תמונה שמכילה גופן, לוגו, סמל, גרפיקה

התיאור נוצר באופן אוטומטי

**Software Engineering Department**

**Braude College**

**Capstone Project Phase A – 61998**

**A CNN-Powered method to help in parasite bacterial identification**

**R-8-23-2**

**Cyrine Salame 319009064**

**Amit Shitrit 205725385**

**Supervisor: Zeev Frenkel**

**June 2023**

Table Of Contents

[Abstract 3](#_Toc136448605)

[1. Introduction 3](#_Toc136448606)

[2. Related Work 4](#_Toc136448607)

[3. Background- Microbiology and Bacteria 4](#_Toc136448608)

[3.1 Bacterial 5](#_Toc136448609)

[3.2 Five basic groups of bacteria 5](#_Toc136448610)

[3.3 Common methods for identifying bacteria 7](#_Toc136448611)

[3.3.1 Phenotypic methods 7](#_Toc136448612)

[Gram Staining 7](#_Toc136448613)

[Gram-positive and Gram-negative 7](#_Toc136448614)

[3.3.2 Immunological methods 7](#_Toc136448615)

[3.3.3 Molecular methods 7](#_Toc136448616)

[3.3.4 Mass spectrometry methods 8](#_Toc136448617)

[4. Background – Mathematical 8](#_Toc136448618)

[4.1 Artificial Intelligence 8](#_Toc136448619)

[4.2 Leaning Machine 9](#_Toc136448620)

[4.3 CNNs Model 9](#_Toc136448621)

[4.4 CNNs and Bacteria Images 9](#_Toc136448622)

[4.5 U-Net architecture 9](#_Toc136448623)

[Explanation of architecture 9](#_Toc136448624)

[5. Pseudocode 11](#_Toc136448625)

[5.1 definition 11](#_Toc136448626)

[5.2 Code Section 11](#_Toc136448627)

[Code Explanation 12](#_Toc136448628)

[6.Expected Achievements 12](#_Toc136448629)

[6.1 Outcomes 12](#_Toc136448630)

[6.2 Unique Feature 13](#_Toc136448631)

[6.2.1 Use of U-Net Algorithm 13](#_Toc136448632)

[6.2.2 Easy-to-use software 13](#_Toc136448633)

[6.2.3 Potential cost savings 13](#_Toc136448634)

[6.3 Criteria for Success 13](#_Toc136448635)

[7. Research process 14](#_Toc136448636)

[8. Software engineering documentation 15](#_Toc136448637)

[8.1 Use Case Diagram 15](#_Toc136448638)

[8.2 Activity Diagram 16](#_Toc136448639)

[8.3 System GUI 16](#_Toc136448640)

[References 20](#_Toc136448641)

Abstract

The ability to identify the classification of bacteria becomes a challenging task following the discovery of several thousand different types. In our project, we apply machine learning techniques, especially the Convolutional Neural Network (CNN) model.

As an attempt to improve the speed of prediction and the quality of accuracy in prediction, our system to be a prototype for the global system that in the future will be much more accurate and species with a wider variety of bacteria, our system will help medical staff in hospitals and professionals in the field of microbiology to save money and a lot of time our system. The U-Net architecture which is known for its high quality with microscope images and images with very tiny elements

1. Introduction

Bacteria are single-celled microorganisms that can be found in a wide range of environments, including soil, water, and even living organisms. They are some of the earliest and most primitive forms of life on Earth, and they play critical roles in many biological processes.

Bacteria are incredibly diverse, with an estimated 10 million different species. Some bacteria are beneficial to humans and other organisms, such as those that live in the gut and aid in digestion, while others can cause disease.

There are various laboratory tests that humans use to identify bacterial species based on their physical and chemical characteristics for example Gram staining, Biochemical tests, Serological tests, DNA-based tests, and Culture-based tests.

The most used test for bacterial identification is the Gram staining method. This is a simple and quick test that can differentiate bacteria into two main groups - Gram-positive and Gram-negative - based on the properties of their cell walls. The Gram staining method is widely used in clinical microbiology labs to quickly identify bacterial species and to guide the choice of antibiotic therapy.

Our project aims to revolutionize the way that bacterial infections are diagnosed and treated by providing healthcare professionals with a faster, more accurate, and more efficient diagnostic tool. By improving the Gram staining method using our software, we aim to contribute to the improvement of patient outcomes and the overall public health. Our software has the potential to greatly enhance the diagnostic capabilities of healthcare professionals and improve the speed and accuracy of bacterial identification.

Our project will leverage machine learning techniques, specifically the Convolutional Neural Network (CNN) model, to improve the Gram stain method for bacterial identification.

We will use the U-Net architecture, a widely used model for image segmentation tasks, to accurately identify and differentiate bacterial cells in stained samples. About By training the CNN model on a diverse dataset of Gram-positive and Gram-negative bacteria, we aim to improve the accuracy and speed of bacterial identification.

To facilitate the adoption and usability of our software, a user-friendly graphical user interface (GUI) has been developed that allows healthcare professionals Create an easy interaction and interpret the results.

2. Related Work

The accessibility of technology has resulted in the use of various applications that can help identify health issues at an early stage. Technological advancements such as Machine Learning and Computer Vision have allowed researchers to study the identification of parasites in microscopy images in new and innovative ways.

One example is Springer G-Turra[[1]](#footnote-1), who developed a deep learning-based approach for parasite detection and identification using a convolutional neural network (CNN) architecture. Another example is Rajaraman et al., who developed a deep learning-based approach for detecting and classifying malaria parasites in microscopy images.

The use of CNNs in identifying parasites can accurately classify various types of images, making the identification process faster and more efficient. CNN-based approaches eliminate the need for manual identification by trained personnel, which can be time-consuming and subjective.

Furthermore, CNN-based approaches can easily scale to accommodate large volumes of data, making them suitable for large-scale parasite identification projects[[2]](#footnote-2).

Based on these approaches, we decided to improve the detection technology by the U-Net algorithm by increasing the efficiency and predictive capabilities.

3. Background- Microbiology and Bacteria

This section is dedicated for the description of the history, causes, current situation, including current medical terms and tools for bacterial identification.

The first person to identify bacteria was Antoine van Leeuwenhoek, a Dutch scientist and microscopist, in the late 17th century. In 1676, he observed bacteria for the first time using a microscope of his own design, which had a magnification of up to 300 times. He discovered and described many microorganisms, including bacteria, and is the father of microbiology. Leeuwenhoek's observations helped to establish the field of microbiology and revolutionized our understanding of the microbial world.

Microbiology is the study of all living organisms that are too small to be visible with the naked eye. This includes bacteria, archaea, viruses, fungi, prions, protozoa, and algae, collectively known as 'microbes'. These microbes play key roles in nutrient cycling, biodegradation/biodeterioration, climate change, food spoilage, the cause and control of disease, and biotechnology. Thanks to their versatility, microbes can be put to work in many ways: making life-saving drugs, the manufacture of biofuels, cleaning up pollution, and producing or processing food and drink.

3.1 Bacterial

Bacteria are microbes with a cell structure simpler than that of many other organisms. Their control center, containing the genetic information, is contained in a single loop of DNA. Some bacteria have an extra circle of genetic material called a plasmid rather than a nucleus. The plasmid often contains genes that give the bacterium some advantage over other bacteria. For example, it may contain a gene that makes the bacterium resistant to a certain antibiotic. They can exist as single cells, in pairs, chains or clusters.

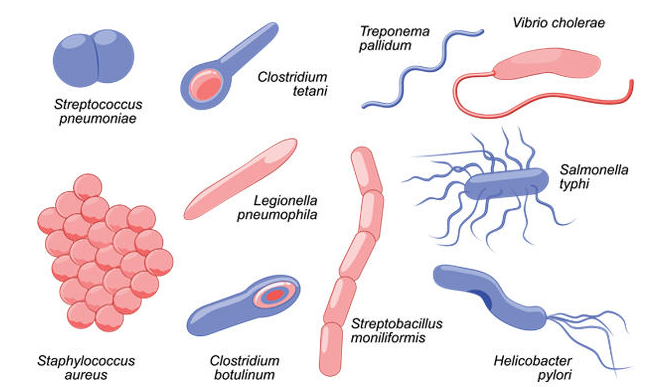
3.2 Five basic groups of bacteria

Bacteria are classified into five basic groups based on their morphology, or shape: 1.Cocci:These are spherical bacteria that can exist singly or in clusters. Examples include Streptococcus and Staphylococcus.

2.Bacilli: These are rod-shaped bacteria that can be found singly or in chains. Examples include Escherichia coli and Bacillus subtilis.

3.Spirilla: These are spiral-shaped bacteria that can be found singly or in chains. Examples include Treponema pallidum and Vibrio cholerae.

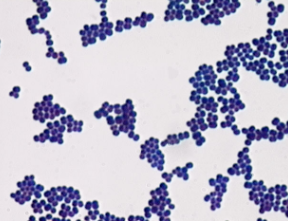
4.Vibrio's: These are comma-shaped bacteria that are typically found singly. Examples include Vibrio parahaemolyticus and Vibrio vulnificus.

****

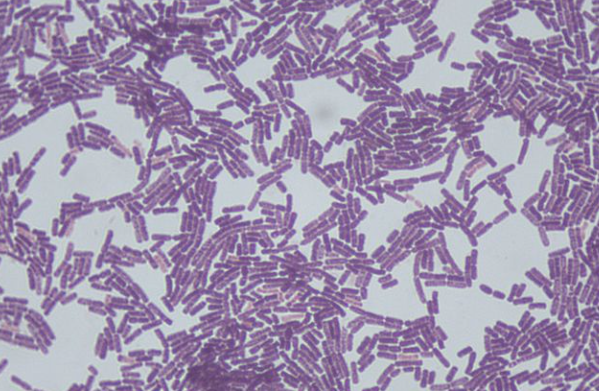
**Figure 1: A variety of common forms forms**

Represents a variety of forms of bacteria

**Figure 2: Gram Positive Cocci form**

****

**Figure 3: Gram Negative Bacilli form**

****

Represents a bacterium with Cocci form

Represents a bacterium with Bacilli form

3.3 Common methods for identifying bacteria

3.3.1 Phenotypic methods

These methods involve observing the physical and biochemical characteristics of bacterial cells. This can include things like the cell shape, size, motility, and growth characteristics on different types of media. These methods are relatively simple and inexpensive, but they can be time-consuming and may not always provide definitive identification.

Gram Staining

Gram staining is a phenotypic method that involves staining bacterial cells with crystal violet and iodine, followed by decolorization with alcohol and counterstaining with safranin. This method is used to distinguish between Gram-positive and Gram-negative bacteria based on differences in the composition of their cell walls. Gram-positive bacteria retain the crystal violet stain, appearing purple, while Gram-negative bacteria lose the stain and appear pink/red with the counterstain.

Gram-positive and Gram-negative

**Gram-positive** bacteria have a thick peptidoglycan layer in their cell wall, which retains the crystal violet stain during Gram staining, giving them a purple color. They also have a relatively simple cell wall structure and lack an outer membrane.

**Gram-negative** bacteria have a thinner peptidoglycan layer and an outer membrane composed of lipopolysaccharides (LPS) in their cell wall, which makes them more resistant to certain antibiotics and other environmental stresses. During Gram staining, router membrane is dissolved by alcohol, and the crystal violet stain is washed out, leading to their characteristic pink or red color when counterstained with safranin**.**

3.3.2 Immunological methods

These methods use antibodies to identify specific bacterial antigens. This can include techniques like enzyme-linked immunosorbent assay (ELISA) and Western blotting. These methods can be very specific and sensitive, but they require specialized equipment and can be expensive.

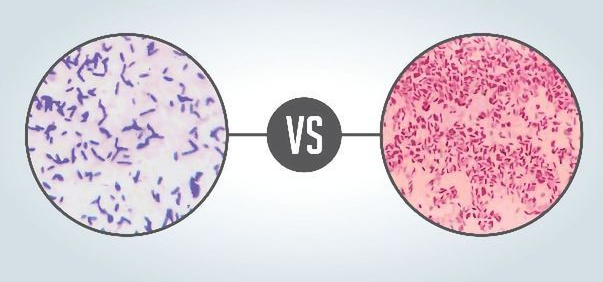
3.3.3 Molecular methods

These methods use DNA or RNA sequencing to identify bacteria at the genetic level. This can include techniques like polymerase chain reaction (PCR) and whole-genome sequencing. These methods can provide highly accurate and specific identification, but they require specialized equipment and can be expensive and time-consuming.

3.3.4 Mass spectrometry methods

This is a newer method that involves ionizing bacterial cells and analyzing the resulting spectra to identify unique biomolecules. This method can be very rapid and accurate, but it requires specialized equipment and expertise.

**Figure 4: Gram Positive and Gram Negative**

****

**תמונה שמכילה טקסט, צילום מסך, עיצוב, קנה מידה

התיאור נוצר באופן אוטומטי**

Left – Gram Positive Right -Gram Negative

**Figure 5: Gram Staining Process**

Represent 4 Stages in Gram Straining

4. Background ****– Mathematical****

This passage provides an overview of artificial intelligence, machine learning, convolutional neural networks, and U-Net architecture.

**4.1 Artificial Intelligence**

AI is the development of computer systems that perform tasks requiring human intelligence. It encompasses subfields like machine learning, deep learning, natural language processing, computer vision, robotics, and expert system. The objective of AI is to create adaptable systems that can learn from experience and perform tasks difficult or impossible for humans alone. Despite its potential in various fields like healthcare, finance, and transportation, AI poses concerns regarding job displacement, bias, privacy and security, and ethical considerations.

**4.2 Leaning Machine**

Machine learning is a type of AI that enables computers to learn and improve without explicit programming. It has three main categories: supervised, unsupervised, and reinforcement learning. Supervised learning uses labeled data to make predictions, while unsupervised learning finds patterns in unlabeled data. Reinforcement learning trains models to make decisions through trial-and-error feedback.

**4.3 CNNs Model**

Convolutional Neural Networks (CNNs) are a popular type of deep neural network used for image-related tasks such as image recognition, object detection, and image segmentation. They are designed to learn spatial hierarchies of features automatically and adaptively from input data. The network architecture of CNNs includes multiple layers, each with a specific purpose. The convolutional layers apply filters to extract features from the input image at different spatial locations. The pooling layers downsample the extracted features to reduce computational requirements. Finally, the fully connected layers use the extracted features to classify the image into different categories.

**4.4 CNNs and Bacteria Images**

Mhathesh, T. S. R. and his team used a CNN to classify bacterial images based on their species. The CNN was trained on a large dataset of labeled images and was able to automatically extract relevant features from the images without manual feature extraction. This approach has the potential to improve the accuracy and speed of bacterial identification and analysis Another approach is to use a CNN for the detection and segmentation of bacterial cells within images. This can be useful for tasks such as counting bacterial cells or identifying regions of interest within an image. Overall, CNNs have shown great potential for the analysis of bacterial images and have the potential to improve the accuracy and speed of bacterial identification and analysis.

4.5 U-Net architecture

The U-Net algorithm is a convolutional neural network (CNN) architecture that is commonly used for image segmentation tasks. It was proposed by Olaf Ronneberger, Philipp Fischer, and Thomas Brox in 2015.

Explanation of architecture

Step number one is Encoder:

The encoder network captures the contextual information of the input image using convolutional layers and pooling layers. It extracts high-level features and reduces the spatial resolution of the feature maps.

Step number two is Decoder:

The decoder network takes the encoded feature maps and performs up sampling operations to reconstruct the original image size. It uses transpose convolutions or up sampling followed by convolutions to increase the spatial resolution. The decoder also incorporates skip connections that allow it to access low-level features from the encoder, preserving fine-grained details.

Step number three is Output and Segmentation Mask:

The U-Net algorithm generates a segmentation mask that assigns a label to each pixel in the input image, representing the predicted boundaries or regions of interest. During training, it compares the predicted mask with the ground truth mask using loss functions like the dice coefficient or cross-entropy. The network parameters are optimized through gradient descent-based algorithms such as backpropagation, minimizing the loss and improving segmentation mask accuracy.

By iteratively updating the weights, the U-Net model learns to produce precise segmentation masks that closely align with the ground truth, enabling accurate pixel-wise segmentation for tasks like cell, organ, or tumor detection in biomedical image analysis.

תמונה שמכילה טקסט, צילום מסך, תרשים, קו

התיאור נוצר באופן אוטומטי

**Figure 7 – U-Net Architecture with Positive Gram image**

תמונה שמכילה טקסט, תרשים, צילום מסך, קו

התיאור נוצר באופן אוטומטי

5.Pseudocode

5.1 definition

Pseudocode is a way of writing out the steps of an algorithm or program using natural language and basic programming concepts, without necessarily adhering to the specific syntax or structure of a particular programming language.

5.2 Code Section

# Preprocessing the data

Load the dataset of gram staining images and their corresponding bacterial species labels Split the data into training, validation, and testing sets Preprocess the images by resizing, normalizing, and augmenting the data

#Designing the U-Net architecture

Define the U-Net architecture with encoder and decoder paths In the encoder path, define convolutional layers with pooling and batch normalization In the decoder path, define upsampling layers and concatenate with corresponding encoder path features Incorporate skip connections to preserve spatial information

# Training and validating the model

Set the hyperparameters for training, such as the learning rate and batch size Define loss function and optimizer Train the U-Net using the training set and hyperparameters Validate the U-Net's performance using the validation set Iterate over the previous steps until the model's performance is satisfactory

# Testing the model

Test the U-Net's performance on the testing set to evaluate its performance on new data Evaluate the model's performance metrics such as accuracy, precision, recall, and F1-score

Return results

# Save the model Save the trained U-Net model and its weights for future use

Code Explanation

Data section: Preparing the Gram Staining dataset by dividing it into training, validation, and testing sets, and applying techniques like resizing, normalization.

U-net Architecture section: Defining the U-Net model with an encoder path, decoder path, convolutional layers, pooling layers and skip connections to capture spatial information

Training Section: Setting hyperparameters, defining the loss function and optimizer, training the U-Net on the training set

Upload Input section : User upload the gram Staining images and click on diagnose

Result Section: Return Results and save the model for future use

6.Expected Achievements

6.1 Outcomes

Our project utilizes a CNN model to improve the accuracy, speed, and efficiency of bacterial identification using Gram staining. We collect a dataset of images of bacteria samples stained using the Gram method and train our model to learn the patterns and features associated with different bacterial species. With our software, healthcare professionals can input an image of a bacterial sample, and the CNN model processes the image to identify the species present, providing faster and more accurate diagnoses. Our project aims to revolutionize bacterial infection diagnosis and treatment by providing healthcare professionals with a powerful and effective tool for bacterial identification. Our software reduces the rate of false positives and false negatives, resulting in better patient outcomes, reduced healthcare costs, and less burden on medical laboratories. By improving the efficiency of bacterial identification, our project has a transformative impact on clinical microbiology, leading to more accurate diagnoses and more efficient healthcare systems.

6.2 Unique Feature

6.2.1 Use of U-Net Algorithm

Our project leverages the state-of-the-art U-Net model, which revolutionizes the identification of bacterial species by combining the power of convolutional neural networks (CNNs) with advanced segmentation techniques. By accurately segmenting bacterial species, it significantly reduces false positives and negatives, which will make us identify the bacteria in the fastest and most qualitative way.

6.2.2 Easy-to-use software

Our software is designed to be user-friendly and accessible to healthcare professionals with varying levels of technical expertise. This feature ensures that the benefits of our project are available to a wide range of users, helping to improve the overall quality of patient care.

6.2.3 Potential cost savings

Another unique feature of our project is its potential to reduce healthcare costs by improving the efficiency of bacterial identification. By reducing the time and resources required for bacterial identification. Overall, the unique features of our project make it a valuable and innovative contribution to the field of clinical microbiology, with the potential to greatly improve patient outcomes and public health.

6.3 Criteria for Success

1. Accuracy: Our software must accurately identify bacterial species to reduce false positives and negatives.  
2. Speed and efficiency of identifying bacteria using our software. It should process images quickly and enable healthcare professionals to make timely treatment decisions.  
3.User satisfaction: The software should be user-friendly and provide accurate results to help healthcare professionals make informed treatment decisions.  
4.Cost-effectiveness: The software must be cost-effective by improving the efficiency of bacterial identification, making it valuable for medical laboratories and healthcare systems. 5.Adaptability: The software should handle a wide range of bacterial species and be flexible enough to be used in various healthcare settings, from hospitals to clinics and beyond.

Our project aims to improve bacterial identification accuracy, speed, and efficiency while meeting healthcare professionals' needs and being cost-effective and adaptable to different healthcare settings.

7. Research process

Problem Definition: Today there are thousands of types of bacteria, and the identification process is becoming more difficult from moment to moment, in addition to this the professional and medical workforce is not able to cope with the work pressure, therefore a fast and efficient system is needed to help meet these needs.

Literature review: In this step, we will review relevant literature on bacterial infections, Gram staining, and diagnostic methods to identify current methods and techniques. This will give us a better understanding of the state-of-the-art in bacterial diagnosis and will help us to identify gaps in the literature that our research can fill.

Data collection: At this stage, we will collect and compile a large data set of bacterial samples that will be used for training and testing the software. This data will include different types of bacteria and may come from different sources, such as clinical samples or environmental samples, in addition to this we will cooperate with pharmaceutical companies and dedicated departments in hospitals to create the best up-to-date and high-quality data obtainable.

Data preprocessing: Once we have collected the data, we will preprocess it to ensure that it is in a format that can be used by the software. This may include cleaning the data, handling missing values, and normalizing the data.

Software development: At this stage we will open the software that improves the Gram staining method. We will use artificial intelligence on the CNN model and with the help of the U-Net algorithm we will be able to improve the accuracy and prediction quality of our system.

Software testing and evaluation: After developing the software, we will test it on the preprocessed data and evaluate its performance using appropriate metrics such as sensitivity, specificity, and accuracy.

Software improvement: Based on the results of the evaluation, we will iterate on the software and make improvements to increase its accuracy and robustness.

8. Software engineering documentation

8.1 Use Case Diagram

***תמונה שמכילה תרשים, טקסט, ציור, קו

התיאור נוצר באופן אוטומטי***Our usage diagram visually describes the user's interaction with our system, in addition to that you can see all the functionality of our system.

תמונה שמכילה טקסט, תרשים, קו, גופן

התיאור נוצר באופן אוטומטי8.2 Activity Diagram

**8.3 System GUI**

1. First screen is Sign up Page , this screen allows you to sign up to our system.

2. Second screen is Sign in Page ; this screen allows you to login into our system.

**תמונה שמכילה טקסט, צילום מסך, מולטימדיה, תוכנה

התיאור נוצר באופן אוטומטי**תמונה שמכילה טקסט, צילום מסך, תוכנה, מולטימדיה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, תוכנה, אתר

התיאור נוצר באופן אוטומטי3. After login to the system, the screen is Home Page, this screen has a variety of options to perform: read articles, update personal details, read the about us and the main function diagnose Bacteria images

תמונה שמכילה טקסט, צילום מסך, תוכנה, תכונות מולטימדיה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, תוכנה, תכונות מולטימדיה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, תוכנה, מערכת הפעלה

התיאור נוצר באופן אוטומטי

תמונה שמכילה טקסט, צילום מסך, תכונות מולטימדיה, תוכנה

התיאור נוצר באופן אוטומטי4.The main Screens is Diagnose page and Result Page, the user can upload up to three images of gram staining and click on diagnose screen.

תמונה שמכילה טקסט, צילום מסך, תוכנה, תכונות מולטימדיה

התיאור נוצר באופן אוטומטי5.Result Page: Show variety of results: Pie Diagram of bacteria type, name, gram positive or negative, Accuracy of identify and Treatment Recommendations.

References

Signoroni, A., Savardi, M., Pezzoni, M., Guerrini, F., Arrigoni, S., & Turra, G. (2018). Combining the use of CNN classification and strength‐driven compression for the robust identification of bacterial species on hyperspectral culture plate images. *IET Computer Vision*, *12*(7), 941-949.

Turra, G., Arrigoni, S., & Signoroni, A. (2017). CNN-based identification of hyperspectral bacterial signatures for digital microbiology. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part II 19* (pp. 500-510). Springer International Publishing.

Rajaraman, S., Antani, S. K., Poostchi, M., Silamut, K., Hossain, M. A., Maude, R. J., ... & Thoma, G. R. (2018). Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images. *PeerJ*, *6*, e4568.

Mhathesh, T. S. R., Andrew, J., Martin Sagayam, K., & Henesey, L. (2021). A 3D convolutional neural network for bacterial image classification. In *Intelligence in Big Data Technologies—Beyond the Hype: Proceedings of ICBDCC 2019* (pp. 419-431). Springer Singapore.

Siddique, N., Paheding, S., Elkin, C. P., & Devabhaktuni, V. (2021). U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, *9*, 82031-82057.‏‏

1. Springer-G-Turra:" CNN-based identification of hyperspectral bacterial signatures for digital" [↑](#footnote-ref-1)
2. Rajaraman et al., “Deep Learning for Parasite Detection and Classification in Microscopy Images.” [↑](#footnote-ref-2)