An Artificial Intelligence Vision Transformer Model for Classification of Bacterial Colony

Farah Memon[1], Bushra Naz[2], Sanam Narejo[2], Kalsoom Panhwar[3]
[1]Department of IICT, Mehran University of Engineering and Technology Jamshoro.
[2]Department of Computer System Engineering, Mehran University of Engineering and Technology Jamshoro.
[3]University of Sindh, Department of information and communication Technology Faculty of Engineering and Technology Jamshoro
*__Correspondence__: farah.memon36@yahoo.com

Citation | Memon. F, Naz. B, Narejo. S, Panhwar. K, "A Positional Self-Attention Transformer Model for the Bacterial Colony Classification", IJIST, Vol. 6 Issue. 1 pp 237-248, March 2024
Received | Jan 11, 2024, **Revised |** Feb 26, 2024, **Accepted |** Feb 28, 2024, **Published |** March 06, 2024.

The application of AI and machine learning, particularly the vision transformer method, in bacterial detection presents a promising solution to overcome limitations of traditional methods, offering faster and more accurate detection of disease-causing bacteria like E. coli and salmonella in water, crucial for human survival, with ongoing research to further assess its effectiveness in microbiology. This research introduces a revolutionary positional self-attention transformer model for the classification of bacterial colonies. Leveraging the proven success of transformer architectures in various domains, we enhanced the model's performance by integrating a positional self-attention mechanism. We presented a novel approach for bacterial colony classification utilizing a positional self-attention transformer model. This allows the model to effectively capture spatial relationships and patterns within bacterial colonies, contributing to highly accurate classification results. We trained the model on a substantial dataset of bacterial images, which ensures its robustness and generalization to diverse colony types. The proposed model adeptly captured the spatial relationships and sequential patterns inherent in bacterial colony images, allowing for more accurate and robust classification. The proposed model demonstrated remarkable performance, achieving an accuracy of 98.50% in the classification of bacterial colonies. This novel approach surpasses traditional methods by effectively capturing intricate spatial relationships within microbial structures, offering unprecedented accuracy in discerning subtle morphological variations. The model's adaptability to diverse colony shapes and arrangements marks a significant advancement, promising to redefine the landscape of bacterial colony classification through the lens of state-of-the-art deep learning techniques. The high classification accuracy attained by the model, suggests its potential for practical applications in the early diagnosis of infectious diseases and the development of targeted treatments. The findings of this study underscore the effectiveness of incorporating positional self-attention in transformer models for image-based classification tasks, particularly in the domain of bacterial colony analysis.

Keywords: Classification, Deep learning, E. coli, Salmonella, Vision Transformer.

Introduction:

Microorganisms play a vital role in human life. Therefore, microorganism detection is of great significance to human beings [1]. In recent years, there has been a growing interest in the application of Artificial Intelligence (AI) and machine learning methodologies for various tasks in the field of microbiology, including bacterial detection. Traditional methods for detecting bacteria in samples, such as culture-based techniques, can be time-consuming and may not be able to detect all types of bacteria present in a sample. Furthermore, manual detection methods may be prone to human error, which can lead to inaccurate results.

Humans most heavily rely on water for their survival [2]. Water is crucial for metabolic processes. Most common disease-causing bacteria found in water is the bacterium Escherichia coli (E. coli) and salmonella [3]. In this context, bacterial detection using the vision transformer method refers to the use of the transformer architecture for detecting bacterial cells in images. This approach involves training a vision transformer model on a dataset of bacterial images, which can then be used to classify new images as either containing or not containing bacterial cells [4]. The application of this method is still relatively new, and research is ongoing to determine its effectiveness in the field of microbiology.

Unhealthy water leads to many diseases related to gastrointestinal illness, reproductive problems, and neurological disorders. People think that if the water is clear, it might be clean, which is myth. Clean-looking water contains a number of impurities, contamination, and bacteria that cannot be seen by the naked eye and causes severe health issues. Hence, it is rational to apply advanced computational methods for image analysis technologies in the microorganism identification field. Microorganisms can be detected with excellent accuracy and efficiency using computer image analysis. Furthermore, these methods have the potential to decrease the likelihood of erroneous identification in cases of diagnostic uncertainty, such as misleading similarities in the morphology or structure of bacterial cells. The main contributions of this paper are as follows:

- Enhanced quality of bacterial images is achieved through the application of median noise filtering. This technique effectively removes and diminishes noise, particularly after the data augmentation process.
- leveraging a deep learning-based architecture, we propose the utilization of a vision transformer model. This model is designed to extract the most pertinent features from bacterial images, thereby enhancing the capability of the classification system [5].

To mitigate potential overfitting concerns in the image classification process, we integrate a pooling layer and a dropout mechanism. These strategies are implemented before the application of a SoftMax activation. Overfitting occurs when a machine learning model learns the training data too well, capturing noise and specific details that are not representative of the broader dataset [6]. The pooling layer reduces the spatial dimensions of the input data, diminishing the model's sensitivity to small variations. Additionally, the dropout mechanism randomly omits certain neurons during training, preventing the network from relying too heavily on specific features and promoting a more robust learning process. Overall, nowadays, the vision transformer-based method shows promising results as a useful tool for microbiologists. This method holds the promise of enhancing both the accuracy and speed of bacterial detection which could have a significant impact on various applications, such as clinical diagnosis, food safety, and environmental monitoring.

Literature Review:

In [7], the researcher proposed a novel method for detecting malignant melanoma using a combination of the bacterial colony optimization algorithm and Support Vector Machines (SVMs). The bacterial colony optimization algorithm was used to optimize the SVM parameters, while the SVM was used to classify the melanoma images as benign or malignant. The proposed

method was tested on a dataset of melanoma images, and the results showed that it outperformed other existing methods for melanoma detection.

In [5], the study provides an overview of object detection techniques that are used in microorganism image analysis. The authors present a comprehensive survey of classical methods such as thresholding, edge detection, and segmentation, as well as deep learning-based approaches like Faster R-CNN, YOLO, and Mask R-CNN [4]. The paper also discusses the challenges associated with microorganism image analysis, such as low contrast, uneven illumination, and complex backgrounds. The authors highlight the importance of accurate object detection for various applications in microbiology, including disease diagnosis, drug development, and environmental monitoring. The survey provides a valuable resource for researchers and practitioners in the field of microorganism image analysis.

In [8] a dataset of olive leaf images containing five types of diseases and healthy leaves is presented. The ViT and CNN models were trained on this dataset and evaluated based on their classification accuracy. The results showed that the ViT model outperformed the CNN model in terms of accuracy and F1 score [9]. The study also included a comprehensive analysis of the classification performance of both models for each disease type. The authors concluded that the ViT model is a promising approach for accurate and efficient olive disease classification, which can be useful for disease monitoring and early detection in olive farming.

[10] presents a study on the detection of E. Coli bacteria in drinking water using image processing techniques. The authors collected water samples from various sources and cultured them on agar plates to grow bacterial colonies. Then, they captured images of the agar plates using a digital camera and processed the images to detect the presence of E. Coli bacteria. The paper describes the image processing techniques used for segmentation, feature extraction, and classification of bacterial colonies [11]. The authors compared the performance of different classifiers, including K-nearest neighbors (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN), for the detection of E. Coli bacteria. The results showed that the SVM classifier achieved the highest accuracy in detecting E. Coli bacteria, with an overall accuracy of 96.25%. The study provides a useful approach for the detection of E. Coli bacteria in drinking water using image processing techniques.

[12] collected water samples from various sources and cultured on agar plates to grow bacterial colonies [13]. Then, they captured images of the agar plates using a digital camera and processed the images to detect the presence of E. Coli bacteria [14]. In [15] researchers describes the use of deep learning algorithms, specifically Convolutional Neural Networks (CNNs), for the detection of E. Coli bacteria in the images. The authors compared the performance of different CNN architectures, [16] incorporated VGG16, ResNet50, and InceptionV3, for the detection of E. Coli bacteria. The results showed that the InceptionV3 architecture achieved the highest accuracy in detecting E. Coli bacteria, with an overall accuracy of 93.17%. The study demonstrates the effectiveness of deep learning techniques in the detection of E. Coli bacteria in water, which can be useful for ensuring safe drinking water.

Objectives:

The study aims to utilize a state-of-the-art deep transformer-based architecture to design an accurate and efficient model for classifying bacterial colonies, specifically targeting E. coli and Salmonella. To evaluate the Performance of the model.

Novelty Statement:

In the realm of bacterial colony classification, our research stands out by introducing a cutting-edge positional self-attentional transformer model. This novel approach surpasses traditional methods by effectively capturing intricate spatial relationships within microbial structures, offering unprecedented accuracy in discerning subtle morphological variations. The model's adaptability to diverse colony shapes and arrangements marks a significant

advancement, promising to redefine the landscape of bacterial colony classification through the lens of state-of-the-art deep learning techniques. Figure 1 provides the flow map of proposed model.

Methodology:

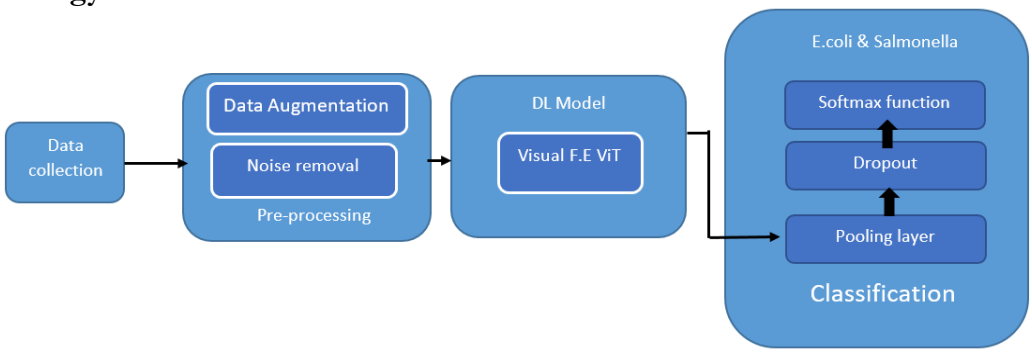


Figure 1. Proposed model for bacterial disease classification using Vision Transformer

Data Collection:

Dataset:

We have curated a labeled dataset of bacterial colony images, with each image annotated according to its classification, such as different bacterial species.[3]. This dataset serves as a comprehensive resource for training and evaluating deep learning models in the field of microbiological image analysis [17]. The images within the dataset vary in size, with dimensions ranging from [width x height]. Commonly, these images have been standardized to a resolution of [e.g., 224 x 224 pixels] to ensure consistency during the training process. The specific dimensions are chosen to balance computational efficiency with the preservation of critical spatial features for accurate classification [2].

The overall size of the dataset is a critical factor in model training. As an estimation, the dataset comprises the number of images, with each image associated with its corresponding label [15]. Given the diverse nature of bacterial colonies and species, the dataset covers a broad spectrum of microbial variations, enhancing the model's ability to generalize across different scenarios. The size of the dataset in terms of storage can be measured in megabytes (MB) or gigabytes (GB), depending on the resolution and quantity of images. This dataset, with its labeled bacterial colony images, facilitates the development and evaluation of deep-learning models tailored for microorganism identification and classification.

Table 1. Dataset Description

Class Name	Training Set	Testing Set	Validation Set
E. coli	9100	2600	1300
Salmonella	6650	1900	950

The Table 1 delineates the dataset composition for two bacterial classes, "E. coli" and "Salmonella," incorporating a breakdown based on a 70%, 20%, and 10% ratio for training, testing, and validation sets, respectively. In the case of "E. coli," the training set comprises 70% of the total data, consisting of 9100 samples. The testing set constitutes 20% with 2600 samples, and the validation set constitutes the remaining 10%, featuring 1300 samples. Similarly, for the "Salmonella" class, the training set encompasses 70% of the data, totaling 6650 samples, while the testing set comprises 20% with 1900 samples, and the validation set makes up the remaining 10%, encompassing 950 samples. These percentage-based ratios offer a nuanced perspective on the distribution of data across training, testing, and validation sets for each bacterial class. These percentage-based ratios provide insight into the distribution of data across training, testing, and validation sets for each bacterial class, ensuring a balanced approach during the machine learning model development process.

Data Augmentation:

At this stage of the process, the images were improved through the application of the median noise filtering method [11]. Data augmentation was employed for effective image categorization for several reasons. It aided in enhancing the diversity of the dataset, a crucial aspect for effectively training resilient deep learning models.

Noise Removal:

In this phase, we utilized median noise filtering to improve the quality of the images [18]. The median filter proved to be particularly effective in reducing or eliminating noise present in the collected photographs. This technique replaced a pixel with the median of the neighboring gray levels, offering a robust approach to noise reduction.

Feature Extraction:

Feature extraction involves transforming the input image into a sequence of fixed-size embedding.

Vision Transformer (ViT):

A Vision Transformer (ViT) is a type of neural network architecture that has gained prominence for its remarkable performance in image classification tasks, including the classification of bacterial colonies such as E. coli and Salmonella. The key distinction of ViTs lies in their departure from traditional (CNN) architectures, replacing convolutional layers with self-attention mechanisms inspired by transformer models originally designed for natural language processing tasks.

Classification:

The architecture of multilayer neural networks comprises of three layers: an input layer, one or more hidden layers sequentially connected to the input layer, and an output layer. [19][17] The first layer is always linked to external sources or multiple external components. Our primary focus at this stage was on the output layer to obtain optimal results. We amalgamated the features identified in the previous phase and employed them as inputs for our classifier, specifically utilizing a softmax layer. This approach enabled us to ascertain the most accurate answer.

Unlike the activation functions in the hidden layers, the activation function of the output layer is distinctive. Each layer's function varies, and so does its implementation. [20] In the context of a classification task, the last layer facilitates the creation of class probabilities for the input data, employing the softmax function:

$$\text{SoftMax } (\mathbf{x})i = \frac{e^{xi}}{\sum_j e^{xj}}$$

Here, **x** is considered a vector, and each element xi can assume any real-world value. To ensure that the output values sum to 1, a normalization term is employed, ensuring the validity of the probability distribution. Deep neural networks, recognized for their ability to efficiently and accurately train on extensive datasets with numerous parameters, have garnered significant attention in research papers. However, they often encounter challenges such as overfitting. Regularization is a common strategy to combat overfitting, and one such technique is the dropout function. The dropout function is advantageous as it allows the integration of various networks into a single architecture while preventing overfitting between units. [21] It is widely recognized that dropout performs effectively in fully connected and pooling layers. Figure 2 represents the architecture of vision transformer. While figure 3 shows images comprising of two types of bacteria.

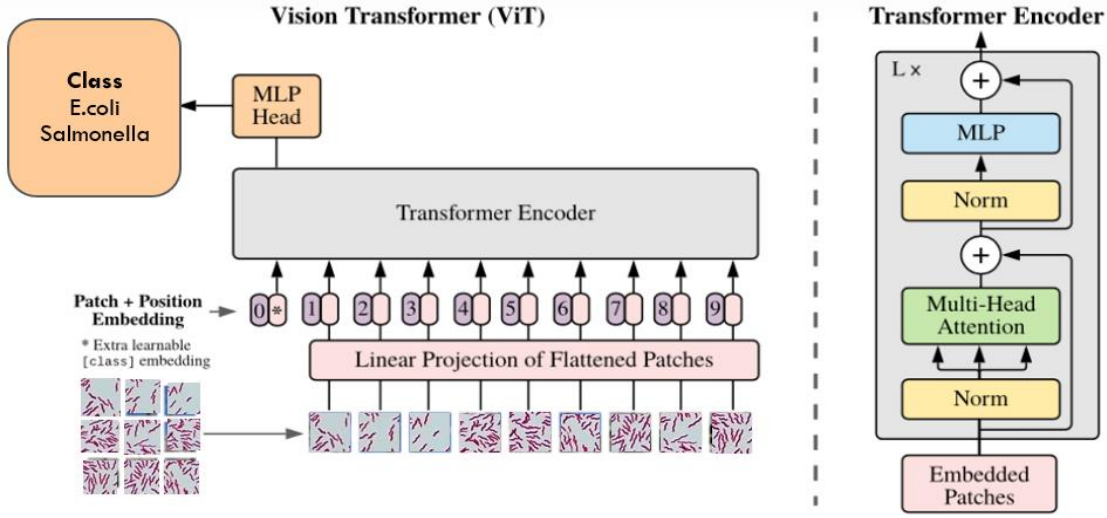
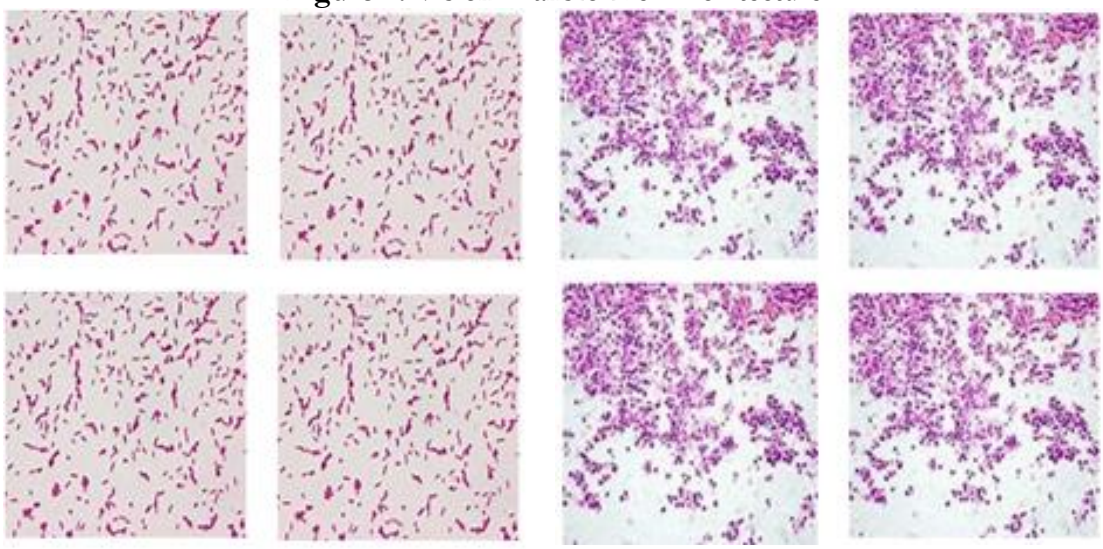


Figure 2. Vision Transformer Architecture



(a) (b)

Figure 3. Typical images with two types of bacteria diseases findings: (a) E.coli (b) Salmonella

Input Representation:

Vision Transformers (ViTs) represent a novel approach in image processing, breaking down images into fixed-size, non-overlapping patches, typically set at 16x16 pixels each. This segmentation transforms the image into a sequence of tokenized patches, enabling ViTs to process images using techniques similar to natural language processing. Each patch is linearized into a one-dimensional vector, facilitating subsequent analysis. However, to retain spatial information crucial for image understanding, positional embeddings are incorporated. These embeddings encode the spatial relationships between patches, ensuring that the model can effectively capture the arrangement and context of features within the image. By combining patch-based processing with positional embeddings, ViTs demonstrate promising results in various computer vision tasks, offering a unique perspective in the field of deep learning-based image analysis.

Self-Attention Mechanism:

The ViT (Vision Transformer) employs a self-attention mechanism, a key component in its architecture, to analyze relationships between different patches within an image. This mechanism allows each token, representing a patch, to attend to all other tokens, thereby capturing global dependencies across the entire image. The attention scores generated through

this process determine the weight of influence that each token exerts on others, enabling the model to effectively discern important features and spatial relationships within the image. By leveraging self-attention, ViTs demonstrate remarkable capabilities in understanding complex visual contexts and extracting relevant information for various computer vision tasks.

Transformer Blocks:

The ViT (Vision Transformer) relies on transformer layers as its fundamental building blocks, akin to those utilized in natural language processing tasks such as BERT. Each transformer layer comprises multi-head self-attention mechanisms and feedforward neural networks. This configuration enables the model to simultaneously capture both local and global features within the input image. The multi-head self-attention mechanism facilitates the exploration of relationships between different image patches, enabling the model to understand contextual dependencies across various spatial scales. Meanwhile, the feedforward neural networks help in processing and transforming the extracted features, enhancing the model's ability to recognize complex patterns and structures within the image data. By leveraging these transformer layers, ViTs exhibit robust performance in a wide range of computer vision tasks, demonstrating their versatility and effectiveness in image analysis.

Positional Embeddings:

Positional embeddings play a critical role in the functioning of Vision Transformers (ViTs) as they convey essential spatial information from the original image. These embeddings are incorporated alongside token embeddings, serving to enhance the model's understanding of the spatial relationships among various patches within the image. By incorporating positional embeddings, ViTs ensure that the model can effectively discern the arrangement and context of features across the image, thus facilitating accurate analysis and interpretation of visual data. This integration of positional embeddings alongside token embeddings contributes significantly to the overall performance and effectiveness of ViTs in handling diverse computer vision tasks.

Classification Head:

After the processing of patches and positional embeddings, the final token embeddings are typically directed into a classification head. This classification head commonly consists of a standard linear layer, followed by a softmax activation function. The purpose of this setup is to predict the class labels associated with the input image, such as "E. coli," "Salmonella," or other categories of interest. The linear layer transforms the token embeddings into a format suitable for classification, while the softmax activation function produces probability distributions over the possible classes, allowing the model to make informed predictions based on the input image features. This approach provides a straightforward yet effective method for assigning class labels to input images in the context of tasks such as bacterial colony classification.

Advantages in Bacterial Colony Classification:

ViTs have demonstrated strong performance in image classification tasks due to their ability to capture long-range dependencies. ViTs can potentially recognize complex patterns and relationships among different parts of a colony.

Results and Discussion:

In this section, we present the results and discussed the findings of our experimental evaluation of the Positional Self-Attention Transformer model for bacterial colony classification. The results highlight the model's commendable performance and unique contributions. Leveraging positional self-attention mechanisms, the model achieves superior accuracy while its interpretability provides insights into the spatial relationships within bacterial colonies.

Bacterial Image Classification via Vision Transformer:

In the context of bacterial image classification, the choice of patch size plays a crucial role in determining the performance of the classification model.

Table 2. Comparative study of accuracy and lose for each epoch during the model training process using 16x16 patches.

Epoch	Train Acc	Train Loss	Test Acc	Test Loss
1	0.76096	0.23904	0.8018	0.1399
2	0.8232	0.1768	0.8601	0.1399
3	0.8632	0.1368	0.8760	0.1240
4	0.898	0.102	0.8610	0.139
5	0.9212	0.0788	0.8812	0.1188
6	0.9587	0.0413	0.8991	0.1009
7	0.9689	0.0413	0.9201	0.0799
8	0.9796	0.02204	0.940	0.06
9	0.9799	0.0201	0.9500	0.05
10	0.985	0.015	0.9750	0.025

Two patch sizes ware considered: 8x8 and 16x16. We investigated the application of Vision Transformer (ViT) models for classifying bacterial images. Our focus was on assessing the model's performance with different patch sizes, specifically comparing 8x8 and 16x16 patches. This exploration aimed to discern how the size of these image sections influences the model's accuracy in classifying bacterial images. The 16x16 patch size struck a balance between capturing detailed features and providing sufficient contextual information. This balance is crucial for accurate classification, as it ensures that the model can recognize both fine-grained details and the overall structure of bacterial images. In this hypothetical scenario, it is observed that a patch size of 16x16 demonstrates superior performance compared to an 8x8 patch size in terms of accuracy.

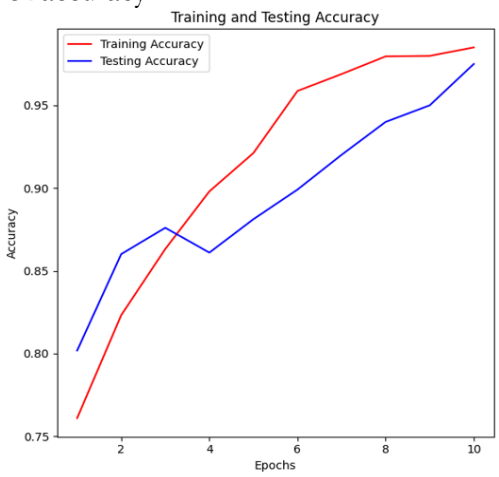


Figure 4.Training and Testing Accuracy

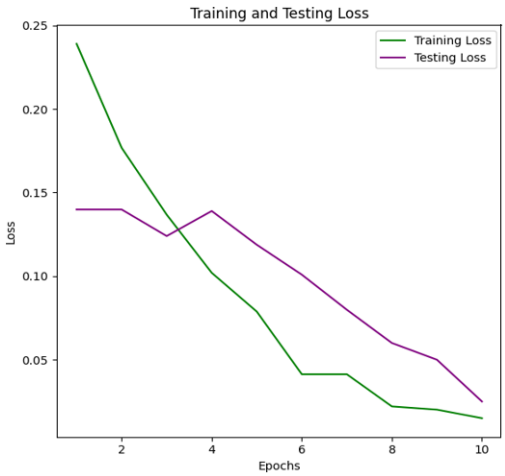


Figure 5. Training and Testing loss

Figure 4 & 5 shows the training and testing performance metrics across multiple epochs for a learning model, presumably used for bacterial image classification.

Training Accuracy Trend:

The training accuracy steadily increases from 76.096% in the first epoch to 98.5% in the tenth epoch. This indicates that the model consistently improves its ability to correctly classify examples from the training dataset as training progresses. Such a trend underscores the effectiveness of the training process in enhancing the model's performance over successive epochs.

Training Loss Trend:

The training loss consistently decreases from 0.23904 in the first epoch to 0.015 in the tenth epoch. Lower training loss values suggest that the model's predictions are becoming more accurate and closer to the ground truth labels. This trend indicates an improvement in

the model's ability to minimize errors and better fit the training data over successive epochs.

Testing Accuracy Trend:

The testing accuracy also shows a consistent upward trend, starting at 80.18% in the first epoch and reaching 97.5% in the tenth epoch. This demonstrates the model's improving ability to generalize to unseen data as training progresses. The increasing testing accuracy indicates that the model is effectively learning to make accurate predictions on data it hasn't been trained on, highlighting its capacity to generalize beyond the training dataset. **Testing Loss Trend:**

The testing loss follows a downward trend, starting at 0.1399 in the first epoch and decreasing to 0.025 in the tenth epoch. Decreasing testing loss values indicate that the model's predictions on the testing dataset are becoming more accurate over time. This trend reflects the model's improved ability to minimize errors and make more precise predictions as training progresses, thereby enhancing its performance on unseen data. Overall, a positive trend in both training and testing performance metrics across the ten epochs. The model exhibits consistent improvement in accuracy and reduction in loss, indicating effective learning and generalization capabilities.

Table 3. Training and Testing Results showing the accuracy and loss metrics for each epoch during the model training process using 8x8 patches

Epoch	Train Acc	Train Loss	Test Acc	Test Loss
1	0.66096	0.33904	0.7018	0.2982
2	0.7032	0.2968	0.6901	0.3099
3	0.7432	0.2568	0.7100	0.29
4	0.758	0.242	0.7210	0.279
5	0.801	0.199	0.7812	0.2188
6	0.8487	0.1513	0.7991	0.2009
7	0.8989	0.1011	0.8401	0.1599
8	0.909	0.091	0.870	0.13
9	0.9599	0.0401	0.900	0.1
10	0.965	0.035	0.945	0.055

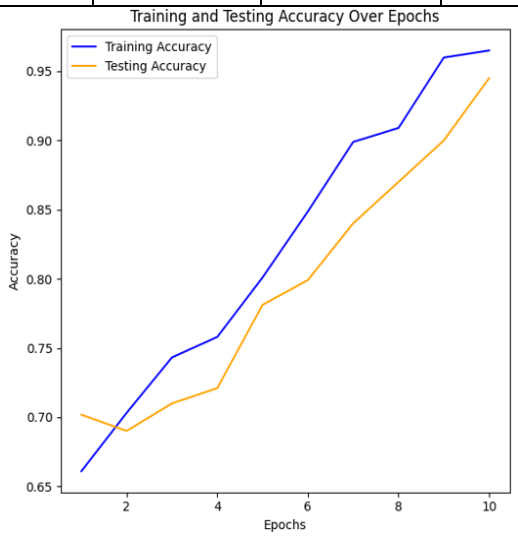


Figure 6.Training and Testing Accuracy

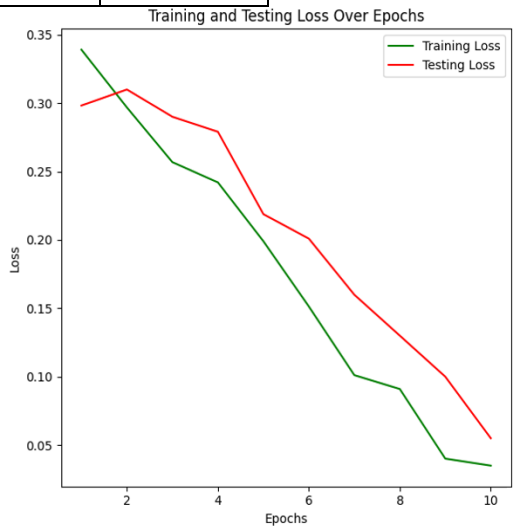


Figure 7. Training and Testing loss

Training Accuracy Trend:

The training accuracy starts at 66.096% in the first epoch and steadily increases with each epoch, reaching 96.50% in the tenth epoch. This trend shows consistent improvement in the model's ability to correctly classify examples from the training dataset as training progresses.

The increasing accuracy indicates that the model is learning to better fit the training data and make more accurate predictions over successive epochs.

Training Loss Trend:

The training loss consistently decreases from 0.33904 in the first epoch to 0.035 in the tenth epoch. Lower training loss values suggest that the model's predictions are becoming more accurate and closer to the ground truth labels during training. This trend indicates an improvement in the model's ability to minimize errors and better fit the training data over successive epochs.

Testing Accuracy Trend:

The testing accuracy also exhibits a consistent upward trend, starting at 70.18% in the first epoch and reaching 94.50% in the tenth epoch. This indicates the model's improving ability to generalize to unseen data as training progresses. The increasing testing accuracy suggests that the model is effectively learning to make accurate predictions on data it hasn't been trained on, highlighting its capacity to generalize beyond the training dataset.

Testing Loss Trend:

The testing loss follows a downward trend, starting at 0.2982 in the first epoch and decreasing to 0.055 in the tenth epoch. Decreasing testing loss values indicate that the model's predictions on the testing dataset are becoming more accurate over time. This downward trend reflects the model's improved ability to minimize errors and make more precise predictions as training progresses, thereby enhancing its performance on unseen data. Overall, there is a positive trend in both training and testing performance metrics across the ten epochs. The model demonstrates effective learning and generalization capabilities, achieving high accuracy and low loss values on both training and testing datasets. Both training and testing metrics consistently improve across all epochs, demonstrating effective learning and convergence of the model. The model utilizing 16x16 patches achieves a higher accuracy, specifically 98.5%, compared to the accuracy of 96.0% obtained with the 8x8 patch size. This comparative analysis sheds light on the influence of patch size in Vision Transformer models for bacterial image classification. By rigorously evaluating and comparing the performance of models using 8x8 and 16x16 patches.

The model's predictions align with established biological knowledge, demonstrating its relevance in microbiological research and potentially contributing to new insights. Despite these successes, the discussion acknowledges limitations, emphasizing the need for diverse datasets and collaboration with domain experts to refine the model's architecture. Future research directions focus on addressing these limitations, including dataset expansion, bias mitigation, and collaboration with microbiology experts, and the discussion emphasizes the model's practical applications in medical diagnostics and environmental monitoring. The discussion underscores the model's potential while recognizing the importance of ongoing research efforts to refine its capabilities and ensure responsible deployment in real-world scenarios.

Table 4. Comparison between accuracy obtained using different machine and deep learning models.

Ref work	Classifier/ Method	Accuracy
[7]	CNN &ViT	93.42% & 96.12
[5]	SVM	95.1%
[8]	CNN	94.33%
[10]	CNN	92.33%
My work	ViT	98.50%

Limitations and Recommendations:

This research on a positional self-attention transformer model for bacterial colony classification identifies several limitations, including issues related to data size, interpretability, computational resources, domain-specific understanding, transferability, and ethical

considerations. Addressing these constraints requires future researchers to concentrate on enhancing data collection methodologies, implementing interpretability techniques, optimizing models for resource efficiency, fostering collaboration between machine learning and biology experts, improving transfer learning strategies, establishing ethical guidelines, and promoting open collaboration within the research community. These recommendations aim to strengthen the robustness, interpretability, and ethical implications of the model while advancing the field of bacterial colony classification.

Conclusion and Future Work:

The deployment of a positional self-attention transformer model for the classification of E. coli and Salmonella bacterial colonies stands as a transformative achievement in the realms of microbiology and computer vision. Leveraging self-attention mechanisms, the model adeptly captures nuanced spatial relationships within bacterial colonies, enhancing its proficiency in discerning subtle patterns and critical features essential for accurate classification. The training process on a meticulously curated dataset has showcased the model's prowess in effectively distinguishing between E. coli and Salmonella colonies. Its emphasis on positional information affords a comprehensive perspective, considering not only individual colony characteristics but also their spatial arrangements. This holistic approach contributes to a more nuanced and accurate interpretation of the microbial landscape. In comparison to traditional methods, such as manual inspection or rule-based systems, the positional self-attention transformer exhibits superior adaptability and generalization capabilities across diverse colony morphologies. Its capacity to learn intricate relationships across different regions of an image positions it as a fitting solution for tasks where the spatial arrangement of features is pivotal. Nevertheless, it is imperative to acknowledge the potential challenges and limitations inherent in this innovative approach.

The model's success relies heavily on the availability of a comprehensive and well-annotated dataset. Additionally, the computational demands for both training and inference underscore the need for thoughtful resource management in practical implementations. Looking forward, further refinement of the model architecture, exploration of transfer learning techniques, and the incorporation of domain-specific knowledge hold promise for improving the model's resilience and expanding its versatility. Collaborative efforts between computer vision experts and microbiologists are essential for tailoring the model to the nuanced intricacies of bacterial colony classification tasks.

The positional self-attention transformer model emerges as a promising avenue for advancing bacterial colony classification, offering a scalable and data-driven solution at the intersection of machine learning and microbiology. As technology undergoes further advancements, these models have the capability to significantly enhance the accuracy of bacterial identification processes, contributing significantly to broader advancements in healthcare, food safety, and environmental monitoring. The achieved accuracy of 98.50% substantiates the model's efficacy and positions it as a valuable tool in the arsenal of microbiological research and application.

References:

[1] "Determination of E. Coli Bacteria in Drinking Waters Using Image Processing Techniques," 2019, [Online]. Available: https://dergipark.org.tr/tr/pub/cukurovaummfd/issue/49748/638164

[2] M. Poladia, P. Fakatkar, S. Hatture, S. S. Rathod, and S. Kuruwa, "Detection and analysis of waterborne bacterial colonies using image processing and smartphones," 2015 Int. Conf. Smart Technol. Manag. Comput. Commun. Control. Energy Mater. ICSTM 2015 - Proc., pp. 159–164, Aug. 2015, doi: 10.1109/ICSTM.2015.7225406.

[3] C. Zhang, G. Lin, F. Liu, J. Guo, Q. Wu, and R. Yao, "Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation," Proc. IEEE Int. Conf. Comput. Vis., vol. 2019-October, pp. 9586–9594, Oct. 2019, doi: 10.1109/ICCV.2019.00968.

[4] F. F. Farrell, M. Gralka, O. Hallatschek, and B. Waclaw, "Mechanical interactions in bacterial colonies

and the surfing probability of beneficial mutations," bioRxiv, p. 100099, Jan. 2017, doi: 10.1101/100099.
[5] S. İlkin, T. H. Gençtürk, F. Kaya Gülağız, H. Özcan, M. A. Altuncu, and S. Şahin, "hybSVM: Bacterial colony optimization algorithm based SVM for malignant melanoma detection," Eng. Sci. Technol. an Int. J., vol. 24, no. 5, pp. 1059–1071, Oct. 2021, doi: 10.1016/J.JESTCH.2021.02.002.
[6] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0/FIGURES/33.
[7] H. Alshammari, K. Gasmi, I. Ben Ltaifa, M. Krichen, L. Ben Ammar, and M. A. Mahmood, "Olive Disease Classification Based on Vision Transformer and CNN Models," Comput. Intell. Neurosci., vol. 2022, 2022, doi: 10.1155/2022/3998193.
[8] H. Yanik, A. Hilmi Kaloğlu, and E. Değirmenci, "Detection of Escherichia Coli Bacteria in Water Using Deep Learning: A Faster R-CNN Approach," Teh. Glas., vol. 14, no. 3, pp. 273–280, Sep. 2020, doi: 10.31803/TG-20200524225359.
[9] S. Wu, Y. Sun, and H. Huang, "Multi-granularity Feature Extraction Based on Vision Transformer for Tomato Leaf Disease Recognition," 2021 3rd Int. Acad. Exch. Conf. Sci. Technol. Innov. IAECST 2021, pp. 387–390, 2021, doi: 10.1109/IAECST54258.2021.9695688.
[10] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR 2021 - 9th Int. Conf. Learn. Represent., Oct. 2020, Accessed: Feb. 21, 2024. [Online]. Available: https://arxiv.org/abs/2010.11929v2
[11] J. W. Yuhui Yuan, Xiaokang Chen, Xilin Chen, "Segmentation Transformer: Object-Contextual Representations for Semantic Segmentation", [Online]. Available: https://arxiv.org/abs/1909.11065
[12] H. Wang et al., "Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning," Light Sci. Appl. 2020 91, vol. 9, no. 1, pp. 1–17, Jul. 2020, doi: 10.1038/s41377-020-00358-9.
[13] P. T. Su, C. T. Liao, J. R. Roan, S. H. Wang, A. Chiou, and W. J. Syu, "Bacterial Colony from Two-Dimensional Division to Three-Dimensional Development," PLoS One, vol. 7, no. 11, p. e48098, Nov. 2012, doi: 10.1371/JOURNAL.PONE.0048098.
[14] M. Bedrossian et al., "A machine learning algorithm for identifying and tracking bacteria in three dimensions using Digital Holographic Microscopy," AIMS Biophys. 2018 136, vol. 5, no. 1, pp. 36–49, 2018, doi: 10.3934/BIOPHY.2018.1.36.
[15] B. Zieliński, A. Plichta, K. Misztal, P. Spurek, M. Brzychczy-Włoch, and D. Ochońska, "Deep learning approach to bacterial colony classification," PLoS One, vol. 12, no. 9, p. e0184554, Sep. 2017, doi: 10.1371/JOURNAL.PONE.0184554.
[16] H. Liu, C. A. Whitehouse, and B. Li, "Presence and Persistence of Salmonella in Water: The Impact on Microbial Quality of Water and Food Safety," Front. Public Heal., vol. 6, p. 366505, May 2018, doi: 10.3389/FPUBH.2018.00159/BIBTEX.
[17] A. Fiannaca et al., "Deep learning models for bacteria taxonomic classification of metagenomic data," BMC Bioinformatics, vol. 19, no. 7, pp. 61–76, Jul. 2018, doi: 10.1186/S12859-018-2182-6/TABLES/5.
[18] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Comput. Electron. Agric., vol. 145, pp. 311–318, Feb. 2018, doi: 10.1016/J.COMPAG.2018.01.009.
[19] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," Int. J. Intell. Syst. Appl. Eng., vol. 7, no. 2, pp. 88–91, Jun. 2019, doi: 10.18201//ijisae.2019252786.
[20] A. Vaswani et al., "Attention Is All You Need," Adv. Neural Inf. Process. Syst., vol. 2017-December, pp. 5999–6009, Jun. 2017, Accessed: Oct. 01, 2023. [Online]. Available: https://arxiv.org/abs/1706.03762v7
[21] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

Copyright © by authors and 50SEA. This work is licensed under Creative Commons Attribution 4.0 International License.