

FINAL YEAR PROJECT

NAME:	Charles Robinson
DEGREE COURSE:	MSci Physics
PROJECT TITLE:	Developing a neural network algorithm to perform rapid recognition of bacterial species
YEAR OF SUBMISSION:	2022
SUPERVISOR:	Dr Massimo Antognazzi
NUMBER OF WORDS:	7890



Declaration

The results presented within the Neural Network section of the report were all obtained by myself, Jack McHaffie, and Alejandro Ortega as part of their 30 credit point Master's projects. The investigations and results into classical machine learning were performed solely by myself to allow my project to become 60 credit points. Analysis of all results is my own. Data was collected by Dr Massimo Antognazzi's research group. Parts of the pre-processing code were based on code created by Hollie Hindley and Gabriel Gibb, who were previous year's Masters students. Work was split evenly between Jack, Alejandro and myself, with each of us working on all sections of the code and cloud management. The impact of COVID-19 upon the project was minimal, as the project had always been intended to be computational in basis. Initial plans for myself to collect additional data did not go ahead due to additional strain in the lab due to COVID.

Acknowledgements

I would like to give special thanks to my project partner Alejandro Ortega and Jack McHaffie, whose work ethic and commitment is to be admired. This project would have been much harder without their insights, support and friendship. I would also like to express my gratitude to my supervisor Dr Massimo Antognazzi for his invaluable support, enthusiasm and interest in the project. Furthermore, I am very grateful to all other students present in our weekly meetings, many of whom had helpful ideas and comments.

Abstract

Sub-cellular fluctuation imaging (SCFI) is a new optical method which captures fluctuations related to bacterial metabolic activity. This paper investigates whether these fluctuations are characteristic to bacterial strain and explores their use for classification. The classification potential of SCFI is evaluated by comparing videos of bacteria captured using SCFI with images captured using Brightfield microscopy. Five deep neural network models were implemented for three strains of bacteria: *Neisseria gonorrhoeae*, *Moraxella catarrhalis*, and *Streptococcus pneumoniae*. The best model produced a class average test accuracy of 96.6%. Further, classical computer vision and machine learning approaches were explored to aid interpretability of these results. Results indicate that SCFI imagery contains distinct information which could be used for bacterial classification.

Contents

1	Introduction	5
2	Theory	5
2.1	Bacteria and Classification	5
2.1.1	Live-cell imaging	6
2.1.1.1	Brightfield microscopy	6
2.1.1.2	Sub-cellular fluctuation imaging	7
2.2	Machine learning	8
2.2.1	Decision trees	8
2.2.2	Artificial Neural Networks	9
2.2.2.1	Multi-layer perception networks	9
2.2.2.2	Convolutional neural networks	10
2.2.2.3	Recurrent neural networks	11
3	Methodology	12
3.1	Data	12
3.1.1	Bacteria imaging data	12
3.1.2	Data pre-processing	13
3.1.3	Data augmentation	14
3.2	Neural Network approach	15
3.2.1	Training	15
3.2.2	Single frame models	16
3.2.2.1	Model A: Brightfield only	16
3.2.2.2	Model B: SCFI 1st frame	17
3.2.2.3	Model D: Brightfield and SCFI 1 frame dual channel	17
3.2.3	Multi frame frame models	17
3.2.3.1	Model C: SCFI 10 frames	17
3.2.3.2	Model E: Brightfield and SCFI 10 frames dual channel	17
3.3	Classical ML approach	17
3.3.1	Feature engineering	18
3.3.2	Feature importance	18
3.3.3	Multi frame models	19
4	Results and Discussion	19
4.1	Neural Network results	20
4.1.1	Model A: Brightfield only	21
4.1.2	Model B: SCFI 1st frame	23
4.1.3	Model C: SCFI 10 frames	23
4.1.4	Model D: Brightfield and SCFI 1 frame dual channel	24
4.1.5	Model E: Brightfield and SCFI 10 frames dual channel	24
4.2	Classical machine learning results	27
5	Conclusions and Further Research	28

Appendices	30
A Appendix A: Model architecture diagrams	30
B Appendix B: Model results	35

1 Introduction

The rapid and accurate classification of bacterial species is essential to aid food and water safety testing [1] [2], clinical diagnoses [3], and biosecurity [4]. In particular, rapid and cost-effective classification can help determine whether antibiotics should be prescribed for a particular treatment. Bacterial infections claim the lives of over 6.7 million people annually [5], yet current classification methods, such as medium culture, biochemical reactions and serological tests, can be slow and costly. According to Public Health England (PHE), 20% of patients are being treated with antibiotics unnecessarily while waiting for results [6]. This is not only costly but leads to increased antimicrobial resistance [7]. Therefore, it is crucial to develop novel methods to classify bacteria rapidly, inexpensively and accurately.

Due to the affordability of computing power and improved tooling, machine learning (ML) has become the prominent approach in bacterial classification, providing advantages over manual methods which are slow and prone to error. Recent publications have shown that deep learning Convolutional Neural Networks (CNNs) and Long Short-term Memory (LSTMs) can achieve state of the art accuracy on classification of bacterial species [8]. However, deep learning can lack the interpretability of classical machine learning techniques such as decision trees [9]. Interpretability is essential in medical setting to uncover unconscious bias in the development of models.

Some inexpensive and straightforward methods of bacterial classification have recently appeared, such as Raman spectroscopy [10] and the MALDI-TOF mass spectrometer [11] [12]. There are, however, limitations in the range of bacterial species that these can classify. For example MALDI-TOF mass spectrometry has difficulties discriminating closely related bacterial species such as *Escherichia coli* and *Shigella* [13]. Therefore, novel classification methods are required to increase coverage over bacterial species.

In the field of live-cell imaging (the study of living cells using time-lapse microscopy) sub-cellular fluctuation imaging (SCFI) [14] is a new optical method which could potentially offer a faster and more cost-effective method of bacterial classification. SCFI images fluctuations linked to a bacteria's metabolism, which may be characteristic to different bacterial strains.

The overall aim of this paper is to investigate whether SCFI is a viable method for classification of bacterial strains. Additionally, to aid with the interpretability of deep learning models, a classical machine learning approach using decision trees will be used. The aim of this method is to evaluate which characteristics of SCFI are used in the classification. Hence, the objectives of this work are to: develop a deep learning approach aimed at achieving accuracy, develop a classical machine learning approach using decision trees, and compare the use of SCFI for classification with a standard imaging method, Brightfield microscopy. For this purpose, a dataset consisting of SCFI videos and Brightfield images containing three strains of bacteria has been used.

2 Theory

In this work we use data from two different live-cell imaging techniques to image bacteria, namely Brightfield microscopy and Sub-cellular fluctuation imaging. Images are then classified using three distinct machine learning approaches: Convolutional neural networks, long short term memory networks, and decision trees. This section gives a brief explanation of these techniques and approaches.

2.1 Bacteria and Classification

In this work three bacteria were imaged: *Neisseria gonorrhoeae* (MS11), *Moraxella catarrhalis* (MX2) and *Streptococcus pneumoniae* (SN2). These bacteria are all indicators of a (potentially severe) infection, therefore designing

methods to classify these bacteria is essential.

Nesseria Gonorrhoeae is a causative pathogen of the sexually transmitted disease gonorrhoea. It has an estimated 106 million infections each year and a high risk of complication from infection due to increase in antibiotic-resistant strains.

Moraxella catarrhalis is found in the human respiratory system. It can cause pneumonia and meningitis by infecting the lower respiratory tract of adults with chronic lung disease.

Streptococcus pneumoniae causes pneumonia in adults and children. There has been a rise in drug resistant strains due to genetic mutations that allow drugs to block it from binding [15].

Different methods are used to detect and classify bacteria. Microbiologists manually observe phenotypic characteristics of bacteria such as colour, shape or size. As some bacteria species are extremely similar other methods, such as gram staining, biochemical testing or motility testing amongst others, are required in order to successfully classify different bacteria species [16].

However, the classification of bacteria using these methods is time consuming and prone to error. Consequently, several ML techniques have become popular as alternatives to traditional methods. Machine learning has proven to be more fast, accurate, reliable, and cheap than the aforementioned classification methods [17]. ML models are often trained on bacteria data from different sources, a primary source used in classification being live-cell imaging [18].

2.1.1 Live-cell imaging

Live-cell imaging studies living cells using time-lapse microscopy. This technique is used to obtain a better understanding of living cells through the study of cellular dynamics [19].

Different types of microscope can be used, such as transmitted light microscopy, fluorescent and confocal microscopy. Some of these techniques require labelling in order to support the imaging of a particular bacteria characteristic. Labelling involves chemically attaching a molecule to aid detection of a molecule within the bacteria. Label free live-cell imaging is important as the more a bacterium is altered the less its behaviour will match its native state.

2.1.1.1 Brightfield microscopy

Brightfield microscopy is a common type of transmission light microscopy imaging. It is a popular technique for classifying bacteria as it is very simple and versatile [20] [8] [21]. The Figure 1 below summarises the principles of a Brightfield microscope.

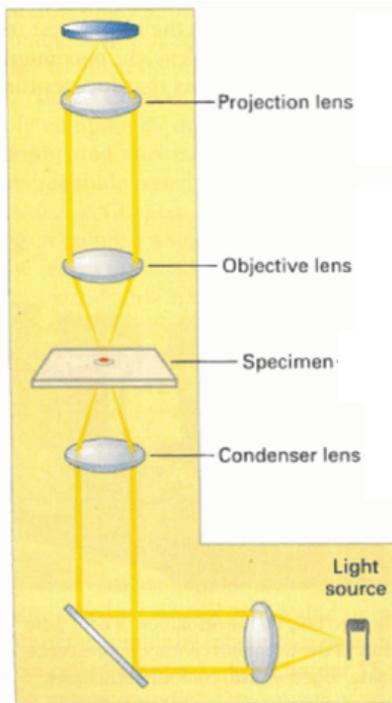


Figure 1: Principles of a Brightfield microscope. Adapted from [22]

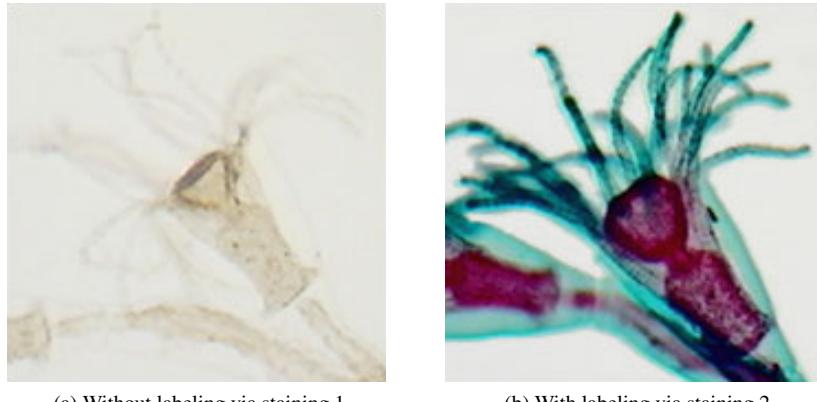


Figure 2: Example of Brightfield imaging showing the effect of staining to increase contrast of sample. Adapted from [23]

For Brightfield microscopy a sample is illuminated by a transillumination light source (often a halogen lamp). In essence, it consists of the shining of a halogen lamp through a cell from below and imaged from above. This technique can provide results with a low contrast, or definition, for many biological samples an example of which is shown in Figure 2. This is due to very few living cells absorbing light to a great extent. As a result it provides less information to classify a sample. Samples often require labelling via staining to increase contrast, which adds another step and makes the process more complex and resource-intensive [24].

2.1.1.2 Sub-cellular fluctuation imaging

Sub-cellular fluctuation imaging is a recently-developed and innovative method of imaging bacteria. It poses significant advantages over Brightfield microscopy as it does not require bacteria to be labelled. This reduces the

complexity of the technique and enables it to be used for rapid high throughput classification [14]. Bacterial discrimination using SCFI has previously been shown to be possible [25].

SCFI can image sub-cellular fluctuations which are nanomechanical motions at characteristic frequencies. The frequencies have a strong dependence on the cell's temperature and metabolic state, indicating that they are biologically driven [14]. Figure 3 shows an example of an E.coli bacterium being imaged by SCFI.

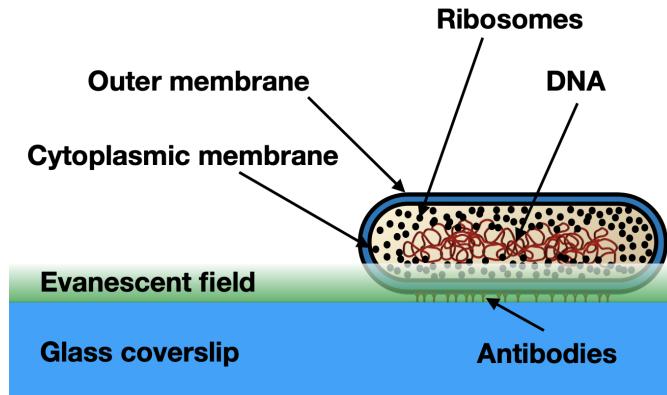


Figure 3: E.coli bacterium being imaged by SCFI. The bacterium is bound with antibodies to an amino function-alised coverslip and then suspended in a flow cell. An evanescent field is created by the total internal reflection of a laser (1mw, 561nm) shone through a high numerical aperture lens. The field decays into the bacterium illuminating 100nm into the surface of the bacterium. Metabolic activity within the bacterial cytoplasm scatters the evanescent field. By measuring the scattered light, the sub-cellular fluctuations can be imaged. Adapted from [14]

2.2 Machine learning

Machine learning (ML) is a leading method in classifying bacteria [17]. Machine learning is a powerful set of techniques which create programs learned from patterns in data without them being explicitly programmed.

A feature is a characteristic of the phenomenon being analysed [26], and ML models learn how to discriminate based on features. For instance, when trying to classify giraffes and horses the height of the animal could be used as a feature.

Feature engineering is the process of creating additional features from data to improve the accuracy of ML models. By way of example, given an image of an bacterium in addition to the raw RGB image, the following features could be created:

- Edges - the outline of the bacteria.
- Segmentation - the area of the bacteria.
- Thresholding - removing colours below or above a particular brightness in order to delete the background.
- Histogram - the number of pixels of each colour in the image.

2.2.1 Decision trees

Neural networks have been shown to often outperform traditional ML techniques like decision trees [27]. However, neural networks often lack the interpretability which decision trees can provide, with interpretability being particularly useful in determining why a model works. This balance between performance and understanding is known as the accuracy-interpretability trade-off [28].

Decision trees are a ML method in which data is continuously split according to a certain parameter. Given a set of features, decision trees iteratively split based on a feature value test. The splitting is completed when the stopping criteria is met. Stopping criteria can be the depth of the tree, the purity of the node or the predictor values for all records which are identical.

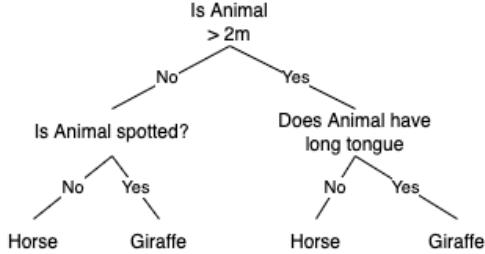


Figure 4: An example of a decision tree classifying horses and giraffes. Nodes represent decisions, branches are choices, and leaves are the predicted class.

While decision trees can be advantageous over neural networks as they can be more easily interpreted, they cannot create their own features as they require well engineered features to predict accurately.

Random forest decision trees (RFDTs) are an ensemble technique. Ensembles are combinations of multiple models used in order to improve performance. RFDTs use multiple decision trees and vote on the final classification.

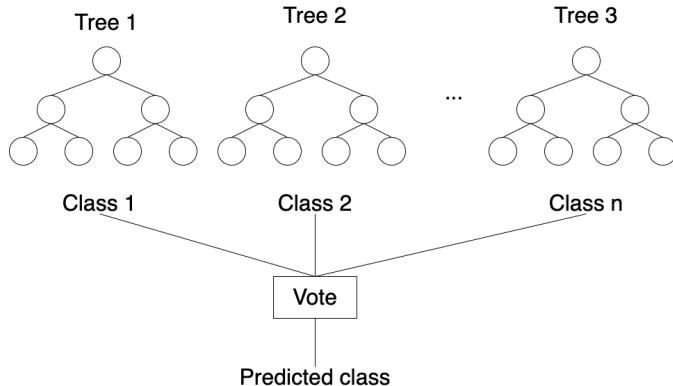


Figure 5: Structure of a random forest decision tree. Multiple independent decision trees are trained and predictions from each are averaged to produce the final prediction.

2.2.2 Artificial Neural Networks

Artificial Neural Networks (ANNs) are a computation model design based on a simplified view of biological neurons found in the brain. The fundamental computation unit is an artificial neuron which has input and output connections. A neuron computes a nonlinear function of the weighted average of the input connections. The weight is passed to each output connection. A neural network is a series of connected neurons used for computation.

2.2.2.1 Multi-layer perception networks

Multi-layer perception networks (MLPs) are a class of ANNs which can distinguish non-linearly separable data. They are arranged in layers with each adjacent layer fully connected to the next layer. This is also known as being densely connected. There are three types of layer:

- Input layer - the initial layer which passes data to the next layer without processing it.
- Hidden layer - layers in the middle of the network. These perform computations and transfer information to the output layer.
- Output layer - the final layer which transfers the result of the computation, the prediction, out of the network.

A network alone has no discriminatory power and must be trained on data to learn how to discriminate new data. Training initially assigns random weights to each input in order to attach more importance to certain connections than others. The inputs are multiplied by their weights and summed with a bias. An activation function is applied to this value to give the output value which is passed to the next layer. This activity repeats in a forward movement through layers of the network in a process known as forward-propagation.

When a final output is determined an error is calculated using a cost function relating to how far the output is from the expected output. The network then moves backwards and uses this error value to update the weights, where neurons contributing most to the error have their weights minimised, through a process known as back- propagation. This has the effect of either diminishing or magnifying the inputs which is a way of quantifying how important a feature is towards the ultimate result. One full pass forwards and backwards through the data is known as an epoch, and training will consist of many of these.

The number of hidden layers is known as the depth of a network. The deeper a network, the more features it can learn at different levels of abstraction [29]. The number of neurons in a layer is known as the width of a layer. Wider networks can memorise more information.

How well a network performs on new data is referred to as generalisation. If a network does not generalise well it is said to have 'overfit' the data. Rather than learning a general function, the network has memorised the training data and performs poorly with new data.

Several techniques have developed to prevent over fitting:

- Data augmentation - amplifying existing data by applying random transformations such as translating, rotating and skewing. This can help make a model more robust and better able to generalise to new data.
- Data splitting - by only training on a portion of the data and keeping a portion unseen, we can test if the model has generalised well. Often data is divided into training, test and validation sets.
- Regularisation - a range of techniques which simplify a model. Simpler models are less prone to overfitting. In neural networks, a dropout layer is a layer which randomly disables some connections between neurons.

2.2.2.2 Convolutional neural networks

Convolutional neural networks (CNNs) are the core technique used in our methodology. We selected CNNs over other ANN architectures as there is a strong prevalence of CNNs achieving state of the art performance with classification tasks. For example, a study classifyied ten species of marine microbes [30] using a 1-dimensional CNN, with all methods achieving 95% accuracy. Another study MALDI-TOF MS identified 6685 strains of anaerobic bacteria and achieved an accuracy of 84% by species [31]. However, there are limitations, as a ML approach using MALDI-TOF MS found that taxonomic information is not well preserved [32].

CNNs are a class of ANNs which are inspired by how neurons in the visual cortex of animals are connected. CNNs have been shown to work well on grid-like data. A gray scale image can be thought as a 3D grid of the rows, column and colour of each pixel. A video is a 4D grid of images over time.

CNNs introduce specialist layers:

- Convolutional layers - apply filters to create feature maps which summarise features across their receptive field. A filter can be thought of as a square matrix. The filter is progressively applied over every grid point in the image performing an elementwise multiplication and summing the results to make the output for that grid point. The combination of all the filtered grid points is the feature map.
- Pooling layers - follow convolutional layers summarising the output of the network by looking at outputs around it (the 'neighbourhood'). This reduces the spacial size of the representation. Max pooling is the most commonly used and selects the maximum output from the neighborhood. Max pooling layers also provide translation invariance.
- Dense layers - these layers are fully connected and often at the final layers of the network which produce the end result.

Three core properties of CNNs make them generalise well on vision problems:

- 3D volume - layers have width, height and depth. Neurons are only connected to a small region of the layer before it known as the receptive field.
- Spatial structure - CNNs take into account the spatial structure of the data. Pixels which are close together spatially remain close in the network, while MLPs flatten out an image.
- Shared weights - filters are replicated across the entire visual field sharing the same parameterization and form a feature map. This allows features to work regardless of translation in the image.

2.2.2.3 Recurrent neural networks

Recurrent neural networks (RNN) and specifically long short-term memory (LSTM) networks are a core approach used in this study. LSTMs were selected over other time domain models such as transformers as they have been shown to successfully classify bacteria from video [33]. Notably, this classification was faster than traditional methods used. Another example used a 'Long Short-Term Memory' (LSTM) network for antibiotic susceptibility testing of bacteria [34].

An RNN is a class of ANN which can contain cycles such that the output of a hidden layer can be fed back into the input of a previous layer.

This property allows these networks to exhibit temporal dynamic behaviour as they can maintain memory over time, making them well suited for problems with sequences of data. For example with a video, information in previous frames provide context to help classify what is happening in the current frame.

RNNs have an issue known as the 'vanishing gradient' which means they have difficulties learning relationships which are several steps apart [35]. LSTMs are a class of RNN designed to avoid the vanishing gradient problem.

An LSTM network is a RNN with special units which can maintain information in memory for longer periods of time. [36]. LSTMs give more control over memory then a RNN.

LSTMs can be combined with other types of network such as a CNN to build specialist networks.

3 Methodology

3.1 Data

3.1.1 Bacteria imaging data

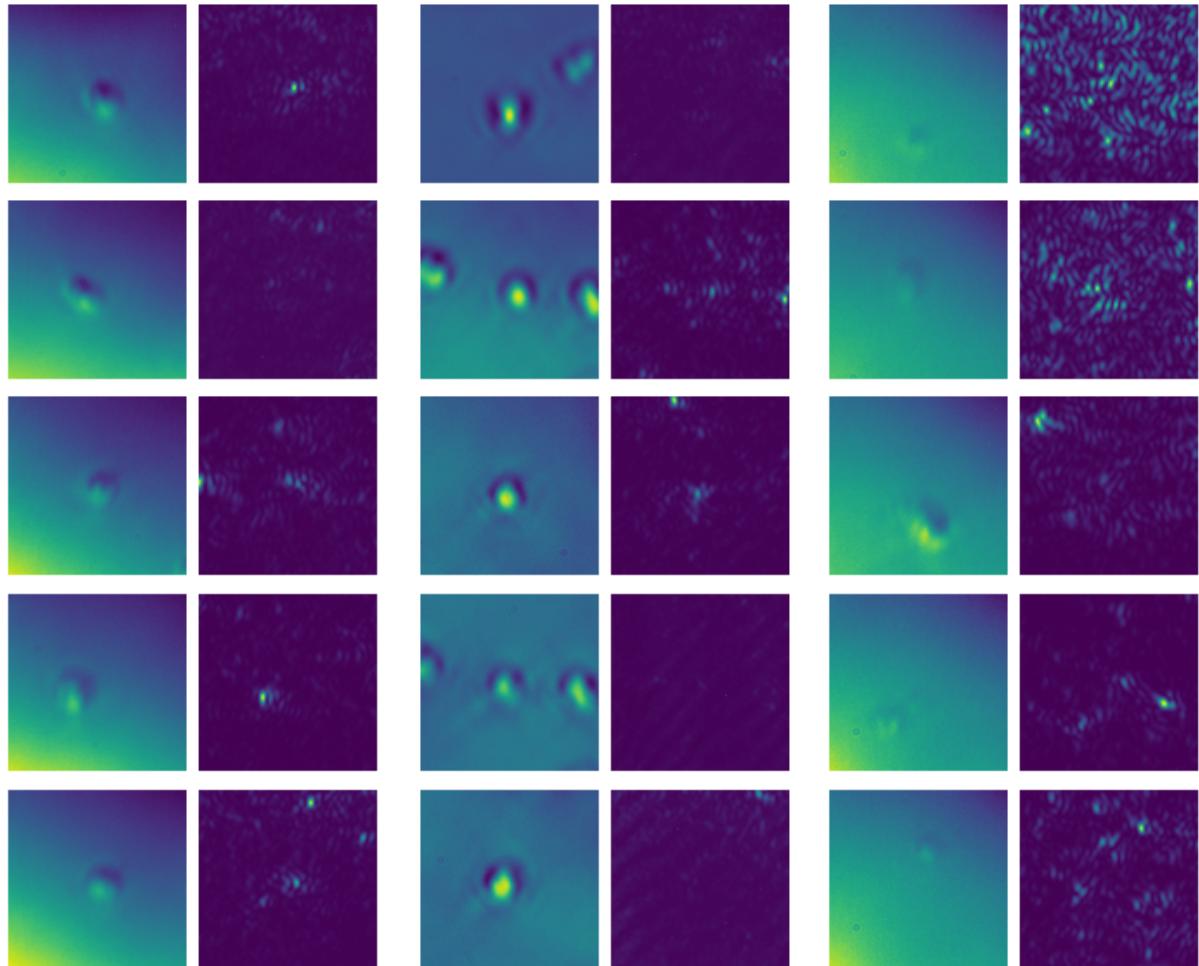


Figure 6: Sample of Brightfield images and first frame from SCFI videos for each bacteria classes. From left to right: MS11 Brightfield, MS11 SCFI, MX2 Brightfield, MX2 SCFI, SN2 Brightfield, SN2 SCFI

Data was sourced from the University of Bristol Research Data Storage Facility (RSDF) server. A sample of data is shown in Figure 6. This data consists of 594 Brightfield images and 594 10 second (at 20 frames per second) SCFI videos for three strains of bacteria:

- *N. gonorrhoeae* (MS11).
- *M. catarrhalis* (MX2).
- *S. pneumoniae* (SN2).

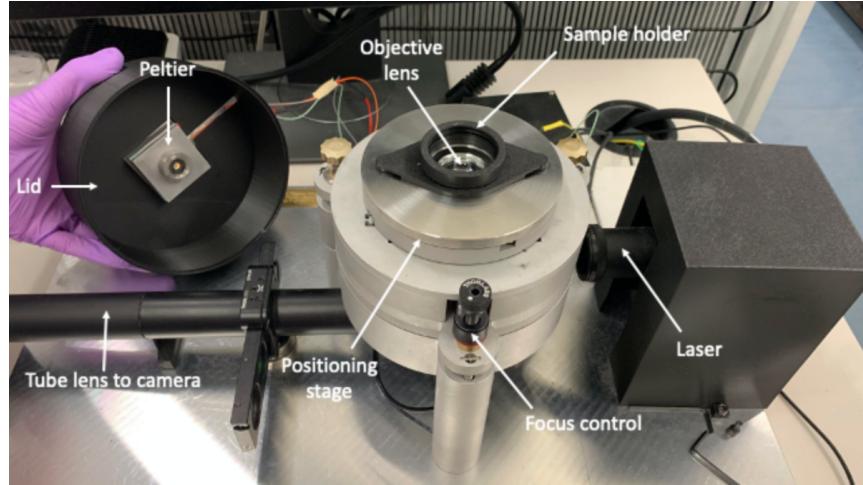


Figure 7: The SCFI Microscope used by [25] to capture both Brightfield and SCFI images. Adapted from [25]

Data was collected from multiple sources. This may have introduced experimental inconsistencies. Optical artefacts, which are present in the data, can be explained by: exposure, laser, scattering intensity changes, and other factors. Steps were taken to minimise experimental inconsistencies in the data. Outlier detection and data preprocessing techniques designed to minimise these limitations are described in Section 3.1.2.

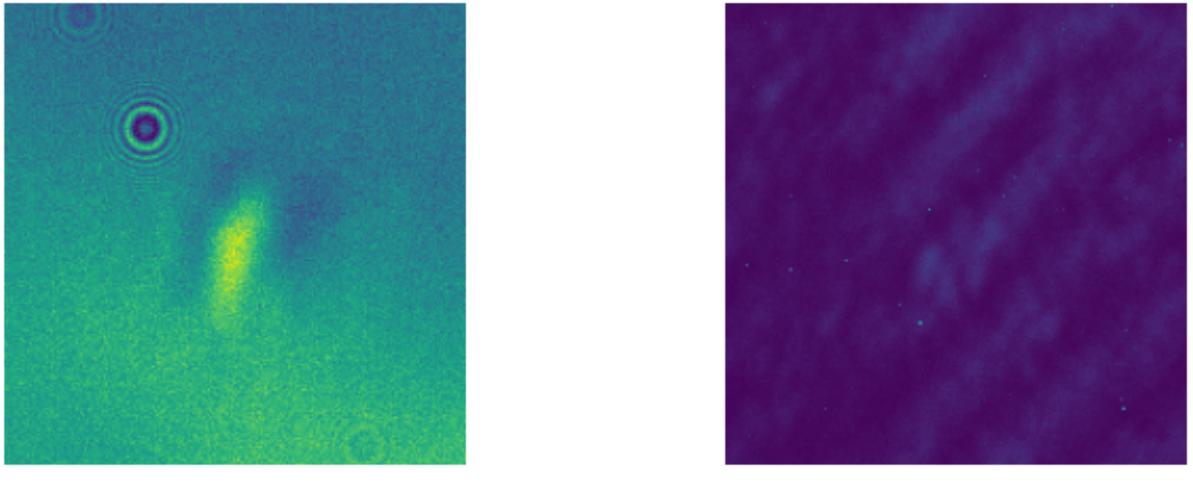
An example of the method used to collect the data is shown by [25], a summary of their method being as follows. Each sample was cultured on HBHI agar plates for 16 hours at 37°C before imaging. MS11 and SN2 were kept at 37°C, whilst MX2 was imaged at room temperature. Since fluctuations are temperature dependent, it was important that the temperature was controlled during imaging. Each sample was exposed to the SCFI setup similar to that shown in Figure 7. By turning off the laser and turning on the LED Brightfield imaging was also taken of each sample.

Fluctuations were quantified by calculating the central coordinates of each bacteria. Pixel intensities were normalised by averaging the intensity across the region of interests for all frames. Optical artefacts in the data can be explained by: exposure, laser, scattering intensity changes and other factors.

3.1.2 Data pre-processing

Data was labeled according to the bacteria class. Gray-scale images and video frames were supplied in TIFF format and for our neural network approach pixel colour values were normalised between 0 and 1. This is important as it ensures each of the models' input parameters have a similar data distribution and makes convergence faster while training the network [37].

Optical artefacts present in the data may lead to poor generalisation of the ML models. Artefacts may be caused by exposure, laser and scattering intensity change [14]. Examples of artefacts are shown in Figures 8a and 8b. As neural network models are 'black boxes', they may use these artefacts to discriminate between bacteria. For example, if artefacts appear in images of MS11 but not in other class images this could be used by a model to classify MS11.



(a) Brightfield artefact

(b) SCFI artefact

Figure 8: Brightfield image with three visible diffraction pattern artefacts and SCFI frame with visible diagonal patterns

To minimise models using artefact information for classification, videos were cropped from 300x300 pixels to 200x200 pixels around bacteria centers. This reduces the information provided to the model and helps to filter some artefacts. Experiments were performed to find the optimal size for cropping, with 200x200px found to be approximately the optimal size for this cropping. Experiments consisted of observing where image intensity drops off from superposition of frames by class and by running a preliminary image recognition model for different sizes of frame.

Manual outlier detection was performed and five videos were removed from the data set due to experimental errors. In particular, some SN2 SCFI images appear to have been over exposed during the experiment. The removal of these outliers from our data set is important since it minimises the presence of systematic errors in the work.

Data was split according to a train-test-validation split of 80:10:10 equally across each class. 80:10:10 was selected as it this is a common ratio used in similar training problems [38] [39]. Training data was used to train the model, validation data was used to tune the parameters of the model, and the test set was used to evaluate the model's performance.

3.1.3 Data augmentation

Data augmentation can lead models to generalise better to unseen examples. Augmentation methods should not impact the features a model uses to classify but instead make a model robust to changes in the input data such as the orientation of the bacteria on the slide. A number of data augmentation methods were tried:

- Vary the brightness of the image by $\pm 10\%$
- Vary the contrast of the image by $\pm 10\%$
- Vertical and horizontal flip.
- Random rotation by $\pm 10\%$
- Random crops by $\pm 10\%$ and zeroing everything outside the cropped region.
- Random shifts by $\pm 10\%$

For each type of data augmentation, model A and model B were trained with the original data, plus the augmented data. It was found that only flips and small rotations improved the model's accuracy. This indicates that other augmentation techniques reduced the discriminatory power of the models by altering characteristic features used for classification. This can be due to:

- Different bacterial strains have different pixel intensities and varying the brightness and contrast of an image will alter these pixel intensities, making it harder for models to classify.
- Random rotations and crops were found to remove parts of the image which may be used by the models for classification. While the bacteria should be rotation independent, other background features of the images may not be.

Therefore, only horizontal and vertical flips as well as small rotation were used as data augmentation techniques in the final data pipeline. A random rotation of 0.1 rad (6 degrees) was found to improve classification, perhaps due to small changes in bacterial alignment in the flow cells. Small changes in alignment change the fluctuation patterns and so could confuse models.

3.2 Neural Network approach

We chose two different deep neural network architectures because there were two different input shapes. Single frame models are based on CNNs and take a single frame of either Brightfield or SCFI imagery. Multi-frame models are based on LSTMs and CNNs and take a sequence of SCFI frames.

Methods were implemented using TensorFlow Keras. TensorFlow Keras was chosen due to its wide use in industry, broad range of tooling and integration with cloud computing services.

Network architecture was inspired based on existing models in bacteria classification, high performing models in similar domains and neural architecture search. Neural architecture search (NAS) is a technique used to help design networks and works by training many different models to find useful architectures. Autokeras was used for this purpose, a module which performs NAS with Keras models. The target of the NAS was to minimise the validation loss (detailed in 9). Models were not as accurate as models designed by hand but they provided insight into building our models as they ran tests on different techniques and methods. For example, after trialing 200 different models it was found that a normalisation layer was a useful addition to models in order to increase accuracy.

3.2.1 Training

Neural network training can be memory and compute intensive. Using cloud computing can speed up training by orders of magnitude, enabling more network architectures being tried. Amazon Web Services (AWS) and Google Cloud (GC) are two of the major cloud computing providers. After evaluating both, we decided to use GC due to better collaborative tools such as google colab and closer integration with TensorFlow. Training was performed on a NVIDIA Tesla K80 GPU with 12GB of memory on google compute engine.

$$-\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C 1_{y_i \in C_c} \log(p_{model})[y_i \in C_c] \quad (1)$$

Figure 9: Equation for categorical cross entropy where the double sum is over the observations i , whose number is N , and the categories c , whose number is C . p_{model} is the probability predicted by the model for the i th observation to belong to the c th category.

During training, the training procedure seeks weights for the neural network to minimise the loss function. We choose categorical cross entropy (shown in Equation 9) as the loss function as it is widely used for multi-class classification problems [40].

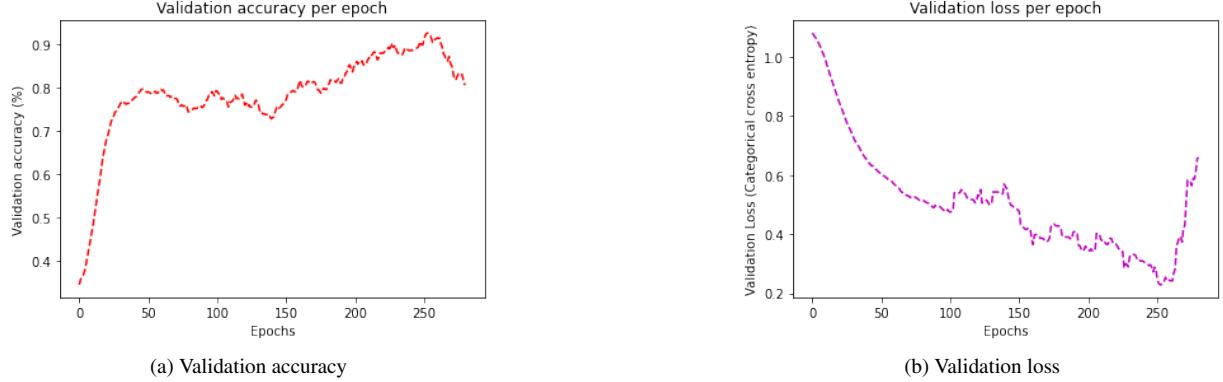


Figure 10: Validation accuracy and loss learning curves for the Brightfield+SCFI 10 frame model (E). Early stopping points are triggered to stop if the validation loss does not decrease beyond its current minimum with a tolerance of 30 epochs. The model will finish training, which happens around epoch 290, and the best model is chosen as the point at which loss is minimised.

In order to minimise overfitting, we implemented early stopping. Early stopping consists of calculating the loss function on the validation set each epoch as we train the model. Once there is no further improvements on our validation set we stop training. Early stopping points can be seen in Figures 10a and 10b.

Finally, the data was randomly shuffled. This was performed to ensure that the order of the data was representative of the overall distribution of the data.

3.2.2 Single frame models

3.2.2.1 Model A: Brightfield only

To act as a baseline to help evaluate the SCFI models, a Brightfield only model was created.

A CNN was chosen as the core architecture as it is a standard approach in high accuracy bacterial classification [41]. The model began as a basic image recognition structure with an input, conv2D, maxPooling, flatten and dense layers. A CNN input size was modified to 200x200 to accept a cropped Brightfield image from our pre-processing pipeline.

The network's hyperparameters (number of hidden layers, dropout, initial network weights, etc.) were tuned by trial and error. Throughout the model was optimised to maximise accuracy on the validation set.

We found that adding drop out layers between the dense layers improved the performance of the model on the validation set, with a dropout value of 0.4 found to be optimal.

In the output layer, class labels were one-hot encoded into [0,0,1], [0,1,0], [1,0,0], where a 1 in a particular position represents a class. One-hot encoding is used as it provides better re-scaling and avoids the model using rank as a predictive feature as the data is non-ordinal.

The final architecture of the model can be seen in Appendix A Figure 21.

3.2.2.2 Model B: SCFI 1st frame

A secondary model using only the first frame of the SCFI video was also created. The first frame was chosen as a consistent reference point for each video.

Initially, the architecture of model A was used as the image size of Brightfield and SCFI was the same. Again, hyperparameters were tuned via trial and error. Neural architecture search was also attempted and highlighted that adding spatial dropouts layers in addition to regular dropout increased accuracy. This is because regular dropouts randomly drops pixel values this could damage features being identified as adjacent pixels are highly related. Spatial dropout layers drop entire feature maps instead of individual elements. This is better as it ensures features stay intact.

The final architecture of the model can be seen in Appendix A Figure 22.

3.2.2.3 Model D: Brightfield and SCFI 1 frame dual channel

A final model was created using both Brightfield imagery as well as the first frame from the SCFI videos. The rational is that if one image source proves particularly discriminatory against a class and the other image source can better discriminate other classes. By combining both sources, a more robust model can be created.

As both Models A and B share the same input size they are combined in a concatenate layer. Trial and error was used to tune the models hyperparameters.

The final architecture of the model can be seen in Appendix A Figure 24.

3.2.3 Multi frame frame models

3.2.3.1 Model C: SCFI 10 frames

To evaluate the effectiveness of SCFI, a model was created to exploit the SCFI video data. Ten frames were selected from the beginning of the SCFI videos. Experiments were performed with different numbers of frames but ten was found to give high accuracy with lower training time. Frame placement was also experimented with, testing if frames should be consecutive or evenly spaced throughout the video. Videos were created of each strategy for a sample of the SCFI data in each class. Videos were manually evaluated and it was determined that consecutive frames displayed a clearer fluctuation pattern compared with evenly distributed frames. Frames were taken from the start of each SCFI video as a consistent reference point.

An LSTM was chosen as a base model as it has been shown to be successful in other bacterial classification studies [16]. Initially, we started with Model B but replaced layers with time distributed layers. Hyperparameters were tuned through a process of trial and error. The final architecture of the model can be seen in Appendix A Figure 23.

3.2.3.2 Model E: Brightfield and SCFI 10 frames dual channel

Model E is a dual channel approach designed to use both Brightfield and SCFI imagery. Similarly with model D, we expect this model to have higher accuracy due to more data to discriminate between classes.

Both models were combined in a concatenation layer as they share the same input size. Trial and error was used to tune the models hyperparameters. The final architecture of the model can be seen in Appendix A Figure 25.

3.3 Classical ML approach

A classical ML approach was also designed to aid interpretability of SCFI using manually engineered features. As such it was not trained on Brightfield data.

3.3.1 Feature engineering

For non neural network methods features were engineered from the dataset. This was important as unlike CNNs, decision trees do not perform well on raw pixel data. As such lower dimensional features are required. OpenCV [42] was selected as a computer vision library to extract these features as it is a primary library used in computer vision problems.

Decision trees cannot easily process sequential data. So four approaches were used to build features from videos:

- Mean pixel value across all frames, this feature represents the intensity of fluctuations.
- Standard deviation per pixel across all frames followed by mean pixel value across the result. The standard deviation compresses all frames into a single image that represents the spread of fluctuations.
- Bitwise AND across all frames followed by mean pixel value across the result. The Bitwise AND acts like a filter to remove noise. Only pixels with an intensity value during the entire sequence go through. By taking the mean value, a feature is created which represents the intensity of fluctuations for active regions.
- Fourier transform followed by mean pixel value. A Fourier transform transforms an image from the spatial domain into the frequency domain. This highlights strong frequencies in the image which should represent the sub-cellular fluctuations.

For images and single frames, three approaches were used to build features:

- Edge detection using Sobel filter [43] followed by mean pixel value of edges. Edge detection was tried to identify core structures which may aid classification. Edges were considered as a feature as they have successfully been used by other bacteria classification approaches [44].
- Region detection with watershed transformation [45] and isolating regions greater than 1000px, followed by mean pixel value in region. Region detection is a common process in bacteria classification [44] [46] [47]. By isolating only large regions we effectively focus on the foreground and exclude the background. This has an effect of denoising which helps to improve performance and increase discriminatory power by focusing on the fluctuations.
- Thresholding using every 10th percentile followed by mean pixel value. This also acts as a filter to denoise low intensity spots in the image.

$$x_{new} = \frac{x - \mu}{\sigma} \quad (2)$$

All features were scaled using a standard scalar transformation (SST), as seen in Equation 2. SST assumes features are normally distributed and scales them such that the distribution is centered around 0, with a standard deviation of 1. Scaling our features is important as some ML methods are sensitive to scaling. By scaling all features each is given an equal weight. Scaling is also critical in performing principle component analysis as it measures the variance of features against one another.

3.3.2 Feature importance

In order to evaluate which characteristics of SCFI images can be used for classification, we performed three methods to evaluate feature importance.

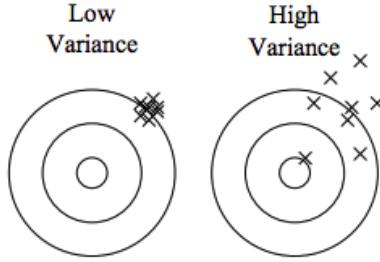


Figure 11: Example of how features with low variance provide no discriminatory power.

Initially, features with low variance were identified since they cannot help to discriminate between classes as seen in Figure 11.

Thereupon, a recursive feature ranking (RFE) was performed. Training the model with smaller and smaller sets of features determines the importance of each.

Then, a principal component analysis (PCA) was performed. PCA reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

3.3.3 Multi frame models

Two decision tree models were trained using all the features. Other machine learning models (such as Logistic regressions, K-nearest-neighbors and Histogram Gradient Boosting classifiers) were experimented with but found to not be accurate.

Decision trees were selected since they have been shown to effective in previous bacteria classification studies [48]. Further, they can be more easily interpretable as to how features are used to classify classes. Two types decision tree models were created: decision trees and random forests.

The scikit learn decision tree and random forest implementation were used [49]. The number of estimators was selected by examining the learning curve. A high number of estimators was used, meaning that many different specific models were tried. All models were 10-fold cross validated, and this was repeated and averaged three times to give high precision scores.

4 Results and Discussion

Table 1 summarises the performance of the best model in each category. The most important scores are test accuracy and F1 score. Results are discussed individually in more detail below.

Table 1: Performance summaries of best models in each category.

Best model	Test accuracy (%) ¹	Test error ²	Precision ³	Recall ³	F1 score ³
NN Single image ⁴	79.8	0.41	0.83	0.80	0.82
NN Video ⁵	96.6	0.08	0.97	0.97	0.97
Classical Video ⁶	64.6	1.1	0.47	0.46	0.47

¹ Mean test accuracy over all classes ² For Neural Networks, error is characterised by loss, specifically categorical-cross-entropy. For Classical ML, error is root mean square error (RMSE) ³ Mean over classes ⁴ Model A ⁵ Model E

⁶ Random Forest model. Only done for SCFI video data

4.1 Neural Network results

Test set accuracies and losses are shown in Table 2. Training accuracies for all models were on average proportionally 20% higher than test accuracies, indicating some level of overfitting. Precision and recall values by model and class are shown in Figure 12a and Figure 12b.

Table 2: Performance summaries of best Neural Network models.

Model	Test Accuracy (%) ¹	Test loss ¹
A: Brightfield only	79.8	0.41
B: SCFI 1st frame	69.7	0.94
C: SCFI 10 frames	53.8	1.08
D: Brightfield and SCFI 1 frame dual channel	92.4	0.23
E: Brightfield and SCFI 10 frames dual channel	96.6	0.08

¹ Mean across classes

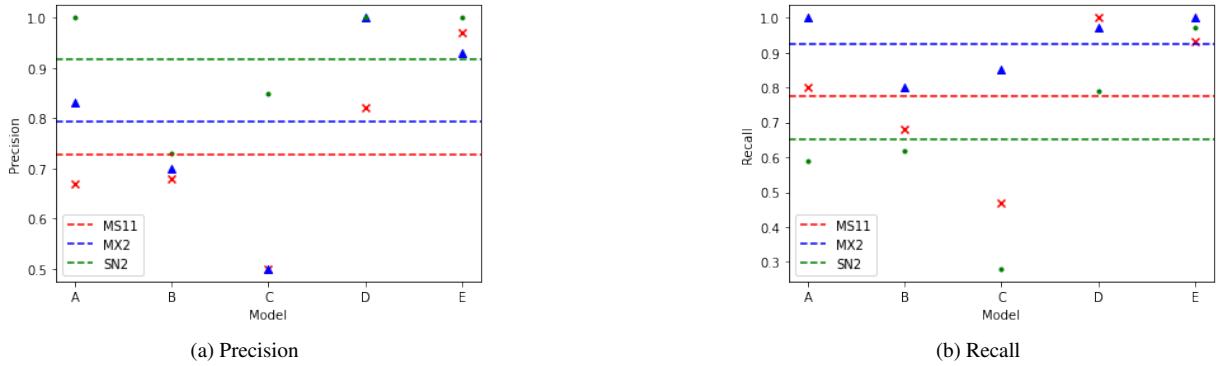


Figure 12: Precision and Recall rates broken down by model and class. Dotted lines indicate average values for each class across models. The SN2 class has highest precision and lowest recall and the MX2 class has the highest combined precision and recall (Table 3).

The Brightfield image only model (A) achieved high accuracy and low loss and establishes a good baseline to compare with SCFI. The SCFI 1st frame model (B) is the most directly comparable and had a 10.1% lower test classification accuracy.

The dual channel models performed the best with the Brightfield and SCFI combined LSTM model (E) obtaining the highest test accuracy. We can see that model E, which contains temporal data, is an improvement on model D, which does not. This was expected as the temporal data includes more information about the characteristic frequencies captured by SCFI.

However, we do not see the same pattern with the SCFI 1st frame (B) and SCFI LSTM (C). Model C includes ten frames compared with model B which contains only one. It was expected that additional frames would aid classification as more information about the characteristic frequencies is included. This indicates a possible methodical error and is further discussed in Section 4.1.3.

Table 3: Average F1 score across Neural Network models by class

Class	Average F1 score across models
MS11	0.75
MX2¹	0.85
SN2	0.76

¹ The MX2 class outperforms other classes in every model and on average

The average F1 scores across models are shown in Table 3. These scores show that the MX2 class was most accurately classified, with both MS11 and SN2 classes being classified approximately equally. This indicates that the MX2 class has more distinct and recognisable features than the other classes. From visual inspection of Figure 6 this may be explained by MX2 having brighter poles compared with MS11 and SN2.

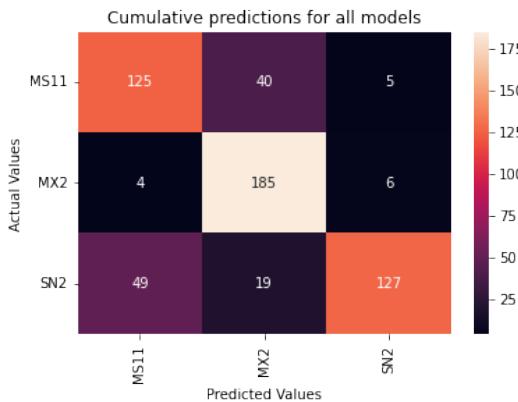


Figure 13: Cumulative confusion matrix made from predictions from all models.

A cumulative confusion matrix made of the sum of all predictions is shown in Figure 13. From this we can see the most misclassifications were SN2 as MS11 and MS11 as MX2.

4.1.1 Model A: Brightfield only

Table 4: Performance breakdown by class for Model A: Brightfield only

This model was designed to act as a baseline to compare SCFI classification accuracy against.

Class	Precision	Recall	F1-Score	Occurrences
MS11	0.67	0.80	0.73	40
MX2	0.83	1.00	0.91	40
SN2	1.00	0.59	0.74	39

The highest precision and lowest recall were for the SN2 class shown in Table 4. A precision of 1 means that every time a prediction was made as SN2 it was always SN2, and a recall of 0.59 means only 59% of SN2 samples were correctly classified.

This could indicate that images in the SN2 class can be split between two types, one which has a certain feature and one without. This feature is distinguishing for SN2 and is not present in other classes, and so when the model recognises the feature the image is classified as SN2 with 'certainty'. The model may have used this feature to solely predict whether a image was SN2 as it was present in over half the test data. This idea is backed up by the

class probability values in the final dense layer where all predictions made as SN2 are low variance, meaning the model is near certain an image is SN2 when it has classified it as such.

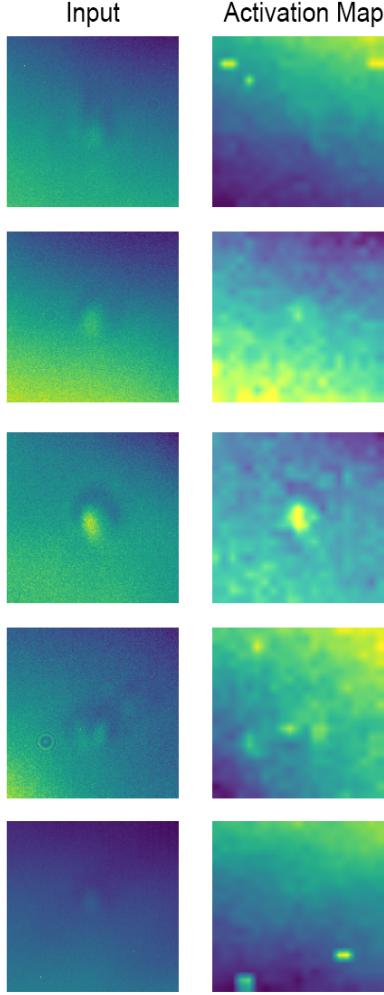


Figure 14: Activation maps for a sample of images from SN2 class of brightfield model A. Lightest regions indicate pixels which the model paid most attention to in determining it's classification. Activation maps were produced by GradCAM.

When observing Brightfield examples (see Figure 26) we can see clear differences between SN2 and the other classes in that the bacteria appear much fainter. This could be the feature detected, so that when a model sees a faint bacteria it determines class as SN2, giving a high precision. A minority proportion of SN2 images have more visible/brighter bacteria so the model could be making a decision that this feature normally indicates SN2. The more visible examples would be harder to tell apart from particularly the MS11 examples, and so would cause a low recall rate. In the vast majority of samples, the MX2 bacteria has a brighter central spot and a darker, more uniform background which make it stand out over the other two classes, which are more similar. These more distinct features are easier for the model to figure out which would explain the high F1 score. The activation maps in Figure 14 support this conclusion. An activation map highlights which parts of the image the CNN used to determine it's classification. Activations for SN2 pay attention heavily to the background and when samples of a bright central spot appear (e.g. furthest left) they are misclassified.

Overall, the model is better at distinguishing the MX2 class from the MS11 and SN2 classes, indicated by the high

F1 score and the similarities in the MS11 and SN2 scores.

4.1.2 Model B: SCFI 1st frame

Table 5: Performance breakdown by class for Model B: SCFI 1st frame

Class	Precision	Recall	F1-Score	Occurrences
MS11	0.68	0.68	0.68	40
MX2	0.70	0.80	0.74	40
SN2	0.73	0.62	0.67	39

Model B was designed to determine if a single frame of SCFI provides enough information to classify a bacteria.

Model B is less accurate overall than model A. This suggests that Brightfield images contain more discriminatory information than a single frame from a SCFI video. We can note from Table 5 that compared to model A:

- F1 scores are lower in all classes, but particularly much lower in MX2 class.
- MS11 precision is similar, MX2, SN2 precision is significantly lower.

These results indicate that model B was unable to extract strongly discriminatory features. On visual inspection of SCFI data samples in Figure 6, differences between classes in a single SCFI frame appear less pronounced compared with Brightfield. This result is expected as a single frame of a SCFI video does not contain the sub cellular fluctuations which we expect to be the main discriminatory feature when classifying bacteria.

4.1.3 Model C: SCFI 10 frames

Table 6: Performance breakdown by class for Model C: SCFI 10 frames

Class	Precision	Recall	F1-Score	Occurrences
MS11	0.50	0.47	0.49	40
MX2	0.50	0.85	0.63	40
SN2	0.85	0.28	0.42	39

Model C was designed to determine if multiple frames of SCFI imagery provides enough information to classify bacterial classes and to compare accuracy to brightfield imagery.

The LSTM ten frame model for SCFI data had the worst performance. In Table 6 we observe that the model had an improvement in precision for SN2 but a very low recall. This indicates that the model returned few results but was mostly correct when it did. Conversely, a higher recall for MX2 was seen, implying that MX2 may have a larger temporal component than the other classes.

We observed the worst performance for model C compared with other models. This was a surprising result as we expected the fluctuations to be a discriminatory feature and visible over the ten starting frames. We speculate that this result indicates an issue with the model architecture. Possible issues include:

- Hyperparameters were poorly selected.
- Layer structure poorly chosen.
- An undiscovered technical error.

Additional attention is required to address this issue and we recommend further study to understand this fully.

4.1.4 Model D: Brightfield and SCFI 1 frame dual channel

Table 7: Performance breakdown by class for Model D: Brightfield and SCFI 1 frame dual channel

Class	Precision	Recall	F1-Score	Occurrences
MS11	0.82	1.00	0.90	40
MX2	1.00	0.97	0.99	40
SN2	1.00	0.79	0.89	39

The Brightfield and SCFI 1 frame model had a high accuracy of 92.4%. MX2 had near perfect classification accuracy with only one MX2 example being misclassified as MS11 as seen in the confusion matrix in the Appendix B Figure 30.

The accuracy of this combined method is higher than both the individual Brightfield images and SCFI single frame models. This indicates that the model is taking different features from both, as the increased accuracy must come by using the discriminatory power of features from both Brightfield and SCFI images. This confirms that SCFI captures different aspects of bacteria compared to Brightfield imaging.

4.1.5 Model E: Brightfield and SCFI 10 frames dual channel

Table 8: Performance breakdown by class for Model E: Brightfield and SCFI 10 frames dual channel

Class	Precision	Recall	F1-Score	Occurrences
MS11	0.97	0.93	0.95	40
MX2	0.93	1.00	0.96	40
SN2	1.00	0.97	0.99	39

Model E was designed to determine if using features from both Brightfield and ten frame SCFI aided classification. Model E had the highest test set accuracy of 96.6% compared with all other models. This was expected as the model was given more data than any of the other models, and therefore had more features to use to discriminate between classes.

The LSTM outperformed model C (SCFI ten frame model) and model D. This further indicates methodical error was introduced during model C training. If this was not the case we would have expected model D to outperform model E.

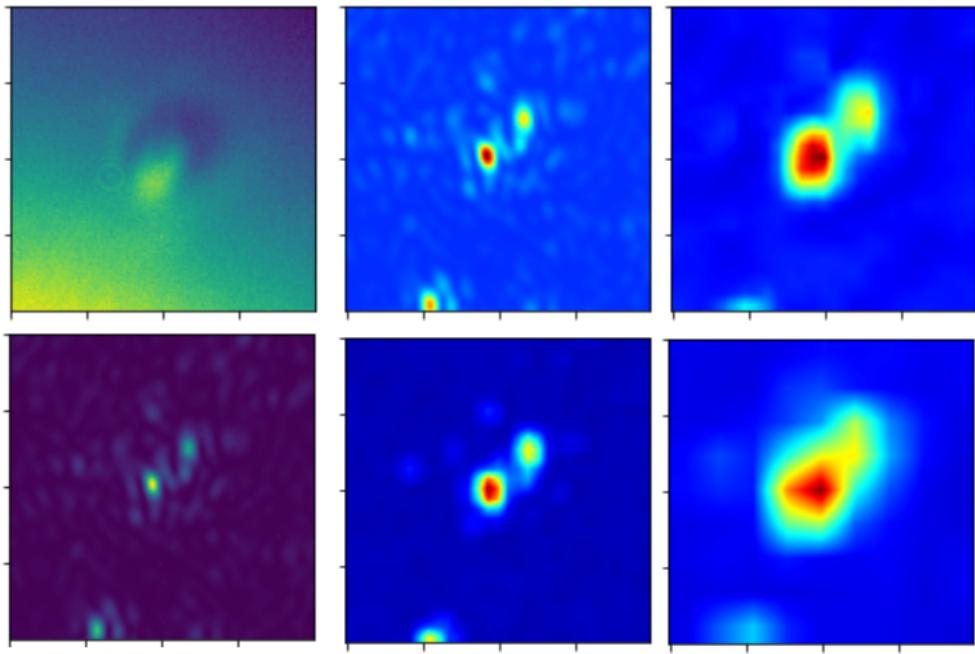


Figure 15: Activations maps at different training intervals for model E produced by GradCAM. Top Left: Brightfield image Bottom Left: example SCFI frame from a sample. Middle: Activation map after 0 (top) and 30 (bottom) epochs. Right: Activation map after 60 (top) and 90 (bottom) epochs.

Activation maps showing which regions of the input carry the highest weights are shown in Figure 15. As training time progresses the model pays highest attention to a region approximately the size of the Brightfield image. The model also pays less attention to the background as time progresses, as can be seen by the diminishing importance of the region at the bottom of the frame. This is a significant result as it shows the network is using the core of the image rather than potential background artefacts, to discriminate between classes.

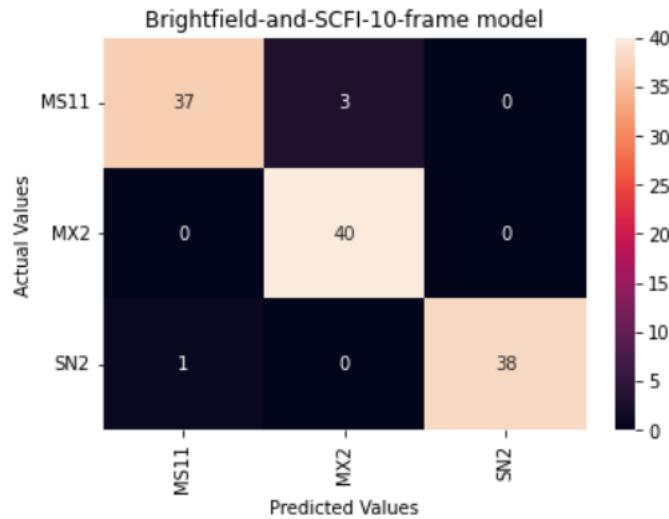


Figure 16: Confusion matrix for model E.

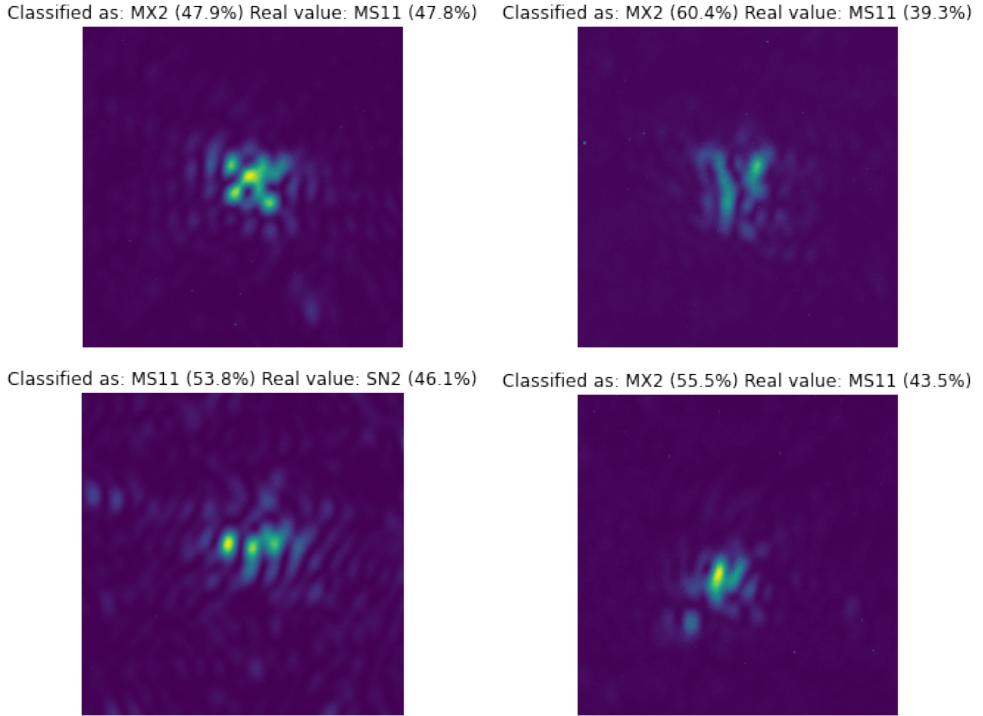


Figure 17: Misclassifications for model E with model's certainty in brackets.

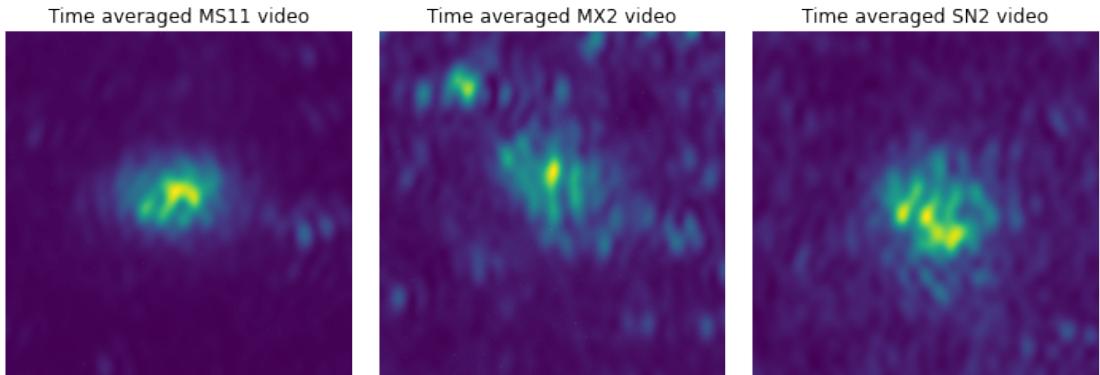


Figure 18: Time averaged videos by class indicating the regions with highest pixel values across videos. These were created for each class by averaging all SCFI frames for that class. Brighter regions indicate where pixel values change most significantly. MS11 videos fluctuate strongly in the center of the image with a sharp drop off and minimal fluctuation outside this central pattern. MX2 videos have a much wider spread of fluctuations, with a small central point. SN2 videos have a large and bright central region but also exhibit the wide spread of fluctuations seen in MX2. These help highlight potential features used by the CNN to classify bacteria.

The confusion matrix Figure 16 highlights that only four samples were misclassified. MS11 was misclassified as MX2 three times, these misclassifications being shown in Figure 17. To evaluate why these misclassifications occurred we visually compared both the SCFI and Brightfield input images to the average class images, with average class images for SCFI being shown in Figure 18. From visual inspection we are unable to fully determine why these misclassifications occurred. However, we hypothesize that the model incorrectly identified two poles in the image and therefore classified as MX2 rather than SN2.

Misclassifications such as these highlight a limitation of our model as we lack the power to fully interpret cases

when the model failed. Interpretability is essential in medical settings as since it helps to highlight unconscious bias in models. Section 4.2 aims to aid interpretability using manually engineered features to give us insight into the benefits and limitations of using SCFI imagery for bacterial classification.

4.2 Classical machine learning results

In the classical ML approach, two separate decision trees (Decision trees and Random Forests) were trained on engineered features. The top performing model was a Random Forest which achieved an accuracy of 64.6% when tested on the testing set.

This result is expected as a Random Forest is an ensemble technique and is widely recognised that ensembling improves the performance of models.

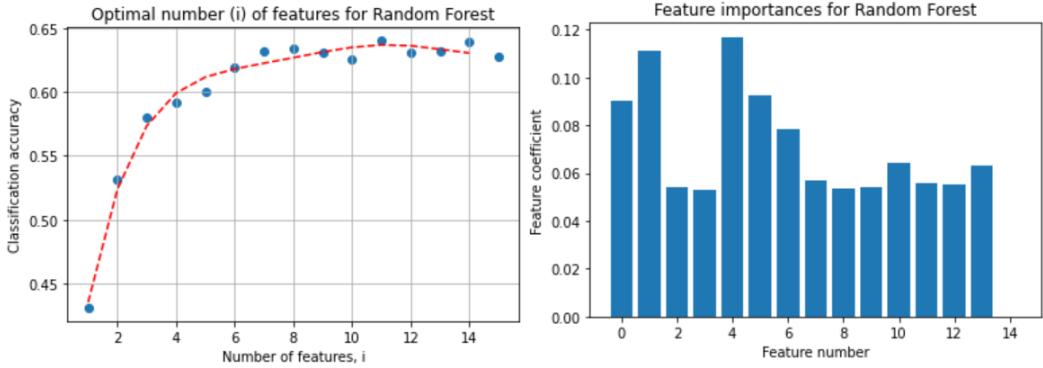


Figure 19: Final random forest feature importance. The final model used 11 of the 15 features for classification.

The final random forest selected 11 of the 15 features as important, as shown in Figure 19. It highlights that three features accounted for most of the variance. These were pixel-by-pixel variance, mean pixel value of image per frame, and total pixel variance per frame. Analysing these one by one can build an understanding of what information SCFI imagery can provide to classify bacteria.

- Pixel-by-pixel variance: this feature captured how widely fluctuations varied over time. As this feature can be used to discriminate between bacteria this indicates that fluctuations are a characteristic feature of the bacteria strains studied. This aligns with the results of existing studies [25].
- Mean pixel value per frame: this feature captures the strength of fluctuations. Again this correlates with existing literature [25] suggesting that the strength and frequency of fluctuations are characteristic to bacterial strains.
- Total pixel variance per frame: this is similar to mean pixel value but captures the range of fluctuations per frame.

The other features had a lower importance and as such had lower discriminatory power. By analysing these we can understand the strengths and weaknesses of SCFI imagery when used for classification:

- Edge detection: as this feature did not discriminate well, it indicates that the strength and frequency of fluctuations matter more than their geometric structure.
- Fourier transform: it was expected that a Fourier transform may aid classification as it captures frequency information about the fluctuations. However, we found it hard to isolate key features in the Fourier transform image which resulted in noise from additional frequencies such as optical artefacts.

- Region detection, thresholding, and bitwise AND: these features were expected to help isolate the poles of the fluctuations. We found these features isolated the poles well in some imagery but failed in others and obscured the poles as such they damaged the main discriminatory features and proved to be poor features themselves.

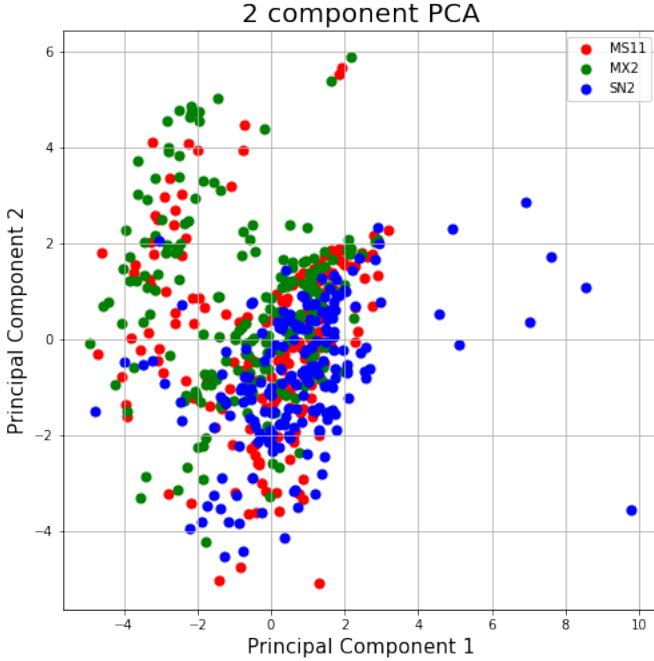


Figure 20: A 2 component PCA accounted for a 0.49 variance (0.26 and 0.23 respectively) in the data. There is significant visual separation between the MX2 class and the other classes. However, there is no clear separation between SN2 and MS11 classes.

A random forest was also trained on a two component principle component analysis shown in Figure 20. It produced an accuracy of 0.49, less than the model trained on the raw features. This is expected, as by visual inspection of the two principle components in Figure 20 there is significant visual separation between the MX2 class and the other classes. However, there is no clear separation between SN2 and MS11 classes.

5 Conclusions and Further Research

It was theorised that sub-cellular fluctuations imaged by SCFI would have enough discriminatory power to classify the three bacterial species. Our best model achieved 69.7% accuracy when only using SCFI imagery and a 96.6% accuracy when combined with brightfield imagery. We therefore find that SCFI with the addition of Brightfield imagery contains distinct information which could be used for bacterial classification.

We developed a deep learning approach using a LSTM and CNN architecture resulting in an 96.6% accuracy. This significantly outperformed the classical ML approach inline with what is seen by other bacterial classification techniques [11] [10]. We have shown the potential of this technique for classification and recommend further research in denoising techniques such as neural denoising [50]. This will help reduce systematic errors in data which the model may be using for classification. Further, as the dataset used was relatively small we also recommend further research in using other architectures better suited to smaller datasets, aswell as the collection of more data. Gated recurrent units (GRU) have been shown to offer improved performance over LSTMs with smaller datasets [51].

We found a surprising result in that model C, an LSTM trained on the first ten frames of SCFI, performed worse than expected. We expected the model to outperform model B which was trained on only a single SCFI frame as ten frames contained fluctuation information. We deduce that this was due to an incorrect selection of hyperparameters. This is a limitation of our research and further research is required to address this issue.

We investigated the benefits and limitations of SCFI at providing discriminatory features by building a classical ML model with engineered features to provide greater interpretability. We found that three features provided much of the variance for classification. These features indicated that SCFI strengths were in mapping power and frequency differences between bacteria. However, we found that structural difference on where the poles were located did not provide discriminatory power in our analysis. In the dataset used, these features achieved an accuracy of 64.6% over the three bacterial strains. In order to generalise this conclusion a wider dataset containing more species is required to understand whether these features can discriminate between other bacteria strains.

We compared the accuracy of classification based on SCFI and Brightfield data and found that a Brightfield-only model produced higher accuracy over a SCFI-only model. However, by combining both techniques a model with 96.6% accuracy was achieved, based on the relative ease of capturing both SCFI and Brightfield imagery in the experimental methods described in Section 3.1.1. This supports the conclusion that SCFI imagery contains distinct information which could be used for bacterial classification.

Appendices

A Appendix A: Model architecture diagrams

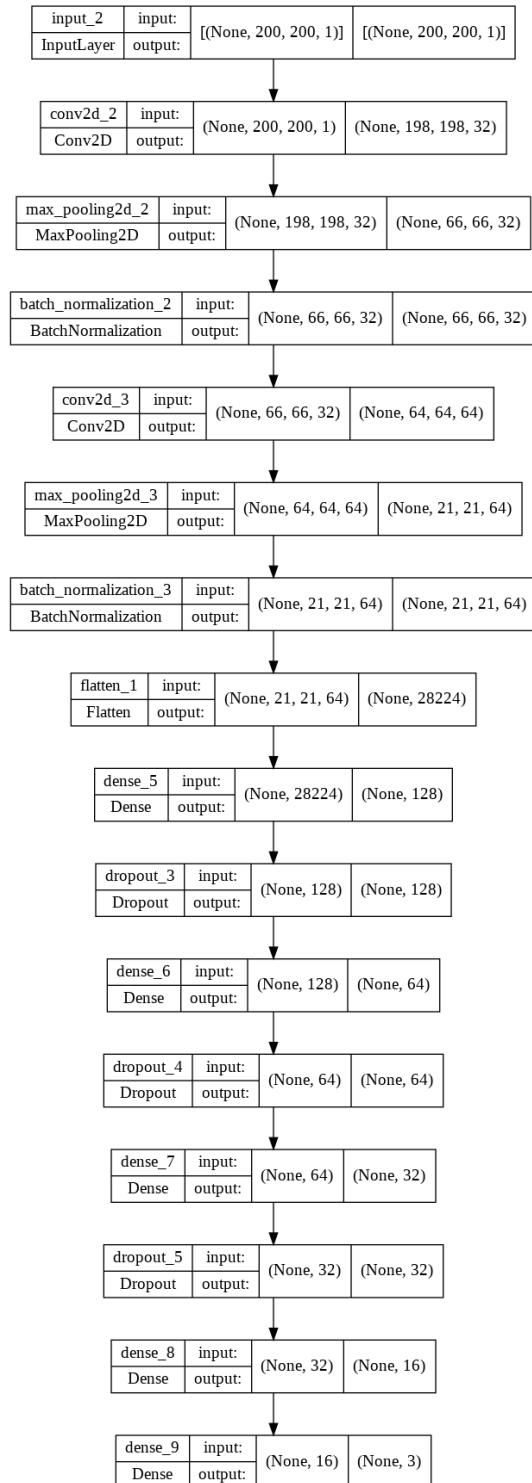


Figure 21: Neural network architecture diagram for Model A: Brightfield only

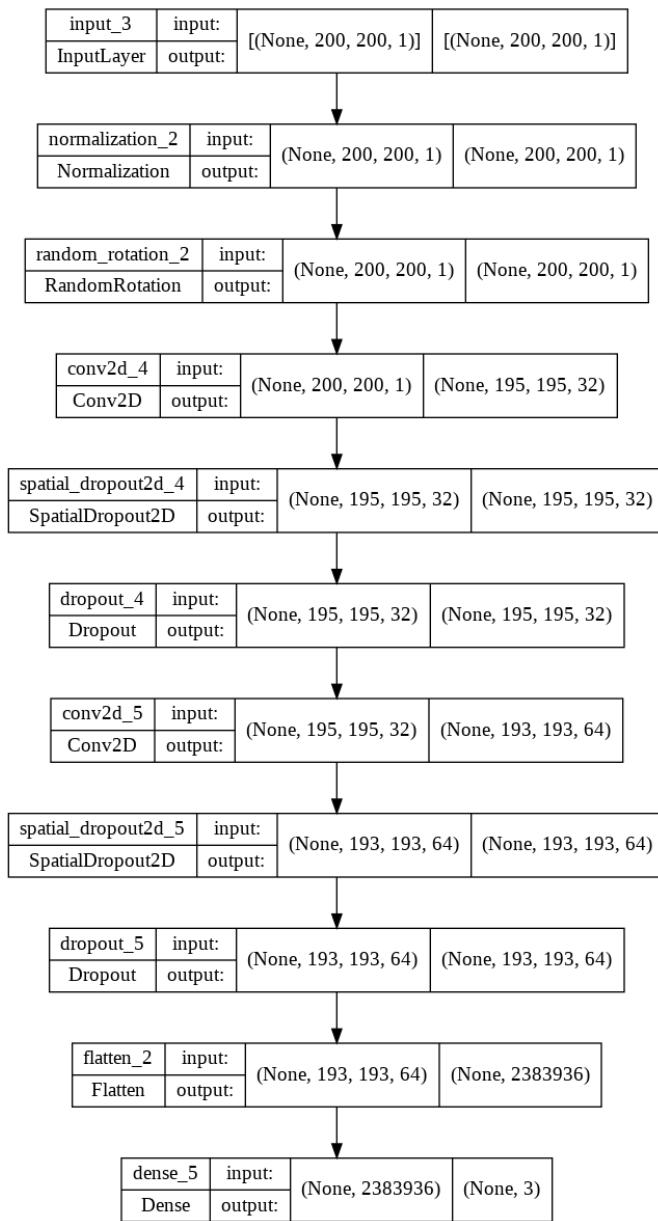


Figure 22: Neural network architecture diagram for Model B: SCFI 1st frame

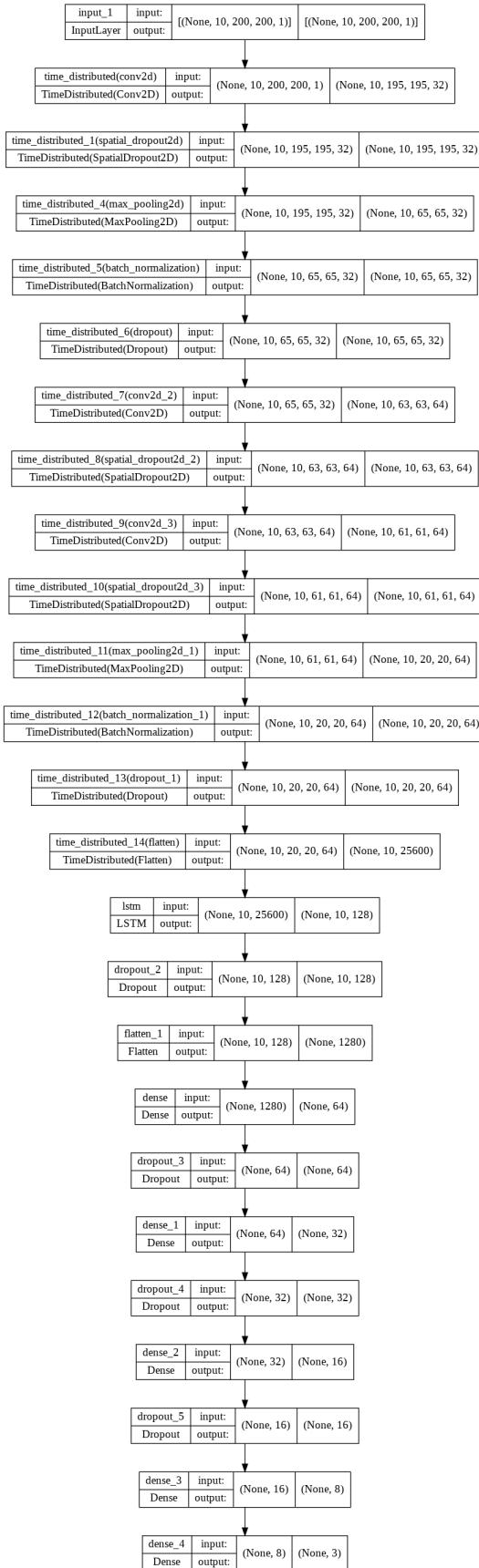


Figure 23: Neural network architecture diagram for Model C: SCFI 10 frames

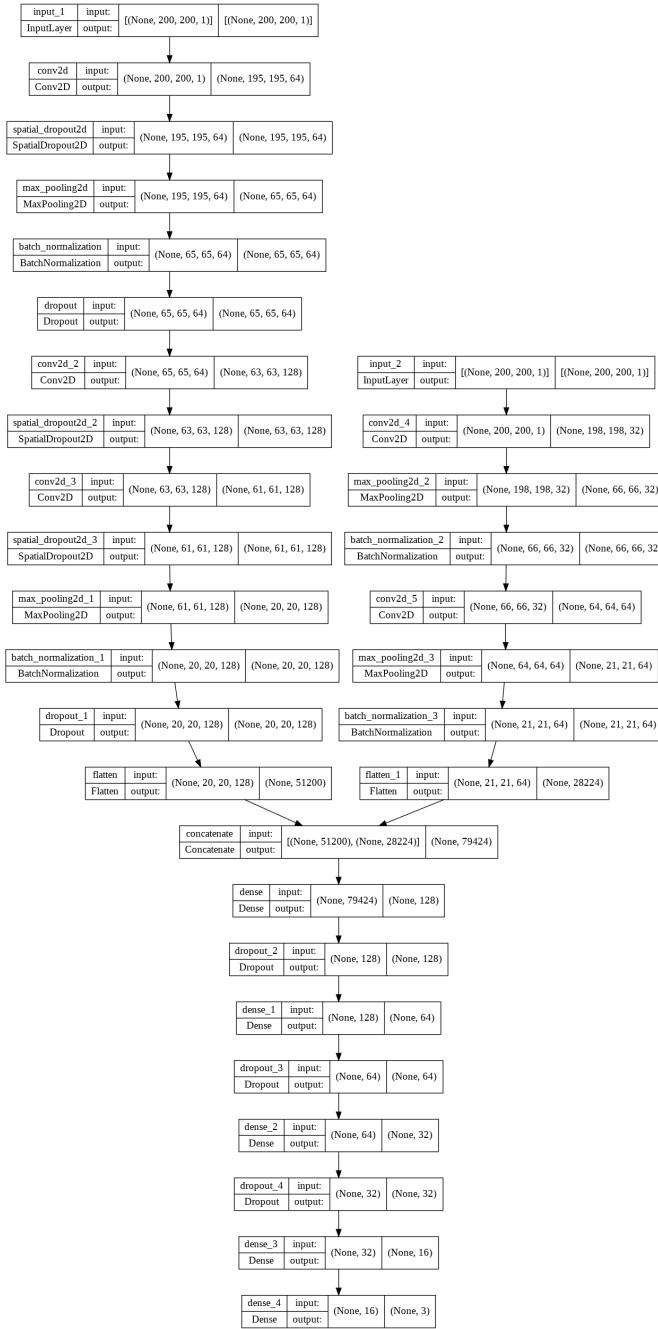


Figure 24: Neural network architecture diagram for Model D: Brightfield and SCFI 1 frame dual channel

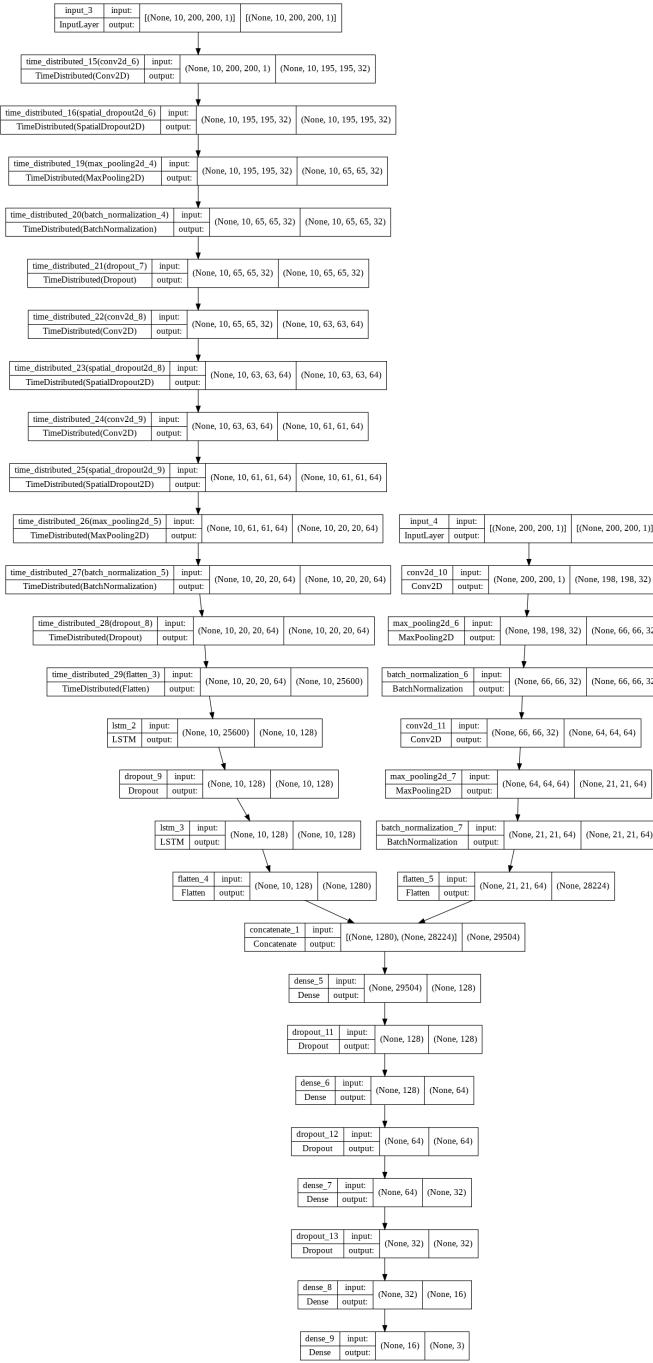


Figure 25: Neural network architecture diagram for Model E: Brightfield and SCFI 10 frames dual channel

B Appendix B: Model results

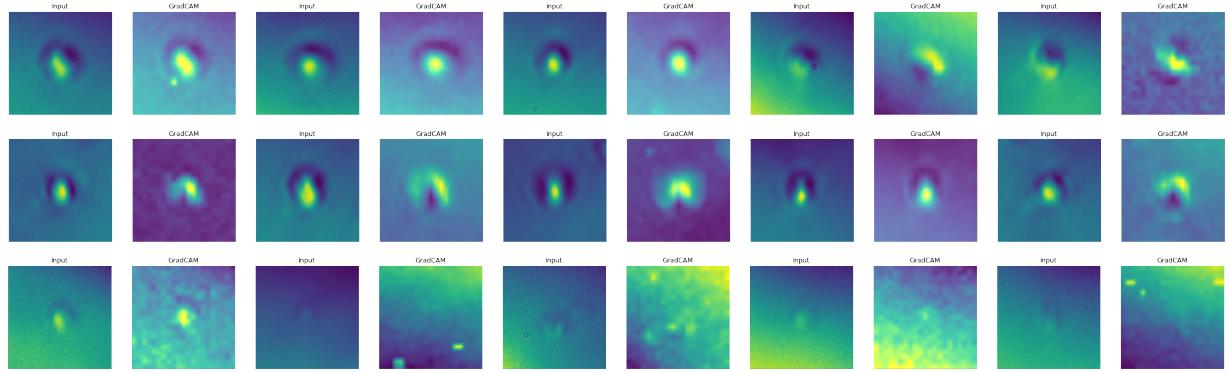


Figure 26: Activation maps for a sample of images from each class of brightfield model A. From top to bottom: MS11, MX2, and SN2. Lightest regions indicate pixels which the model paid most attention to in determining it's classification. Activation maps were produced by GradCAM.

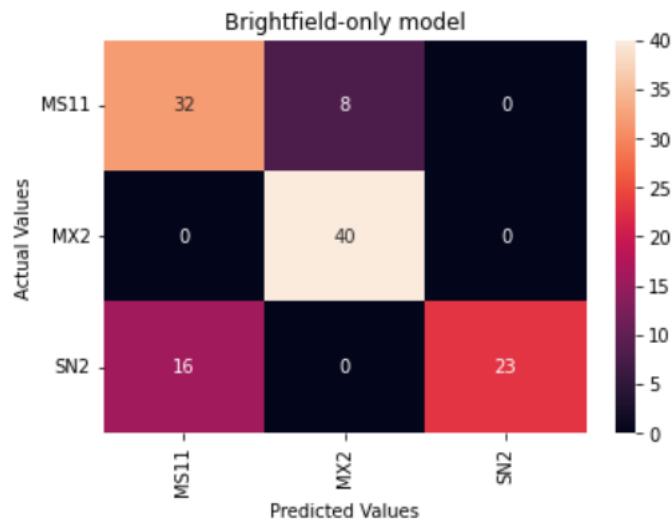


Figure 27: Confusion matrix for brightfield model A

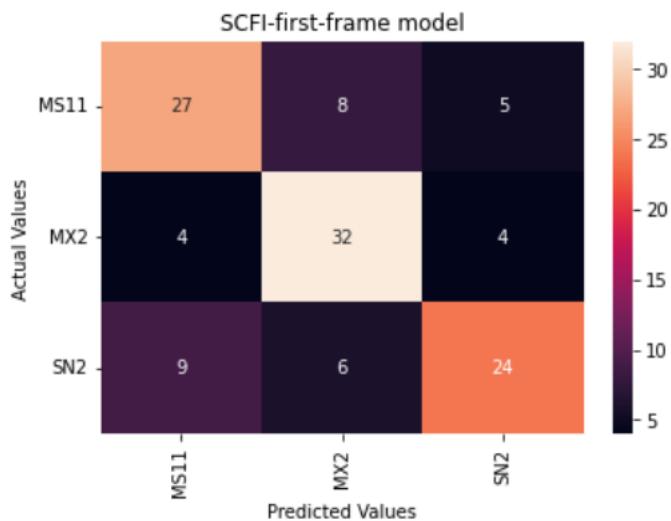


Figure 28: Confusion matrix for single frame SCFI model B

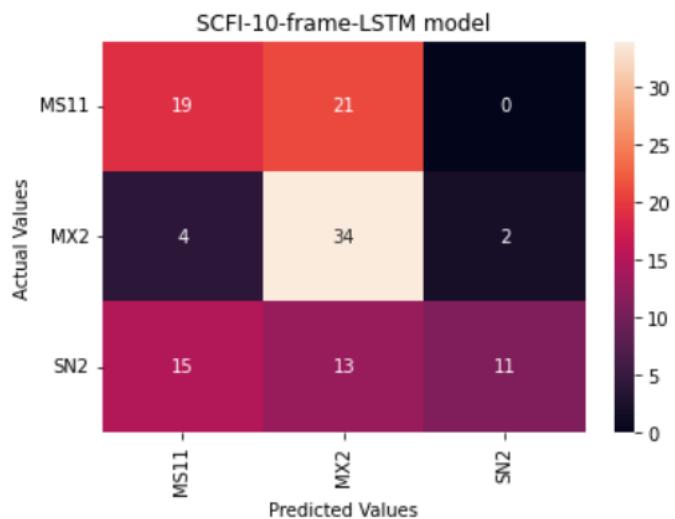


Figure 29: Confusion matrix for 10 frame SCFI model C



Figure 30: Structure

References

- [1] Stephen P Oliver, Bhushan M Jayarao, and Raul A Almeida. “Foodborne pathogens in milk and the dairy farm environment: food safety and public health implications”. In: *Foodborne Pathogens & Disease* 2.2 (2005), pp. 115–129.
 - [2] Ronald H Schmidt and Gary E Rodrick. *Food safety handbook*. John Wiley & Sons, 2003.
 - [3] Florian P. Maurer et al. “Advances in Rapid Identification and Susceptibility Testing of Bacteria in the Clinical Microbiology Laboratory: Implications for Patient Care and Antimicrobial Stewardship Programs”. eng. In: *Infectious disease reports* 9.1 (Mar. 2017). PMC5391540[pmcid], pp. 6839–6839. ISSN: 2036-7430. DOI: 10.4081/idr.2017.6839. URL: <https://doi.org/10.4081/idr.2017.6839>.
 - [4] Jenny Göransson et al. “Rapid Identification of Bio-Molecules Applied for Detection of Biosecurity Agents Using Rolling Circle Amplification”. In: *PLOS ONE* 7.2 (Feb. 2012), pp. 1–9. DOI: 10.1371/journal.pone.0031068. URL: <https://doi.org/10.1371/journal.pone.0031068>.
 - [5] Rodrigo DeAntonio et al. “Epidemiology of community-acquired pneumonia and implications for vaccination of children living in developing and newly industrialized countries: A systematic literature review”. In: *Human vaccines & immunotherapeutics* 12.9 (2016), pp. 2422–2440.
 - [6] F Christiaan K Dolk et al. “Antibiotics in primary care in England: which antibiotics are prescribed and for which conditions?” In: *Journal of Antimicrobial Chemotherapy* 73.suppl_2 (Feb. 2018), pp. ii2–ii10. ISSN: 0305-7453. DOI: 10.1093/jac/dkx504. eprint: https://academic.oup.com/jac/article-pdf/73/suppl_2/ii2/25088019/dkx504.pdf. URL: <https://doi.org/10.1093/jac/dkx504>.
 - [7] C. Lee Ventola. “The antibiotic resistance crisis: part 1: causes and threats”. eng. In: *P & T : a peer-reviewed journal for formulary management* 40.4 (Apr. 2015). PMC4378521[pmcid], pp. 277–283. ISSN: 1052-1372. URL: <https://pubmed.ncbi.nlm.nih.gov/25859123/>.
 - [8] Hongda Wang et al. “Early detection and classification of live bacteria using time-lapse coherent imaging and deep learning”. In: *Light: Science Applications* 9.1 (July 2020). DOI: 10.1038/s41377-020-00358-9. URL: <https://www.nature.com/articles/s41377-020-00358-9>.

- [9] Fenglei Fan et al. “On Interpretability of Artificial Neural Networks: A Survey”. In: *arXiv.org* (2020). DOI: 10.48550/arXiv.2001.02522. URL: <https://arxiv.org/abs/2001.02522>.
- [10] Liang Wang et al. “Applications of Raman Spectroscopy in Bacterial Infections: Principles, Advantages, and Shortcomings”. In: *Frontiers in Microbiology* 12 (2021). ISSN: 1664-302X. DOI: 10.3389/fmicb.2021.683580. URL: <https://www.frontiersin.org/article/10.3389/fmicb.2021.683580>.
- [11] Neelja Singhal et al. “MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis”. In: *Frontiers in Microbiology* 6 (2015). ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00791. URL: <https://www.frontiersin.org/article/10.3389/fmicb.2015.00791>.
- [12] Sascha Sauer and Magdalena Kliem. “Mass spectrometry tools for the classification and identification of bacteria”. In: *Nature Reviews Microbiology* 8.1 (Jan. 2010), pp. 74–82. DOI: 10.1038/nrmicro2243. URL: <https://pubmed.ncbi.nlm.nih.gov/20010952/>.
- [13] Walter Florio et al. “Recent Advances and Ongoing Challenges in the Diagnosis of Microbial Infections by MALDI-TOF Mass Spectrometry”. In: *Frontiers in Microbiology* 9 (May 2018). DOI: 10.3389/fmicb.2018.01097. URL: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01097/full>.
- [14] Charlotte R Birmingham et al. “Imaging of sub-cellular fluctuations provides a rapid way to observe bacterial viability and response to antibiotics”. In: *bioRxiv* (2018). DOI: 10.1101/460139. eprint: <https://www.biorxiv.org/content/early/2018/11/02/460139.full.pdf>. URL: <https://www.biorxiv.org/content/early/2018/11/02/460139>.
- [15] William Wiley Navarre and Olaf Schneewind. “Surface Proteins of Gram-Positive Bacteria and Mechanisms of Their Targeting to the Cell Wall Envelope”. In: *Microbiology and Molecular Biology Reviews* 63.1 (Mar. 1999), pp. 174–229. DOI: 10.1128/mmbr.63.1.174-229.1999. URL: <https://pubmed.ncbi.nlm.nih.gov/10066836/>.
- [16] Shallu Kotwal et al. “Automated Bacterial Classifications Using Machine Learning Based Computational Techniques: Architectures, Challenges and Open Research Issues”. In: *Archives of Computational Methods in Engineering* (Oct. 2021). DOI: 10.1007/s11831-021-09660-0. URL: <https://link.springer.com/article/10.1007/s11831-021-09660-0>.
- [17] Rani Oomman Panicker et al. “Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods”. In: *Biocybernetics and Biomedical Engineering* 38.3 (2018), pp. 691–699. DOI: 10.1016/j.bbe.2018.05.007. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0208521618301311?via%3Dihub>.
- [18] Mingxing Zhang et al. “Non-invasive single-cell morphometry in living bacterial biofilms”. In: *Nature Communications* 11.1 (Dec. 2020). DOI: 10.1038/s41467-020-19866-8. URL: <https://www.nature.com/articles/s41467-020-19866-8>.
- [19] Monya Baker. “Taking a long, hard look”. In: *Nature* 466.7310 (Aug. 2010), pp. 1137–1138. DOI: 10.1038/4661137a. URL: <https://ui.adsabs.harvard.edu/abs/2010Natur.466.1137B/abstract>.
- [20] Nan Meng, Hayden K. H. So, and Edmund Y. Lam. “Computational single-cell classification using deep learning on bright-field and phase images”. In: *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)* (May 2017). DOI: 10.23919/mva.2017.7986833. URL: <https://ieeexplore.ieee.org/document/7986833>.

- [21] Christoph Spahn et al. “DeepBacs: Bacterial image analysis using open-source deep learning approaches”. In: (Nov. 2021). DOI: 10.1101/2021.11.03.467152. URL: <https://www.biorxiv.org/content/10.1101/2021.11.03.467152v1.full>.
- [22] unknown. *Figure 6: Principle of a bright field microscope (a), a phase-contrast...* Feb. 2019. URL: https://www.researchgate.net/figure/Principle-of-a-bright-field-microscope-a-a-phase-contrast-microscope-b-and-a_fig5_333825066.
- [23] 2022. URL: <https://micro.magnet.fsu.edu/optics/olympusmicd/anatomy/micdbrightfield.html>.
- [24] Ganna Platonova et al. “Spectroscopic Approach to Correction and Visualisation of Bright-Field Light Transmission Microscopy Biological Data”. In: *Photonics* 8.8 (Aug. 2021), p. 333. DOI: 10.3390/photonics8080333. URL: <https://arxiv.org/abs/1903.06519>.
- [25] H. Hindley. “Developing a convolutional neural network based on the SCFI method to perform rapid recognition of bacterial species”. In: (2021).
- [26] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [27] Md Zahangir Alom et al. “A State-of-the-Art Survey on Deep Learning Theory and Architectures”. In: *Electronics* 8.3 (Mar. 2019), p. 292. ISSN: 2079-9292. DOI: 10.3390/electronics8030292. URL: <http://dx.doi.org/10.3390/electronics8030292>.
- [28] Ulf Johansson et al. “Trade-off between accuracy and interpretability for predictive in silico modeling”. In: *Future Medicinal Chemistry* 3.6 (2011). PMID: 21554073, pp. 647–663. DOI: 10.4155/fmc.11.23. eprint: <https://doi.org/10.4155/fmc.11.23>. URL: <https://doi.org/10.4155/fmc.11.23>.
- [29] Yann Lecun. *A Theoretical Framework for Back-Propagation*. Aug. 2001. URL: https://www.researchgate.net/publication/2360531_A_Theoretical_Framework_for_Back-Propagation.
- [30] Yaoyao Liu et al. “Rapid and accurate identification of marine microbes with single-cell Raman spectroscopy”. In: *Analyst* 145 (9 2020), pp. 3297–3305. DOI: 10.1039/C9AN02069A. URL: <http://dx.doi.org/10.1039/C9AN02069A>.
- [31] Ying Li et al. “Application of MALDI-TOF MS to rapid identification of anaerobic bacteria”. In: *BMC INFECTIOUS DISEASES* 19.1 (Nov. 2019). DOI: 10.1186/s12879-019-4584-0.
- [32] Thomas Mortier et al. “Bacterial species identification using MALDI-TOF mass spectrometry and machine learning techniques: A large-scale benchmarking study”. In: *Computational and Structural Biotechnology Journal* 19 (2021), pp. 6157–6168. ISSN: 2001-0370. DOI: <https://doi.org/10.1016/j.csbj.2021.11.004>. URL: <https://www.sciencedirect.com/science/article/pii/S2001037021004694>.
- [33] H. Yu et al. “Phenotypic Antimicrobial Susceptibility Testing with Deep Learning Video Microscopy”. In: *Anal Chem* 90.10 (May 2018), pp. 6314–6322.
- [34] R. Iriya et al. “Rapid antibiotic susceptibility testing based on bacterial motion patterns with long short-term memory neural networks”. In: *IEEE Sens J* 20.9 (May 2020), pp. 4940–4950.
- [35] Y. Bengio, P. Simard, and P. Frasconi. “Learning long-term dependencies with gradient descent is difficult”. In: *IEEE Transactions on Neural Networks* 5.2 (Mar. 1994), pp. 157–166. DOI: 10.1109/72.279181. URL: <https://ieeexplore.ieee.org/document/279181>.

- [36] Junyoung Chung et al. “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”. In: *arXiv.org* (2014). DOI: 10.48550/arXiv.1412.3555. URL: <https://arxiv.org/abs/1412.3555>.
- [37] Parth Sane and Ravindra Agrawal. “Pixel Normalization from Numeric Data as Input to Neural Networks”. In: *arXiv.org* (2017). DOI: 10.48550/arXiv.1705.01809. URL: <https://arxiv.org/abs/1705.01809>.
- [38] Yun Xu and Royston Goodacre. “On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning”. In: *Journal of Analysis and Testing* 2.3 (July 2018), pp. 249–262. DOI: 10.1007/s41664-018-0068-2. URL: <https://link.springer.com/article/10.1007/s41664-018-0068-2>.
- [39] Borislava Toleva. *The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems*. May 2021. URL: https://www.researchgate.net/publication/351939082_The_Proportion_for_Splitting_Data_into_Training_and_Test_Set_for_the_Bootstrap_in_Classification_Problems.
- [40] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 1st ed. Springer, 2007. ISBN: 0387310738. URL: <http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738%3FSubscriptionId%3D13CT5CVB80YFWJEPWS26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0387310738>.
- [41] Tumun Shaily and S Kala. “Bacterial Image Classification Using Convolutional Neural Networks”. In: *2020 IEEE 17th India Council International Conference (INDICON)* (Dec. 2020). DOI: 10.1109/indicon49873.2020.9342356. URL: <https://ieeexplore.ieee.org/document/9342356>.
- [42] Apr. 2022. URL: <https://opencv.org/>.
- [43] N. Kanopoulos, N. Vasanthavada, and R.L. Baker. “Design of an image edge detection filter using the Sobel operator”. In: *IEEE Journal of Solid-State Circuits* 23.2 (Apr. 1988), pp. 358–367. DOI: 10.1109/4.996. URL: <https://ieeexplore.ieee.org/document/996/>.
- [44] Marie-Pierre Dubuisson, Anil K. Jain, and Mahendra K. Jain. “Segmentation and classification of bacterial culture images”. In: *Journal of Microbiological Methods* 19.4 (Apr. 1994), pp. 279–295. DOI: 10.1016/0167-7012(94)90031-0. URL: <https://www.sciencedirect.com/science/article/abs/pii/0167701294900310>.
- [45] Serge Beucher and Centre De. *The Watershed Transformation Applied To Image Segmentation*. July 2000. URL: https://www.researchgate.net/publication/2407235_The_Watershed_Transformation_Applied_To_Image_Segmentation.
- [46] Carlos Garcia-Perez et al. “Efficient Detection of Longitudinal Bacteria Fission Using Transfer Learning in Deep Neural Networks”. In: *Frontiers in Microbiology* 12 (June 2021). DOI: 10.3389/fmicb.2021.645972. URL: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.645972/full>.
- [47] 2020. URL: <https://simtk.org/projects/lab-img-segment>.
- [48] Anna Plichta. *Methods of Classification of the Genera and Species of Bacteria Using Decision Tree*. 2020. URL: https://www.researchgate.net/publication/338252197_Methods_of_Classification_of_the_Genera_and_Species_of_Bacteria_Using_Decision_Tree.

- [49] 2021. URL: <https://scikit-learn.org/stable/index.html>.
- [50] Chunwei Tian et al. “Deep Learning on Image Denoising: An overview”. In: *arXiv.org* (2019). DOI: 10.48550/arXiv.1912.13171. URL: <https://arxiv.org/abs/1912.13171>.
- [51] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. *An empirical exploration of recurrent network architectures*. 2015. URL: <https://research.google/pubs/pub45473/>.

Certification of ownership of the copyright in a typescript or manuscript

Project Report presented as part of, and in accordance with, the requirements for the Final Degree of MSci at the University of Bristol, Faculty of Science.

I hereby assert that I own exclusive copyright in the item named below. I give permission to the University of Bristol Library to add this item to its stock and to make it available for consultation in the library, and for inter-library lending for use in another library. It may be copied in full or in part for any bona fide library or research worked, on the understanding that users are made aware of their obligations under copyright legislation, i.e. that no quotation and no information derived from it may be published without the author's prior consent.

Author	Charles Robinson
Title	Developing a neural network algorithm to perform rapid recognition of bacterial species
Date of submission	26/04/2022

Signed: (electronically is fine)

Full name: Charles Jack Robinson

Date: 26/04/2022

This project is the property of the University of Bristol Library and may only be used with due regard to the rights of the author. Bibliographical references may be noted, but no part may be copied for use or quotation in any published work without the prior permission of the author. In addition, due acknowledgement for any use must be made.

