistance (AMR). *British Journal of Biomedical Science*, *80*, 11387. https://doi.org/10.3389/bjbs.2023.11387

Tankeshwar, A. (2013, November 15). *Broth Dilution Method for MIC Determination • Microbe Online*. Microbe Online. https://microbeonline.com/minimum-inhibitory-concentration-mic-broth-dilution-method-procedure-interpretation/

Themes, U. F. O. (2017, January 1). Principles of Antimicrobial Chemotherapy. *Basicmedical Key*. https://basicmedicalkey.com/principles-of-antimicrobial-chemotherapy-2/

Ultralytics. (2024). *Ultralytics Documentation*. https://docs.ultralytics.com/

Ventola, C. L. (2015). The Antibiotic Resistance Crisis. *Pharmacy and Therapeutics*, *40*(5), 344–352.

Wenzler, E., Maximos, M., Asempa, T. E., Biehle, L., Schuetz, A. N., & Hirsch, E. B. (2023). Antimicrobial susceptibility testing: An updated primer for clinicians in the era of antimicrobial resistance: Insights from the Society of Infectious Diseases Pharmacists. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *43*(4), 264–278. https://doi.org/10.1002/phar.2781

World Health Organization. (2023). *Antimicrobial resistance*. https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

Appendix 1. Distance-Based Logic for Matching Ground Truth Boxes with Predictions

The following code was used to match each ground truth box with the corresponding prediction. To ensure the correct prediction is matched with the ground truth annotation, the Euclidean distance between the center points of ground truth and predicted boxes is calculated. Subsequently, the nearest predicted box is matched with the ground truth box. The maximum distance allowed for matching a ground truth box with a predicted one was 20 pixels. Predicted boxes that had already been matched were excluded from further matching.

```
distance = cdist(gt_centers, pred_centers)
max_distance = 20  # maximum pixel distance to accept a match

for gt_idx, gt_center in enumerate(gt_centers):
    all_distances = distance[gt_idx]

    # Skip predictions that are already matched
    for pred_idx in matched_pred:
        all_distances[pred_idx] = np.inf

    # Find the closest predicted box
    best_idx = np.argmin(all_distances)
    best_distance = all_distances[best_idx]

    if best_distance <= max_distance:
        matched_gt.add(gt_idx)
        matched_pred.add(best_idx)
```

Appendix 2. Evaluation Logic for Diameter Accuracy

After the matching logic, the diameters of the ground truth and predicted boxes were calculated using the bounding box width and a know pixel-to-mm scale. The following code shows the comparison logic and categorization of diameter errors within 1 mm, 2 mm and 3 mm thresholds.

```
gt_d = round((gt_box[2] * img_gt.shape[1]) / px_per_mm)
pred_d = round((pred_box[2] * img_pred.shape[1]) / px_per_mm)
diff = abs(gt_d - pred_d)

diameter_diffs.append(diff)
gt_diameters.append(gt_d)
pred_diameters.append(pred_d)

total_pairs += 1
if diff <= 1:
    within_1mm += 1
if diff <= 2:
    within_2mm += 1
if diff <= 3:
    within_3mm += 1
```

Appendix 3. Data Management Plan

The research material used in this thesis consisted of a publicly available dataset providing antimicrobial resistance data, which was properly cited in the thesis. To keep a detailed documentation of the process of the thesis, a research diary was created and updated daily.

All materials including the dataset, annotations and thesis related files were stored on a password-protected local computer. The files were stored at the local C: drive and backed up to an external hard drive.

The analysis and results provided in the thesis were produced independently by the student. No third-party content was used without permission. In accordance with HAMK´s data management policy, the material will be securely stored for a period of one year.