

**Development of microbial identification platform for industrial pharmaceutical
microbiology using high-throughput Laser-Assisted Rapid Evaporative Ionisation Mass
Spectrometry (LA-REIMS)**

A thesis submitted by:

Toma Ramonaite

For the degree of the Doctor of Philosophy (PhD)

2023

Imperial College London

Department of Metabolism Digestion and Reproduction

'It's always seems impossible until it's done'

-Nelson Mandela-

Statement of originality

I confirm that this thesis and the associated research are the result of my own efforts and work, and that any concepts or citations from the work of others, whether published or not, have been properly recognised in accordance with the standard referencing agreements of the academic discipline.

Toma Ramonaite

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

I would like to express my sincere gratitude to all those who have supported me throughout the journey of completing this thesis. First and foremost, I am deeply thankful to my supervisors, Prof. Zoltan Takats, Andrew Ray and Miriam Guest for their unwavering guidance, invaluable insights, and constant encouragement. Their expertise and mentorship have been instrumental in shaping this research.

I am also grateful to my postdoctoral mentor Dr. Yuchen Xiang for his valuable feedback and constructive suggestions, which have significantly enhanced the quality of this work. Additionally, I extend my appreciation to my colleagues and friends for their support and fruitful discussions.

A shoutout to my amazing friend Maria Sani who kept me sane during late nights working and shared in the triumphs and challenges of this adventure. Your laughter, support, and occasional coffee breaks have made this journey so much more enjoyable.

I would also like to extend a special acknowledgement for the support to my family and to my incredible partner Alex Lamb for his unwavering belief in me and endless encouragement that has been my constant source of motivation. Lastly, to my furry friend Betty, thank you for providing endless moments of joy, laughter, and distraction throughout the countless hours spent researching and writing. Your playful antics and purring companionship have brought a sense of calm and levity to the stressful moments.

To all those who have contributed to this thesis in big and small ways, thank you for being a part of this wild ride. Your support, encouragement, and occasional distractions have played a significant role in reaching this milestone.

Abstract

Pharmaceutical microbiology studies microorganisms to help with drug production or to control the environment around a procedure. The primary goals of pharmaceutical microbiology are to minimise the presence of microorganisms in a processing environment and to prevent the contamination of water and other starting materials with microorganisms and their by-products. As a result, the pharmaceutical industry has devoted substantial effort to the development of quick, sensitive, and accurate methods of identifying indicator bacteria. Due to their inherent advantages of quick data capture, high sensitivity, and specificity, mass spectrometry methods have emerged as one of the main means of building rapid microbiology platforms and are being utilised more frequently in clinical and pharmaceutical microbiology laboratories.

This research projects focuses on developing a LA-REIMS based microbial detection tool for pharmaceutical microbiology that can offer rapid high-accuracy results in as little as a few minutes without any sample pre-processing, performing analysis directly from a microbial petri dish on the sample in its native form. A large library of microbial spectroscopic fingerprints was generated using the prototype automated LA-REIMS platform and utilised to develop training and validation sets for machine learning based multivariate classification models.

LA- REIMS improves analytical throughput and sensitivity with full automation, maximizing the impact for industrial microbiology and direct-from-sample isolate detection. Using single MS technique, a wide range of microorganisms were identified at a remarkable classification accuracy of 100%.

By utilising direct, rapid, real-time, and high-throughput sampling/ionisation of analytes from direct samples without any sample pre-processing LA-REIMS provides substantially enhanced analytical efficiency and offers a powerful and efficient tool for pharmaceutical microbiology.

Table of Content

Statement of originality.....	3
Copyright declaration.....	4
Acknowledgements.....	5
Abstract	6
Table of Content.....	7
List of Figures.....	12
List of Tables	17
List of Abbreviations.....	19
Chapter 1 Introduction	23
1.1 Microorganisms in pharmaceutical manufacturing.....	23
1.2 Pharmaceutical Microbiology	24
1.2.1 Environmental monitoring.....	25
1.2.2 Water testing.....	25
1.2.3 Product related testing	26
1.3 Microbial identification in the pharmaceutical industry.....	26
1.3.1 Phenotypic methods	27
1.3.2 Genotypic methods	27
1.3.3 Mass Spectrometry and microbial identification	29
1.3.3.1 Matrix-Assisted Laser Desorption/Ionisation (MALDI)	31
1.3.3.1.1 Applications of MALDI-TOF MS in environmental microbiology	33
1.3.3.2 Ambient Ionisation Mass Spectrometry	35
1.3.3.2.1 Desorption Electrospray Ionisation (DESI).....	36
1.3.3.2.2 Direct analysis in real-time (DART).....	38
1.3.3.2.3 Paper Spray Ionisation	39
1.3.3.3 Rapid Evaporative Ionisation Mass Spectrometry (REIMS)	41
1.3.3.3.1 Applications of REIMS in microbiology	41
1.3.4 Spectroscopy based methods for microbial identification.....	44

1.3.4.1 FT-IR Spectroscopy.....	44
1.3.4.2 Raman Spectroscopy	45
1.3.5 Summary of MS based microbial identification approaches in comparison to LA-REIMS 45	
Chapter 2 Hypothesis and Aims	49
Chapter 3 Materials and Methods.....	51
3.1 Health and Safety	51
3.2 Microbiological sample storage	51
3.3 Development of a custom LA-REIMS database.....	51
3.4 Culturing	52
3.4.1 Environmental sample monitoring culturing	52
3.4.1.1 Small-scale model building.....	52
3.4.1.2 Large-scale model building.....	52
3.4.2 Water analysis culturing.....	52
3.4.3 Specified organism culturing for absence or presence experiment.....	53
3.4.4 Fungi culturing.....	54
3.5 Fridge incubation.....	54
3.6 Isolate identification and gene sequencing	54
3.6.1 Sanger sequencing.....	54
3.6.2 Agarose gel electrophoresis.....	56
3.7 LA-REIMS platform	56
3.7.1 Mass Spectrometry.....	56
3.7.2 High-Throughput LA-REIMS Analysis	57
3.7.3 Fungi low-resolution imaging data collection	58
3.7.4 Laser power settings.....	59
3.7.5 Calibration and Cleaning.....	59
3.7.6 Leu- enkephalin preparation	60
3.7.7 Quality control	60
3.8 Data processing and statistical analysis.....	60
3.8.1 Peak picking pipeline	60
3.8.2 1mm LA-REIMS imaging data analysis	61

3.8.3 Multivariate data analysis.....	62
3.8.3.1 Post-processing	62
3.8.3.2 Principal Component Analysis.....	62
3.8.3.3 Support Vector Machine analysis.....	63
3.8.3.4 Data repeatability evaluation	64
3.8.3.5 Recursive Feature Elimination.....	64
3.8.4 Performance evaluation.....	65
3.8.4.1 Stratified K-fold cross validation	65
3.8.4.2 Leave one out (loo) cross validation.....	65
3.8.4.3 Independent data set validation	66
3.8.4.4 Confusion matrix.....	66
3.8.4.4.1 Performance scores	67
3.8.4.5 ROC analysis	68
3.8.4.5.1 Comparing ROC Curve and Confusion Matrix	69
3.9 LA-REIMS Lipid and Metabolite annotation	70
3.9.1 Tentative ID confirmation using MS/MS.....	70
3.10 Untargeted LC-MS assay.....	71
3.10.1 Extraction method	71
3.10.2 RPC UPLC-QTOF Analysis of Lipids	71
Chapter 4 Environmental Sample Analysis (bacteria and yeast).....	76
4.1 Introduction.....	76
4.2 Aims of the chapter	77
4.3 Current industrial pharmaceutical environmental monitoring pipeline.....	77
4.3.1 Potential LA-REIMS pipeline for environmental sample monitoring	78
4.4 Results and Discussion	79
4.4.1 Small-scale multivariate model.....	79
4.4.1.1 Optimisation of Pre-processing parameters	80
4.4.1.1.1 Recursive Feature Elimination cross-validation selection (RFECV).....	83
4.4.1.2 Investigation of sample storage effect on LA-REIMS multivariate model stability	87
4.4.1.3 ROC curve analysis	95
4.4.1.4 Repeatability evaluation	97
4.4.1.5 Reproducibility and stability testing.....	98

4.4.2 Large-scale multivariate model.....	103
4.4.2.1 Large-scale model building and validation.....	105
4.4.2.1.1 Recursive Feature Elimination selection (RFE)	110
4.5 Conclusion	115
Chapter 5 Environmental monitoring (fungi).....	118
5.1 Introduction.....	118
5.2 The aims of the chapter.....	118
5.2.1 Current industrial pharmaceutical environmental monitoring for fungi pipeline	119
5.2.1.1 Potential LA-REIMS pipeline for environmental fungi monitoring	119
5.3 Results and Discussion	119
5.3.1 1mm Low-resolution fungi imaging	119
5.3.2 Development and validation of classification model for fungal species	123
5.4 Conclusion	128
Chapter 6 Other applications for LA-REIMS in pharmaceutical microbiological monitoring.....	131
6.1 Aims of the chapter	131
6.2 Water analysis	131
6.2.1 Introduction	131
6.2.1.1 Current industrial pharmaceutical water analysis pipeline	132
6.2.1.2 Potential LA-REIMS pipeline for water analysis.....	132
6.2.2 Results and Discussion.....	133
6.2.2.1 Data collection from filter membrane using LA-REIMS	133
6.2.2.2 Classification model building for water contaminant identification	135
6.3 Absence or Presence of specified organisms	141
6.3.1 Introduction	141
6.3.1.1 Current industrial pharmaceutical specified microorganisms' detection pipeline	142
6.3.1.2 Potential LA-REIMS pipeline for detection of specified microorganisms.....	143
6.3.2 Results and Discussion.....	144
6.3.2.1 Data collection from the microorganisms in the presence of drug substance	144
6.3.2.2 Classification model building for microorganisms in the presence of drug substance.....	147

6.4 Conclusion	149
Chapter 7 Conclusion, Strengths and Limitations, Future Work.....	152
7.1 Conclusion	152
7.2 Strengths and Limitations	154
7.3 Future Work	154
Appendix	156
Bibliography.....	172
Permissions to Reprint.....	197
List of Conferences presentations.....	201
Grants and Awards	202

List of Figures

Figure 1.1 A flow diagram illustrating the main components of a mass spectrometer.....	30
Figure 1.2 MALDI-TOF MS set up for microbial identification. MALDI-TOF MS is used to identify microorganisms. The sample colony is placed on a metal plate and the matrix is covered. The laser ionises the molecules, which are then accelerated by an electric field. The ions are sorted based on their m/z (flight time) until they reach the detector, where the abundance of each signal is recorded. Figure adapted from Santos, Hildenbrand and Schug, 2016, see section <i>Permission to reprint</i>	32
Figure 1.3 Typical MALDI-TOF MS workflow for routine identification of bacteria in industrial pharmaceutical microbiology laboratories.....	33
Figure 1.4 A schematic representation of a DESI-MS setup used for microbial species analysis.	37
Figure 1.5 Schematic representation for paper spray analysis. The clip transmits applied voltage to the paper substrate, where the ionisation process takes place in an ambient environment. Adapted and reprinted with permission from McBride et al., 2019, see section <i>Permission to reprint</i>	40
Figure 1.6 Overview of simplified high-throughput LA-REIMS setup for microbial classification	43
Figure 3.1 Automated Laser-Assisted REIMS platform (culture plate setup). Tecan liquid handling instrument was integrated with a CO ₂ fibre optic laser system. Laser beam ablates the biomass, the generated aerosol travels into the MS via the transfer tubing. The resulting mass spectra are used for data analysis and multivariate modelling for machine learning-based classification.	58
Figure 3.2. MS data pre-processing pipeline	61
Figure 3.3 Example of PCA displayed as a scatter plot.	63
Figure 3.4. A multi-class confusion matrix describing the meanings of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).....	67
Figure 3.5. A schematics of receiver operating characteristic curve (ROC). AUC, area under the ROC curve.	69

Figure 4.1 Current steps and timeline for identifying bacteria and fungus in the pharmaceutical industry's environmental monitoring.	78
Figure 4.2. Confusion matrix and average performance based of '75% peak filter' RFECV features only model validation using independent validation set, with an overall species-level F1 accuracy of 100%.....	84
Figure 4.3 Box plot of 745.517 m/z, PG(34:2), (1/188) features used as a new matrix for the analysis. RFECV feature developed using '75% peak filter', 'Pre-processing 2', and 'TIC+Log' normalisation pre-processing.....	85
Figure 4.4 PCA scores plots of 12 different pre-processing parameter variations. '75% peak filter' with 'Pre-processing 2' and 'TIC+Log' shows the most separation between species and classes.....	86
Figure 4.5 Experimental design for sample storage in the refrigeration investigation. 525 isolates from 15 species were used, which were then used for classification analysis. The reference model and Validation 1 set were treated under the same conditions prior to the analysis - 24/48-hour incubation at 37°C/35°C. Validation 2,3, and 4 were kept in the 4°C fridge for 24, 48, and 72 hours, respectively.....	88
Figure 4.6 Confusion matrix and performance based on SVM classification model with K-fold validation built using 15 species small-scale training model with an overall species-level accuracy of F1 99.73 %.....	89
Figure 4.7 Confusion matrix and performance based on the independent validation set used to validate the reference model above (Figure 4.6) with an overall species-level accuracy of F1 100%.....	90
Figure 4.8 Confusion matrix and performance with the Validation 2, which was analysed after 24 hours in the 4°C fridge and yielded a species-level accuracy of F1 96.10%.	92
Figure 4.9 Confusion matrix and performance with the Validation 3, which was analysed after 48 hours in the 4°C fridge and yielded a species-level accuracy of F1 97.45%.	93
Figure 4.10 Confusion matrix and performance with the Validation 4, which was analysed after 72 hours in the 4°C fridge and yielded a species-level accuracy of F1 97.45%.	94
Figure 4.11 Example of probability table for classicisation generated for each of the validation sets. The table shows the actual vs expected probability of scoring 1. Green cells represent the highest probability of classification for an individual sample.....	95

Figure 4.12 (a), (b), (c) and (d) shows ROC curves from four validation sets tested against the training/reference model. (a) displays the ROC curve for the original validation set (accuracy 98.66%), which was constructed under the same conditions as the reference model. (b) ROC curve of refrigerated validation set (accuracy 98.00%) analysed on LA-REIMS after 24 hours in the 4°C. (c) shows the ROC curve for a refrigerated validation set (accuracy 100%) analysed on LA-REIMS after 48 hours in the 4°C. (d) shows ROC curve which represents refrigerated validation set (accuracy 98.00%) analysed on LA-REIMS after 74 hours in the 4°C..... 96

Figure 4.13 Data repeatability evaluation between data from same species. Graph indicates that the mean pairwise correlation factor is stable across days for the same species. The LA-REIMS platform demonstrated great stability and data comparability throughout multiple days of data collection. 98

Figure 4.14 Confusion matrixes and average classification accuracy representing two training models build using ‘75% peak filter’ and ‘No peak filter’ pre-processing parameters. 100

Figure 4.15 Confusion matrix and performance of training model and independent validation, based on ‘75% peak filter’, with an overall accuracy for independent validation of F1 95.43%. 101

Figure 4.16 Confusion matrix and performance of training model and independent validation, based on ‘No peak filter’, with an overall accuracy for independent validation of F1 25.71%. 102

Figure 4.17. PCA scores plot displaying PC1 vs PC3, of the microbial species used for to develop large-scale classification training model. 106

Figure 4.18 Confusion matrix and average performance of training model containing 162 microbial classes representing 211 species, with an overall species-level accuracy of F1 98.1%. 108

Figure 4.19 Confusion matrix and average performance of independent validation set containing 48/162 microbial classes representing 211 species, with an overall species-level classification accuracy of F1 100%. 109

Figure 4.20 Macro-averaged ROC curve for independent validation set containing 48/162 microbial classes representing 211 species, with AUC=1.00..... 110

Figure 4.21 Confusion matrix and average performance of RFE feature independent validation set containing 48/162 microbial classes representing 211 species, with an overall species-level classification accuracy of 83.7%.....	114
Figure 4.22 Box plot of RFE feature 679.4642 <i>m/z</i> (PG 29:0) with the highest abundance obtained in <i>Georgenia muralis</i> species in the training model.	115
Figure 6.1 Current industrial pharmaceutical water contaminant analysis process (1a) (current timeline 9-18 days depending on the species) in comparison to proposed LA-REIMS method (proposed timeline - 5 days) (1b).....	133
Figure 6.2. Representative mass spectra of the <i>Citrobacter freundii</i> and <i>Ralstonia picketti</i> water pollutants.	135
Figure 6.3 PCA scores plot of PC1 vs PC2 of the top 10 industrial pharmaceutical water contaminants analysed with LA-REIMS platform directly from filter membrane on the agar plate.....	136
Figure 6.4 Confusion matrix and the average performance of the training model based on SVM classification model build using top 10 industrial pharmaceutical water contaminants with an overall species-level accuracy of F1 99.34 %.	138
Figure 6.5 Confusion matrix and the performance of the independent validation set based on SVM classification model build using top 10 industrial pharmaceutical water contaminants with an overall species-level accuracy of 97.97 %.	139
Figure 6.6 Macro-averaged plot of ROC for AUC for top 10 industrial pharmaceutical water pollutants classification (independent validation set) with AUC = 0.99.....	141
Figure 6.7 Current industrial pharmaceutical absence or presence of specified organisms' workflow. Current timeline 9-18 days depending on the species.....	143
Figure 6.8 Proposed approach for detection of specified microorganisms using LA-REIMS (proposed timeline- 5 days). <i>E.coli</i> (EC), <i>Candida albicans</i> (CA) and <i>Staphylococcus aureus</i> (SA).	143
Figure 6.9 PCA score plot of PC1 vs PC2 of data collected from <i>E.coli</i> , <i>Candida albicans</i> and <i>Staphylococcus aureus</i> cultured with and without the drug substance present.	145
Figure 6.10 Representative mass spectra of <i>E.coli</i> with and without the drug presence during sample preparation prior to the LA-REIMS analysis.....	146

Figure 6.11 Confusion matrix and average performance based on the SVM classification model validated using loo, built using the specified organisms with and without the presence of the drug substance..... 148

Figure 6.12 Confusion matrix and the average performance based on the SVM classification model built using the specified organisms cultured with the presence of the drug substances. 149

List of Tables

Table 3.1 PCR primers list for bacteria, yeast and fungi DNA amplification.....	55
Table 3.2. MS instrument parameters used during LA-REIMS analysis.....	56
Table 3.3. OmniGuide laser settings used in LA-REIMS analysis.....	59
Table 3.4 NPC LC instrument parameters.	72
Table 3.5 NPC MS source parameters.	72
Table 3.6. NPC MS instrument parameters.	73
Table 3.7 NPC LC gradient list.	73
Table 4.1 Microbial species list for small scale multivariate model building. B-bacteria, Y-yeast.....	80
Table 4.2. MS data pre- and post-processing variables and parameters	81
Table 4.3. Pre- and post-processing two-step validation results using “No peak filter”, ‘Pre-processing 1’, ‘Pre-processing 2’, and three different normalisation variations. The best combinations with an F1 accuracy of 98.02% and ROC=1 was ‘No peak filter’ with ‘Pre-processing 2’and “Log” normalisation parameters.....	82
Table 4.4. Pre- and post-processing two-step validation results using ‘75% peak filter’, ‘Pre-processing 1’, ‘Pre-processing 2’ and three different normalisation variations. The best combinations with an F1 accuracy of 100% and ROC=1 was ‘75% peak filter’ with ‘Pre-processing 2’and ‘TIC+Log’ normalisation parameters.	82
Table 4.5. Results from the top two pre-processing parameter combinations using third-stage validation of RFECV features. The model accuracy obtained using only specific features was greatest when ‘75% peak filter’ with ‘Pre-processing 2’ and ‘TIC+Log’ normalisation was used.	83
Table 4.6 The training model and independent validation classification accuracy using ‘No peak filter’ and ‘75% peak filter’ pre-processing along with ‘Pre-processing 2’and “TIC+Log”.	100
Table 4.7 Bacterial and yeast strains representing 211 different species used in large-scale classification model building for industrial pharmaceutical environmental monitoring	104

Table 4.8 List of statistically important biomarkers and their lipid and metabolite annotations.	111
Table 4.9. Average classification results of training model developed using 75 statistically important biological biomarkers.....	113
Table 5.10 Average performance of the independent validation set, consisting of six out of the twelve fungal species, analysed six months after the initial reference model isolates and multiple subculturing. Overall species-level classification accuracy 87%.	127
Table 6.1 The combination of microbes, drug, enrichment medium, and selective agar used in the experiment.....	144

List of Abbreviations

ACN	Acetonitrile
AIMS	Ambient Ionisation Mass Spectrometry
API	Analytical Profile Index
BLAST	Basic Local Alignment Search Tool
CBA	Columbian blood agar
CFU	Colony Forming Unit
COSHH	Control of Substances Hazardous to Health
DART	Direct Analysis in Real Time
DART-MS	Direct analysis in real time mass spectrometry
DDA	Data Dependent Acquisition
DESI	Desorption Electrospray Ionisation
DNA	Deoxyribonucleic acid
ELISA	Enzyme-Linked Immunosorbent Assay
ESI	Electrospray Ionisation
ESI-MS	Electrospray ionisation mass spectrometry
FA	Fatty Acid
FAME	Fatty Acid Methyl Esters
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FT	Fourier Transform
FTIR	Fourier Transform Infrared
GC	Gas Chromatography
HPLC	High-Performance Liquid Chromatography
ID	Identification
IFA	Immunofluorescence Assays
IPA	Isopropanol or propan-2-ol
IR	Infrared
LA	Laser Assisted
LA-REIMS	Laser Assisted Rapid Evaporative Ionisation Mass Spectrometry

LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
LOO	Leave-one-out
LOOCV	Leave-one-out Cross Validation
LPE	Lysophosphatidylethanolamine
LPG	Lysophosphatidylglycerol
LPI	Lysophosphatidylinositol
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MLT	Median Log Transform
MPN	Most Probable Number
MS	Mass spectrometry
MS/MS	Tandem Mass Spectrometry
NCBI	National Centre for Biotechnology Information
NPC	National Phenome Centre
PA	Phosphatidic Acid
PC	Phosphatidylcholine
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PLS	Partial Least Squares
PMF	Peptide Mass Fingerprint
POC	Proof Of Concept
PSI	Paper Spray Ionisation
PSI-MS	Paper Spray Ionisation Mass Spectrometry
PTFE	Polytetrafluoroethylene
QToF	Quadrupole Time of Flight
REIMS	Rapid Evaporative Ionisation Mass Spectrometry
RFE	Recursive Feature Elimination
RFECV	Recursive Feature Elimination, Cross-Validated
RNA	ribonucleic acid

ROC	Receiver Operating Characteristic curve
SAB	Sabouraud agar
SVM	Support Vector Machine
TG	Triacylglycerol
TIC	Total Ion Current
TN	True Negative
TOF	Time of Flight
TP	True Positive
TPR	True Positive Rate
TSA	Trypticase soy agar
TSB	Trypticase soy broth
UPLC	Ultra-Performance Liquid Chromatography
UV	Ultraviolet

Chapter 1

Introduction

Chapter 1 Introduction

1.1 Microorganisms in pharmaceutical manufacturing

Microorganisms can be found everywhere around us, performing a variety of critical functions. They serve a crucial part in human culture and health, including the production of food and beverages, antibiotics, and specific enzymes¹⁻³. However, microbial contamination can have adverse effects on human health, making it essential to maintain microbial control when it comes to pharmaceutical manufacturing⁴. The presence of microorganisms in pharmaceutical products can lead to product degradation, loss of potency, and contamination, which can have serious consequences for patient health⁵. Therefore, quality control measures must be implemented to prevent microbial contamination and ensure the safety and efficacy of pharmaceutical products⁶.

Within the pharmaceutical industry, a range of microorganisms may be encountered during routine operations⁷. Bacteria are one of the most common types of microorganisms found in pharmaceutical settings. They can be introduced from raw materials, personnel, air, or equipment⁸. Gram-positive bacteria are a significant group encountered in pharmaceutical settings that can cause infections and is particularly problematic due to their ability to develop resistance to antibiotics. Gram-negative bacteria are another prevalent group found in the pharmaceutical environment, that mainly indicates potential faecal contamination, emphasizing the need for stringent hygiene practices and quality control measures^{9,10}. Examples of bacteria commonly found in pharmaceutical environments include *Staphylococcus aureus*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Bacillus* species.

Fungal contamination is another concern in the pharmaceutical industry¹¹. They can contaminate raw materials, equipment, surfaces, and air, presenting challenges to product safety and quality. Filamentous fungi, commonly referred to as molds can lead to significant risk and impact factors such as product stability issues, allergic reactions, and infections^{9,12}. Common fungi encountered in pharmaceutical environments include *Aspergillus*, *Penicillium*, *Candida*, and *Alternaria* species. *Aspergillus fumigatus*, in particular, poses a significant risk as it can cause invasive infections, especially in immunocompromised individuals. Yeasts, another type of fungal microorganism, can also contaminate pharmaceutical products, especially those with liquid or semi-solid formulations¹³.

The prevalence and specific types of microorganisms present in pharmaceutical manufacturing can vary based on the manufacturing process, product type, and environmental conditions. Pharmaceutical companies implement rigorous control measures to ensure the safety and quality of their products. These measures include environmental monitoring, sterilisation techniques, and quality control testing, which are designed to prevent and identify microbial contamination^{11,14}.

1.2 Pharmaceutical Microbiology

The study of microorganisms in connection to the production of medications is the focus of pharmaceutical microbiology. This has to do with either controlling the microbial contaminants in a process environment or employing microorganisms to assist in the production of medications¹⁵. Taking care of the completed product's sterility or absence of those particular strains that are considered as unacceptable is the focus of the major issue. This extends to raw materials and water used in the manufacturing process. Pharmaceutical microbiologists are also interested in toxins (microbial by-products such as endotoxins and pyrogens), especially with regard to ensuring that these and other of microbes (which may trigger unfavourable patient reactions) are absent from products¹⁶⁻¹⁸.

If microbiological contamination causes unintended side effects in pharmaceutical formulations, it becomes an issue. Pharmaceutical microbiologists develop control strategies for products, using risk assessment to focus on understanding the likelihood of product contamination, the severity of such contamination, ways to reduce/eliminate it, and apply methods to detect contamination, throughout the product lifecycle¹⁸.

Understanding the product's type, intended usage, and quantity of pollutants is crucial to comprehend the severity. The biggest impact is microbial contamination of sterile injectable products (parenteral), which can cause patient mortality. In contrast, microbially caused smells, off-flavours, or discolorations in other items may cause less harm. Therefore, the fundamental issue with regard to sterile items is any potential microbial contamination. A certain amount of microbial contamination may be acceptable in Pharmaceutical Microbiology nonsterile goods (such as inhalers, tablets, oral liquids, creams, and ointments), with the presence or absence of "objectionable bacteria" microbial enumeration tests and presence/absence tests of specified organisms are of interest. Specified organisms are representative organisms that may pose the highest patient safety risk for the route of administration. The sorts of microorganisms of

concern depends on the product type and administration route^{19–22}. The three main international pharmacopoeias—American, European, and Japanese, that are largely harmonised for microbiological methods—describe some of the crucial tests used to monitor and identify these contaminants. These consist of²⁰:

- Environmental monitoring
- Methods for testing pharmaceutical water
- Antimicrobial susceptibility testing
- Tests for specified microorganisms
- Microbial enumeration methods
- Bacterial endotoxin test
- Sterility test

This research will look into three key aspects of pharmaceutical testing: environmental monitoring, water testing, and product-related testing.

1.2.1 Environmental monitoring

The evaluation of pharmaceutical production facilities includes a critical component called microbiological environmental monitoring. Data from environmental monitoring shows whether cleanrooms are functioning properly. An evaluation of the parameters for heating, ventilation, and air conditioning, which help to make the cleanroom "clean," yields additional information about contamination control. These variables include room air changes, temperature, humidity, pressure, and airflow patterns^{15,16}.

1.2.2 Water testing

The pharmaceutical industry tests a variety of devices that aid in production. This comprises compressed gases (which are checked for microbial count and particles), steam (which is usually checked for bacterial endotoxin as condensate), and most importantly - pharmaceutical-grade water. Water for injection and filtered water are examples of pharmaceutical-grade water. Moreover, the main water (feed water) is also examined. Water samples can underestimate the size of any biofilm communities that may have developed on pipes, but they still provide a thorough picture of the microorganisms that are present in the water^{23–25}.

1.2.3 Product related testing

During the pharmaceutical manufacturing process, four main areas are investigated to establish a sampling and testing regime^{26–28}. This involves selecting, sampling, and testing:

- starting materials
- in-process samples
- final product formulation
- testing of the final product

Pharmacopoeia monographs for the vast majority of starting materials (raw materials) will describe the kind of microbiological tests necessary. The following assays often make up the needed testing on starting materials²⁹:

- Microbial enumeration; tests for total microbial, mould and yeast count
- Presence-absence of particular pre-determined microorganisms³⁰; bile-tolerant, Gram-negative bacteria, *Salmonella*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Candida albicans*, *Clostridia*.

In-process samples, obtained at representative stages or places in the process when there is believed to be a risk, are typically taken, and submitted to the microbiology laboratory for bioburden testing while a product is made. The number of bacteria residing on a surface or in a material is referred to as the bioburden. The total viable count approach is commonly used to assess bioburden³¹.

For both sterile and non-sterile products, the finished product must always be tested for microorganisms. Pharmacopoeia normally defines the testing specifications and requirements for sterile items. This involves the sterility test, which verifies that the product is devoid of microbes that are capable of reproducing, and typically an endotoxin or pyrogen test. When it comes to non-sterile items, there will usually be specifications based on the maximum allowable microbial count as well as a requirement that the product be devoid of specific undesirable bacteria^{19,22,32}.

1.3 Microbial identification in the pharmaceutical industry

Microbial identification plays a crucial role in the pharmaceutical industry to ensure the safety, quality, and efficacy of pharmaceutical products. Identification of microorganisms is essential

for assessing product contamination, monitoring manufacturing processes, and maintaining compliance with regulatory requirements^{33–35}. Over the years, various microbial identification techniques have been developed and employed in the pharmaceutical industry to accurately and rapidly identify microorganisms. These techniques have evolved from traditional phenotypic methods to more advanced molecular and mass spectrometry-based approaches^{36–39}.

1.3.1 Phenotypic methods

Traditional methods that are based on phenotypic identification of microorganisms using culture, staining, and biochemical methods continue to dominate in diagnostic laboratories; however, these methods can exhibit significant variation as a result of various factors such as culture, subculture, or handling and storage^{40,41}.

Plate culture-based procedures are the most often applied, although they can be time-consuming (1-18 days) and labour-intensive. Organisms can be slow growing and intolerant, and it can take many weeks or even months to obtain an appropriate culture, as in the case of *Mycobacterium spp*⁴². Differential staining techniques (Gram stain, acid-fast) are useful for the subjective evaluation of bacterial morphology, although while staining techniques might supplement analysis, the cells of many bacteria are often similar, and additional biochemical tests are required for reliable identification⁴³.

A wide range of biochemical tests, as well as selective media and chemotaxonomic approaches, are available, which are often commercially developed and sold as systems such as the analytical profile index (API) ID strip and VITEK 2 system (bioMérieux inc. , France)^{44,45}. While phenotypic methods have been widely used in the past, they have certain limitations such as subjectivity, time-consuming procedures, and a limited ability to differentiate closely related species or strains. As a result, the pharmaceutical industry has gradually shifted towards more advanced techniques for microbial identification and serological and molecular methods like enzyme-linked-immunoassays (ELISA) and immunofluorescence assays (IFA), which might possibly bypass this phase, are frequently used²⁰.

1.3.2 Genotypic methods

The latest genomic revolution has resulted in a variety of molecular biology methods for microbe detection⁴⁶. These nucleic acid-based approaches provide selectivity determined by a

pathogen's genome's genetic features, which stay mostly stable over time and are able to withstand environmental changes. Molecular techniques depend on amplification (polymerase chain reaction) or probe-based (nucleic acid hybridization, microarrays) techniques to detect microbe-specific genes^{47,48}.

Polymerase chain reaction (PCR) is a popular microbiological detection tool. Traditional approaches may depend on gel electrophoresis or gene probe hybridisation to identify amplicons⁴⁹. However, because no post-PCR analysis is necessary, real-time PCR is more routinely used, minimizing turnaround times, and decreasing contamination. Furthermore, the approach is frequently employed for quantification, which is critical for particular cases^{49,50}.

The most common genomic method used in bacterial classification is 16S rRNA gene sequencing, which has well-conserved primer regions and contains hypervariable areas at the genus or species level⁵¹. As a result, it is highly suited for categorising microorganisms that are hard to culture and those with unclear biochemical experimental methodology. It has also been extensively applied in the analysis of a wide range of microbial samples with next-generation sequencing technologies such as Illumina MiSeq and Roche 454⁵². DNA is extracted from microbial cells using a commercial extraction kit. Kits can lyse cells enzymatically or physically and bind DNA to a matrix. Bound DNA can be eluted, and specialised primer sets can be utilised to target the 16S rRNA gene region⁵³. In the case of Illumina 16S rRNA gene sequencing library preparation, PCR is used to amplify the regions of interest; samples are then processed through the sequencer, which determines the sequence using a fluorescence imaging system^{52,53}. Because of the millions of target sequences generated, data sets can be extremely large, making analysis complex and laborious, with the frequent need for custom bioinformatics tools to categorise bacterial taxonomy⁵⁴.

Although genotypic techniques have been regarded as the "gold standard" for identification, they are rarely used due to time-consuming preparation, a lack of equipment and expertise, difficult analysis, and high prices⁵⁵. Furthermore, studies using a 16S rRNA gene sequencing technique to analyse the microbes are not fully quantitative and frequently over-represent some identifications at the cost of others due to discrepancies in gene copy counts between species. An innovative method for quick identification of microbial species could therefore revolutionise environmental microbiology and vastly enhance microbial classification findings^{56–58}.

1.3.3 Mass Spectrometry and microbial identification

Mass spectrometry (MS) has evolved into an essential instrument for microbial identification, revolutionizing the field of microbiology⁵⁹⁻⁶¹. By analysing the mass-to-charge ratio of molecules, MS allows for the fast and accurate identification of microorganisms based on their unique molecular fingerprints^{62,63,64}. The application of MS in microbial identification has transformed the way microorganisms are characterized, enabling faster, more reliable, and comprehensive identification compared to traditional methods^{65,66}.

The fundamental principle of mass spectrometry involves the ionization, separation, and detection of charged molecules⁶⁷. The process begins by ionizing the sample, converting the molecules into electrically charged species called ions. This ionization can be achieved using various techniques, such as electron impact ionization, electrospray ionization, or matrix-assisted laser desorption/ionization^{67,68}.

Once the sample is ionized, the ions are subjected to an electric or magnetic field, which causes their separation based on their mass-to-charge ratio. This separation is accomplished in a mass analyser, which can be of different types, including time-of-flight (TOF), quadrupole, ion trap, orbitrap or magnetic sector analysers. The mass analyser selectively allows ions of specific masses to pass through and reach the detector, while others are filtered out⁶⁹⁻⁷². TOF analysers offer fast data acquisition rates and high mass accuracy, making them suitable for high-throughput applications such as microbial identification and proteomics⁶³.

The final step in the mass spectrometry process is the detection and analysis of the ions. The ions that reach the detector generate signals that are recorded and analysed to obtain mass spectra⁷³. Mass spectra represent a plot of ion abundance versus the mass-to-charge ratio, providing valuable information about the molecular composition and structure of the analysed sample⁷⁴.

The interpretation of mass spectrometry data depends on extensive databases containing reference spectra of known compounds. By comparing the acquired mass spectra with these databases, researchers can identify the molecules present in their samples. Furthermore, the development of sophisticated data analysis algorithms and statistical methods has facilitated the extraction of valuable information from complex mass spectrometry datasets⁷⁵.

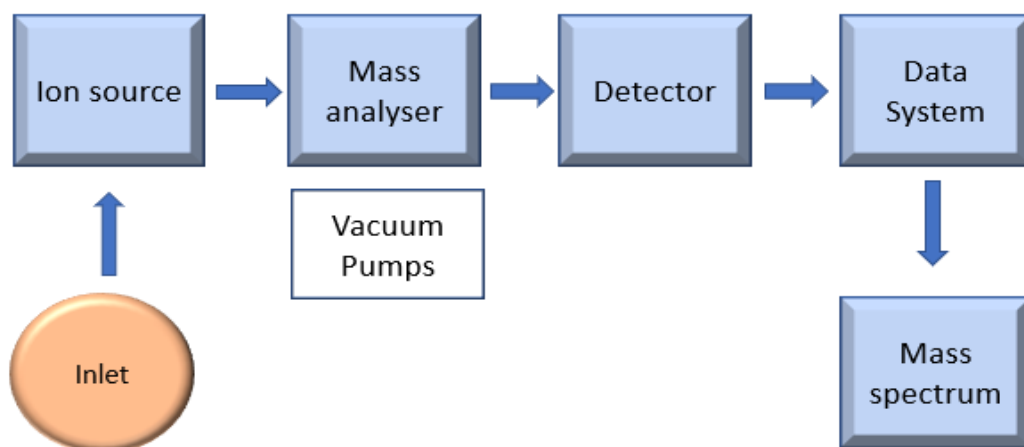


Figure 1.1 A flow diagram illustrating the main components of a mass spectrometer.

Mass spectrometry has an extensive range of applications across various scientific disciplines. In chemistry, it is used for the analysis of organic and inorganic compounds, including small molecules, polymers, and complex mixtures⁷⁶. Mass spectrometry is also invaluable in proteomics, allowing for the identification and characterization of proteins, peptides, and post-translational modifications. In metabolomics, mass spectrometry is employed to study the small-molecule metabolites present in biological samples, providing insights into metabolic pathways and disease biomarkers. Additionally, mass spectrometry has applications in environmental analysis, forensics, pharmaceuticals, and many other fields^{77,78}.

The advancement of mass spectrometry techniques has led to significant improvements in sensitivity, resolution, and accuracy. Modern mass spectrometers can detect and analyse molecules with exceptional precision, even at trace levels⁷⁷. Additionally, the integration of mass spectrometry with other analytical techniques, such as chromatography, enhances the separation and identification of complex mixtures. The integration of MS-based methods with other analytical techniques, such as liquid chromatography (LC), enables more comprehensive characterization of microorganisms. LC-MS permits for the identification and quantification of specific metabolites, lipids, or secondary metabolites produced by microorganisms⁷⁹. This information can be valuable for assessing the virulence potential, metabolic capabilities, or quality attributes of microbial strains^{80–83}.

Electrospray ionisation mass spectrometry (ESI-MS) is another commonly used technique for microbial identification in the pharmaceutical industry^{84–86}. It involves the ionization of analytes in solution using an electrospray source, which generates ions that are subsequently analysed in a mass spectrometer. ESI-MS is particularly useful for the analysis of small molecules, such as metabolites, lipids, and secondary metabolites produced by microorganisms^{80,87}.

Over the past 40 years, a large variety of mass spectrometry techniques, apparatus, and procedures have been used to analyse microorganisms⁸⁸. Combining improvements in ionisation technologies, low-cost compact mass analysers, and bioinformatics, several MS techniques including Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS or ambient ionisation mass spectrometry techniques have emerged for microbial identification in the pharmaceutical industry^{89–92}. Ambient ionization techniques have proven to be valuable tools for microbial identification in the industrial pharmaceutical sector⁹³. These techniques offer rapid, direct, and on-site analysis of microbial samples without the need for extensive sample preparation. By enabling real-time identification, ambient ionization techniques contribute to the timely detection and control of microbial contaminants in pharmaceutical manufacturing processes^{94–97}.

1.3.3.1 Matrix-Assisted Laser Desorption/Ionisation (MALDI)

Matrix-Assisted Laser Desorption/Ionization (MALDI) is an ionisation technique that was developed in the 1980s. With the use of a UV laser beam directed at the sample that has been co-crystallized in a matrix compound, the soft ionisation technique known as MALDI causes analyte desorption and ionisation. The analyte can be vaporised or ionised more effectively because of the matrix's ability to absorb laser energy at a particular wavelength, which reduces fragmentation^{84,98–100}.

By demonstrating that matrix-assisted laser desorption/ionization (MALDI) can produce high m/z ions (500 to 300,000 m/z) for the analysis of biomolecules, Karas, Hillenkamp, and Tanaka's work had a significant impact on biomolecule characterization by expanding the scope of molecular weights and sample types that could be examined. Mass spectrometric analysis of complex, non-volatile biomolecules, such as peptides, proteins, oligonucleotides, and oligosaccharides, is frequently performed with MALDI, enabling applications in a variety of scientific domains^{20,101,102}.

The sample for MALDI MS analysis is composed by mixing or coating with a solvent of an energy-absorbent, organic substance known as matrix. When the matrix crystallises after drying, the sample entrapped within it likewise crystallises. The sample within the matrix is ionised using a laser beam in an automated manner. Desorption and ionisation with the laser beam produce singly protonated ions from analytes in the sample. The molecules are then accelerated at a predetermined potential, where they separate according on their mass-to-charge ratio (m/z). The charged analytes are subsequently detected and analysed using various mass analysers such as quadrupole mass analysers, ion trap analysers, time of flight (TOF) analysers, and so on^{87,103–107}. The choice of analyser depends on aspects such as resolution requirements, mass accuracy, sensitivity, and specific research goals. Instrument configurations combining MALDI with different mass analysers offer flexibility and versatility in addressing various analytical challenges in fields ranging from proteomics and metabolomics to microbiology and pharmaceutical research^{108–111}.

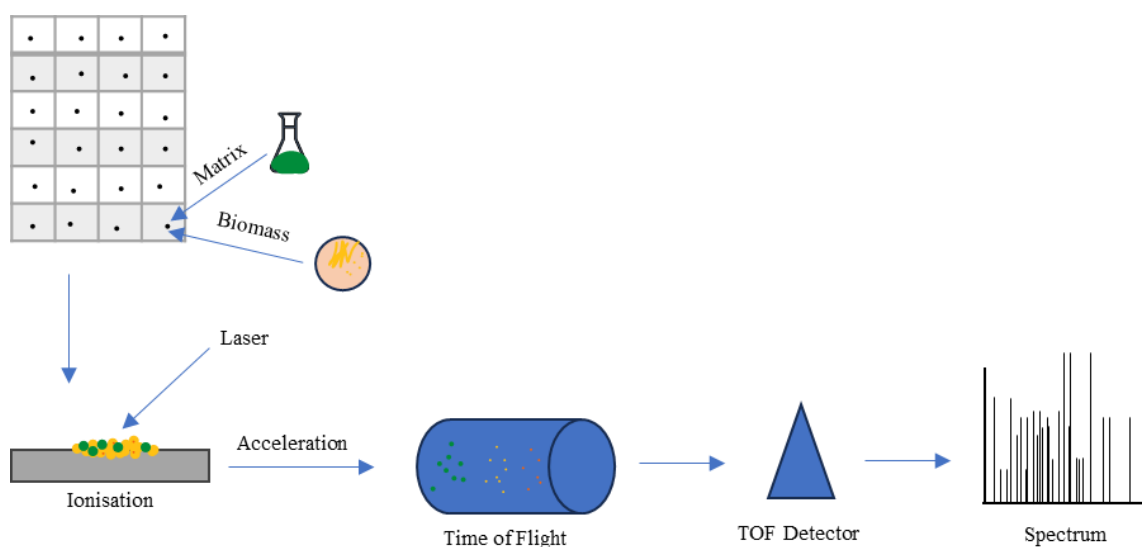


Figure 1.2 MALDI-TOF MS set up for microbial identification. MALDI-TOF MS is used to identify microorganisms. The sample colony is placed on a metal plate and the matrix is covered. The laser ionises the molecules, which are then accelerated by an electric field. The ions are sorted based on their m/z (flight time) until they reach the detector, where the abundance of each signal is recorded. Figure adapted from Santos, Hildenbrand and Schug, 2016, see section *Permission to reprint*.

1.3.3.1.1 Applications of MALDI-TOF MS in environmental microbiology

TOF mass analysers are commonly employed in microbiological applications due to their suitability for the high-throughput analysis. The m/z ratio of an ion is assessed during MALDI-TOF analysis by calculating the time it takes to travel the length of the flight tube^{112,113}. A few TOF analysers have an ion mirror at the back end of the flight tube that reflects ions back through the flight tube to a detector. As a result, the ion mirror not only lengthens the flight tube which improves the mass resolution, but it also corrects minor energy disparities between ions. For analytes in the sample, a characteristic spectrum known as peptide mass fingerprint (PMF) is formed based on the TOF information^{114–117}. Figure below shows a summary of a typical workflow for identification of bacteria in industrial pharmaceutical microbiology laboratories using MALDI-TOF MS system.

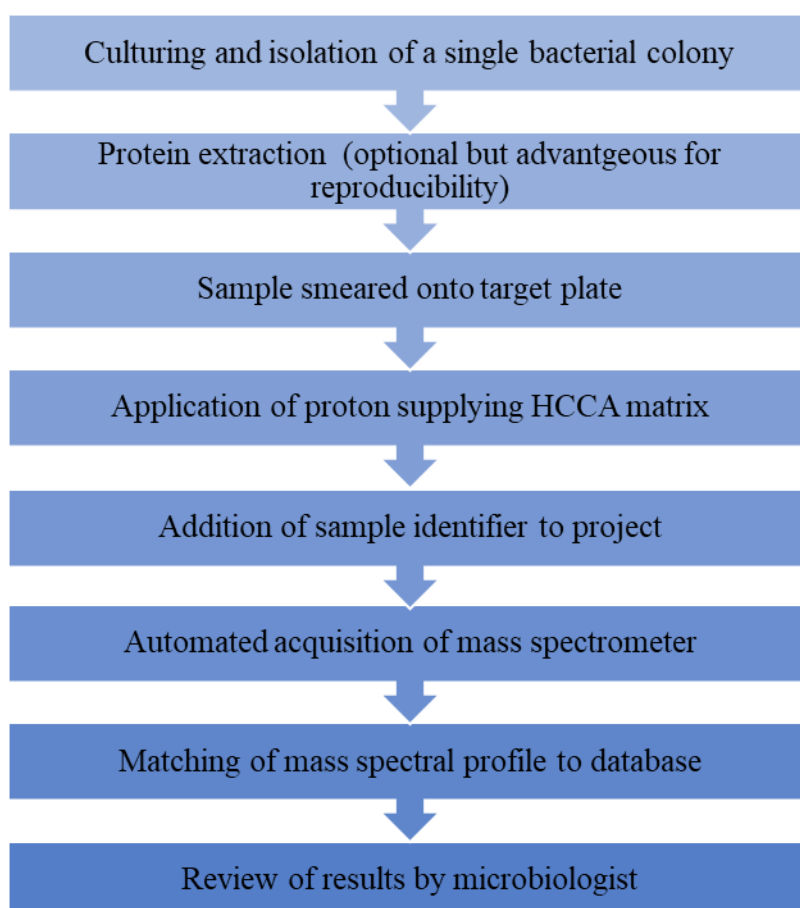


Figure 1.3 Typical MALDI-TOF MS workflow for routine identification of bacteria in industrial pharmaceutical microbiology laboratories.

MALDI-TOF Mass spectrometry (MS) is a technique utilised for bacterial identification through the comparison of the peptide mass fingerprint (PMF) of an unknown organism with those present in a database. Additionally, identification can be achieved by matching the biomarkers of the unknown organism to the proteome database.^{118–120} PMF matching compares the MS spectra of unknown microbiological isolates to the MS spectra of known microbial isolates in the database. For microbe species identification, a typical mass range m/z of 2–20 kDa is utilised, which contains primarily ribosomal proteins as well as a few housekeeping proteins. The identification of a particular microorganism is facilitated by comparing its PMF pattern with the PMFs of ribosomal proteins present in a comprehensive open database. These ribosomal proteins, which constitute approximately 60–70% of the dry weight of a microbial cell and are highly abundant in the mass range of 2–20 kDa, exhibit a distinctive pattern^{89,105,121–123}. Consequently, it is possible to identify the origin of a microorganism up to the level of genus, and frequently, down to the level of species and strain. The utilisation of this technique has become increasingly common in the identification of microorganisms due to its simplicity and feasibility in a microbiological diagnostic setting, as well as the existence of numerous commercially available libraries containing organism PMFs¹²⁴. The practise of employing biomarker masses to match with the predicted molecular masses of proteins from a genome sequence for microbial identification is not commonly used in microbiological diagnostic facilities due to the necessity of possessing comprehensive knowledge of an organism's complete genome sequence prior to establishing a database of its predicted protein molecular masses^{66,125–127}.

MALDI-TOF MS is a revolutionary innovation for identifying, characterising, and typing microorganisms^{128–130}. This approach has primarily been applied in the clinical field, but it also has great environmental potential. Through faster sample turnaround, the incorporation of MALDI systems into clinical microbiology laboratories enhances patient outcomes, decreases hospital stay durations, and shortens the time before correct antibiotic action^{131,132}. Hundreds of systems have been implemented in clinical microbiology laboratories around the world in recent years. The significance of swiftly identifying microorganisms responsible in human infections and, as a result, applying the appropriate therapy is undeniable^{133–136}.

Environmental microbiology is a field where the application of MALDI-TOF MS has yet to be completely examined, although it has been shown that the technology is already making a substantial contribution⁶⁴. MALDI-TOF MS is commonly utilised in proteomics, particularly

microbial typing, but it can also be applied in other investigations such as metabolomics. These two domains, respectively, provide information on the expression of proteins and the diversity of generated metabolites in an organism under specific conditions^{137–139}.

Bruker has used its many years of experience to produce the really innovative MALDI Biotyper System to assist answer important difficulties in microbiology¹⁴⁰. This innovative technology has been implemented in many microbiology laboratories around the world over the last decade for the reliable, fast, and efficient identification of numerous gram-negative and gram-positive bacteria, yeasts, and moulds using an easy-to-use, yet powerful benchtop analyser^{141,142}. It has been established that it is a reliable approach of protein-based identification of intact bacteria and yeast, however the use of MALDI-TOF for mould or filamentous fungi identification is difficult and has not yet attained a high degree of accuracy^{143–145}.

The MALDI Biotyper reference library now contains ~9,607 reference spectra from ~3,239 species representing ~593 bacteria and fungus taxa. The library is constantly maintained and updated in accordance with stringent quality-control processes. There are also separate libraries dedicated to filamentous fungus and mycobacteria. Laboratories can also use their spectra to develop their own reference libraries, for example, on specific local bacteria. These unique libraries can be shared and swapped with other pharmaceutical company plants as well as externally^{146–149}.

MALDI-TOF improved routine microbial identification by providing faster and at a lower cost per sample analysis, however, they do necessitate the employment of a matrix to aid in lysis and ionisation, and in some circumstances, further extraction phases are also required for correct species classification¹⁵⁰. Moreover, the complex nature of environmental samples and the wide range of environmental microorganisms pose challenges. Nevertheless, further advancements, such as the enhancement of protein databases and refinement of sample preparation techniques, are necessary prior to the integration of this methodology as a conventional practice in the field of environmental microbiology^{127,151}.

1.3.3.2 Ambient Ionisation Mass Spectrometry

Ambient ionisation mass spectrometry (AIMS) has become a vital scientific research instrument in a variety of disciplines, including medicines, food, environment analysis, public safety, clinic diagnosis, and pharmaceutical development¹⁵². The cornerstone of AIMS analysis is ambient ionisation technology, which permits direct, rapid, real-time, and high-throughput

sampling/ionisation of analytes from direct samples without any sample pre-processing, and has substantially enhanced analytical efficiency^{153,154}.

With each passing year, the field of AIMS improves and expands, with the development of powerful and unique methodologies, enhancements to existing instruments, and intriguing new applications¹⁵³. Ambient ionization techniques have gained prominence in industrial pharmaceutical microbial identification due to their ability to provide rapid, direct, and on-site analysis. Desorption Electrospray Ionization (DESI), Direct Analysis in Real Time (DART), and Paper Spray Ionization (PSI), are among the prominent techniques utilized for this purpose¹⁵⁵. These techniques contribute to the efficient detection and control of microbial contaminants in pharmaceutical manufacturing, enhancing product safety, quality, and regulatory compliance. The implementation of ambient ionization in industrial pharmaceutical settings supports proactive monitoring and timely interventions to ensure the integrity and safety of pharmaceutical products^{156,157}. The AIMS approach offers several benefits for metabolic profiling of microbial communities. It involves the exclusion of sample pre-treatment, such as metabolite extraction, which mitigates the risk of introducing bias and minimises any negative impacts of disrupting the microbial community from its native habitat. Additionally, this method enhances sample throughput^{158–160}. These characteristics make AIMS ideal for applications where time-critical decisions need to be made or where sampling from complex matrices is challenging¹⁶¹.

1.3.3.2.1 Desorption Electrospray Ionisation (DESI)

DESI-MS is an ambient ionization technique that allows for direct analysis of samples in their native state, without the need for extensive sample preparation. It combines the principles of electrospray ionization and desorption/ionization, enabling the direct sampling and ionization of analytes from various surfaces, including microbial colonies, biofilms, and environmental samples¹⁶².

DESI-MS employs an ionization mechanism that involves the interaction of a spray solvent with the microbial sample surface. The process involves the generation of charged droplets from a solvent that is sprayed onto the sample surface. These charged droplets interact with the analytes on the surface, causing their desorption and ionization. The ions are then transferred into the mass spectrometer for analysis. The mass spectrometer splits the ions based on their mass-to-charge ratio (m/z), allowing for the identification and quantification of the analytes^{163–}

¹⁶⁵. This technique enables the direct analysis of intact cells, colonies, biofilms, or environmental samples, reducing the time and effort required for microbial identification.

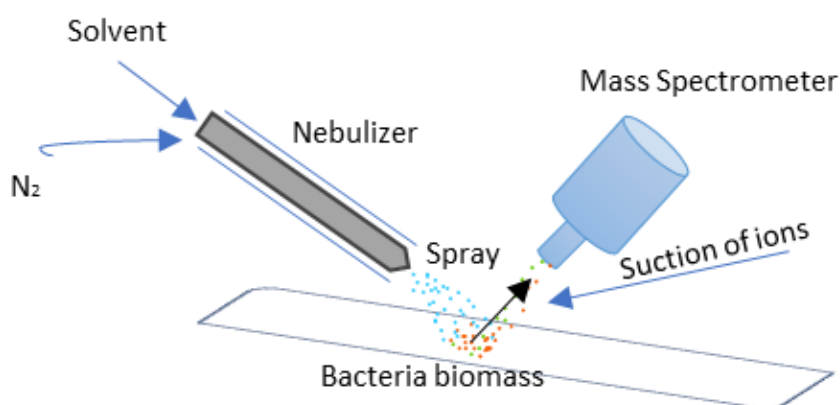


Figure 1.4 A schematic representation of a DESI-MS setup used for microbial species analysis.

One of the key advantages of DESI-MS is its ability to provide comprehensive molecular profiling of microbial samples. It can identify lipids, proteins, peptides, metabolites, and other biomolecules, offering valuable insights into the characteristics and behaviour of microorganisms¹⁵⁶. By mapping molecular distributions within microbial samples, DESI-MS provides spatial resolution, revealing the spatial distribution of specific biomolecules or microbial species^{166,167}.

Putting an emphasis on microbiological investigation the desorption mechanism does not, unfortunately, make it possible to directly analyse surfaces that are either soft or uneven. This is because of the significant gas pressure that is generated upon interaction with the surface. Because of this constraint, the analysis of microorganisms grown on agar media using DESI-MS is a very difficult process^{168,169}. In addition, the high pressure of the gas and spray can cause divots to be formed in the medium, which results in a change to the predetermined distances and angles that are necessary for the process of ionisation to be reliable and effective¹⁷⁰. In addition, unlike bacterial cultures, fungal cultures are exceedingly uneven, and because both aerial hyphae and spores are present on the agar plate and are rapidly influenced by high pressure gas, DESI-MS analysis is not practicable on such samples¹⁷¹. In order to overcome these issues, samples must be prepared for the analysis by transferring biomass on suitable substrate such as a glass slide or an empty petri dish^{172,173}.

One of the drawbacks associated with DESI refers to the considerable quantity of experimental parameters that necessitate optimisation for each experiment. While this feature makes the technique highly flexible and versatile for particular samples and methodologies, there is a lack of understanding regarding the most optimal methodology for microbial analysis. Parameters such as solvent composition, spray angle, and nebulizing gas pressure remain subject to debate^{172,174}. The selection of appropriate mass spectrometers, ionization sources, and data analysis methods is also crucial for accurate microbial identification¹⁷⁵. Furthermore, continuous development and expansion of well-curated databases containing representative mass spectra from various microorganisms are essential to enhance the reliability and coverage of DESI-MS in microbial identification¹⁷⁶.

Overall, DESI-MS is a powerful technique that could be used for the identification of microorganisms in the pharmaceutical industry. It offers several advantages for rapid and direct analysis of microbial samples, eliminating the need for extensive sample preparation and providing real-time results¹⁶¹. Although, it is important to consider limitations when utilising DESI-MS for microbial identification¹⁷⁷. Addressing these challenges through ongoing research and development will help to further improve the accuracy, sensitivity, and applicability of DESI-MS in microbial identification for the pharmaceutical industry and other relevant fields¹⁵⁵.

1.3.3.2.2 Direct analysis in real-time (DART)

The ambient ionisation approach known as direct analysis in real time (DART) ionisation was first developed in 2005 by Cody et al. and has since grown significantly in popularity. DART ionisation enables the direct analysis of sample surfaces of any size, shape, texture, or morphology by desorbing analytes from sample surfaces in its native conditions under ambient conditions using an ionising gas or aerosol. Without the requirement for sample preparation or chromatographic separation, mass spectra are generated in less than a second^{178–180}.

DART ionisation sources generate a corona discharge of helium gas between two ion-containing electrodes. The heated plasma is subsequently released from the ionisation source and creates reagent ions via a series of gas-phase interactions with ambient air, both of which are directed at a solid, liquid, or gaseous sample. Ionised analyte particles are produced when plasma and reagent ions collide with a sample surface, and they enter the mass spectrometer intake via a short reaction zone. The DART MS technique has demonstrated efficacy in diverse analytical domains, that includes amino acids, peptides, controlled substances, explosives,

pharmaceuticals, and oligosaccharides. Furthermore, similar to DESI, this method has been leveraged for the analysis of microorganisms^{181–183}.

DART-MS is a commonly used technique for the identification of microorganisms through the analysis of fatty acid methyl esters (FAME). FAME profiles are effective biochemical biomarkers for bacterial categorization and are analysed in reference laboratories for taxonomy research using gas-chromatography mass spectrometry (GC-MS)¹⁸⁴. Despite its reliability as a classification method, mass spectrometry analysis of FAME signatures is not commonly employed in microbiology laboratories due to the necessary expertise, time-consuming sample preparation, and the increasing availability of more recent, specialised, commercially available mass spectrometry identification systems¹⁸⁵. The utilisation of FAME profiling in conjunction with DART removes the necessity for an extensive chromatographic separation phase and has the capability to produce signatures within the 150–400 m/z range from whole cells of diverse microorganisms in a time frame of four minutes, in contrast to multiple hours required for GC-MS¹⁸⁶.

Overall, DART has made significant strides in microbial identification, offering a rapid and sensitive approach that has impacted various industries. Its ability to provide real-time results, high sensitivity, and species differentiation has paved the way for improved microbial detection and monitoring, benefiting sectors such as healthcare, food safety, environmental protection, and pharmaceutical manufacturing¹⁸⁷. Although, as a single standing technique, DART has its limitations, and it is often used in combination with other techniques and methods to overcome these limitations and provide more robust microbial identification and characterization¹⁸⁸. This causes limitations and reliability and in order to materialise a procedure that is reliable, additional work and optimization would have to be carried out^{182,189,190}.

1.3.3.2.3 Paper Spray Ionisation

Paper spray ionisation mass spectrometry, also known as PSI-MS, is a relatively new analytical technology that enables quick mass spectrometric detection of biological samples with very little or no sample preparation required^{191–193}. PSI-MS is a prospective option for future microbiological assays due to its rapid nature of analysis and low demand for sample preparation. One potential application of PSI-MS is in the detection of different species of bacteria. This novel application exhibits significant clinical potential as it has the ability to distinguish between pathogenic bacteria and their non-pathogenic naturally occurring equivalents, thereby streamlining clinical diagnostics and reducing associated costs^{194–198}.

PSI-MS is an ambient analytical technique that incorporates the direct electrospray analysis and detection of proteins, metabolites, and lipids in biological materials on a porous substrate, often cellulose paper^{199–201}. Paper is largely made up of compacted cellulose fibres that, when folded into sheets, generate an amorphous network of porous channels within the overall structure. Because of these channels, paper may absorb and transport solvent via capillary, which is why it has traditionally been used in chromatographic separations. This amorphous channel structure is essential to the paper spray process and distinguishes it from other forms of ambient ionisation. For quick, direct mass spectrometric analysis, a solvent is added to the substrate, followed by electrospray-like ionisation. The applied high voltage can be either positive or negative and typically varies between 2.0 and 5.5 kV^{202,203}.

The key advantages of PSI-MS are the small sample volume, minimum sample preparation requirements, and faster analytical time. Typical procedures last 30 seconds to 2 minutes, significantly less than the runtimes of liquid chromatography mass spectrometry (LC-MS) methods, which can last up to 20 minutes per injection for specific applications^{196,204–206}.

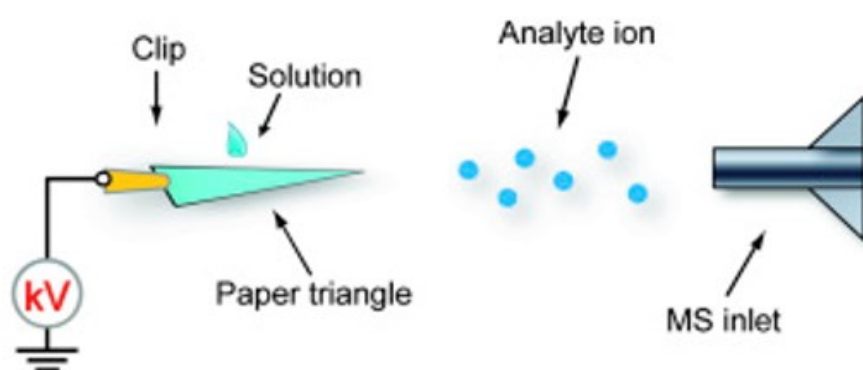


Figure 1.5 Schematic representation for paper spray analysis. The clip transmits applied voltage to the paper substrate, where the ionisation process takes place in an ambient environment. Adapted and reprinted with permission from McBride et al., 2019, see section *Permission to reprint*.

As with other MS techniques, not enough work on paper spray analysis of bacteria has been published, and while it shows potential for use in microbiology laboratories, further optimization studies are needed. The optimisation should include testing different porous media, as these influence ionisation rates, as do solvent mixture ratio²⁰⁷.

1.3.3.3 Rapid Evaporative Ionisation Mass Spectrometry (REIMS)

Rapid Evaporative Ionisation Mass Spectrometry (REIMS) was first introduced as a technology for analysing and identifying intact biological tissue *in vivo* and *in situ*. By merging MS with electrosurgical equipment, REIMS was at first developed exclusively for intraoperative tissue identification and margin measurement. Since then, applications have grown beyond surgery to include the characterization and identification of microbes, as well as the study of food and environmental samples^{208,209}.

Originally, REIMS was developed to provide rapid and accurate identification of biological tissues during surgical procedures^{210,211}. REIMS works by analysing the chemical composition of a sample through the detection of charged molecules or ions. The process involves the use of electrosurgical devices, such as an electrosurgical knife or scalpel, to generate an aerosol from the tissue being analysed. This aerosol is then transferred to the mass spectrometer, where the molecules are ionized and subsequently analysed based on their mass-to-charge ratio^{159,212,213}.

The advantage of REIMS is its ability to provide near-instantaneous analysis, enabling real-time decision-making during surgeries²¹⁴. By rapidly identifying different tissue types, including healthy and diseased tissues, REIMS can assist surgeons in distinguishing tumour boundaries, reducing the risk of leaving behind cancerous cells or damaging healthy tissue. REIMS has shown promising results in various surgical applications, including cancer surgery, where it can aid in precise tissue identification and improve surgical outcomes. It has the potential to enhance the accuracy and efficiency of surgical procedures by providing valuable molecular information in real-time^{215,216}.

While REIMS is primarily used in the medical field, it has also found applications in food analysis, pharmaceutical research, and other areas where rapid molecular identification is crucial^{217,218}. Ongoing research and advancements in the field of REIMS continue to explore its potential and expand its applications across different disciplines²¹⁹.

1.3.3.3.1 Applications of REIMS in microbiology

Initial studies to optimise the REIMS system for microbe identification investigated monopolar and bipolar diathermy capability of producing distinct MS patterns from intact bacteria²²⁰. Monopolar and bipolar diathermy devices are electrosurgical instruments that were initially used for surgical application development and then utilised to generate an aerosol from

microbial samples for analysis in the mass spectrometer. Both monopolar and bipolar diathermy devices were successfully used for microbial detection using REIMS by applying controlled thermal energy to the sample. The generated aerosol contained molecules from the microbial cells, which can then be ionised and analysed in the mass spectrometer for identification and characterization.

The vaporisation of bacteria placed directly on the counter electrode plate, on the other hand, resulted in significant charring and smoke on contact with the monopolar electrode, resulting in weak MS spectra and negatively compromising the robustness of the ion optics. The bipolar forceps were discovered to provide improved electrical current distribution and allowed for smaller amounts of bacterial biomass to be controllably vaporised, thus minimising contamination effects. At the same time, superior MS spectra with rich features and reduced fragmentation characteristics were detected, offering better intensity of molecular species above m/z 1000^{221–224}.

1.3.3.3.1.1 Laser Assisted Rapid Evaporative Ionisation Mass Spectrometry

Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry (LA-REIMS) is an advancement of the REIMS technique that incorporates laser technology into the ionization process. It offers improved spatial resolution, increased sensitivity, and enhanced versatility for molecular analysis^{225–228}.

In LA-REIMS, a laser beam is focused on the sample, leading to localized heating and vaporization of the target molecules. This laser-induced vaporization generates an aerosol, similar to the traditional REIMS method, which is then ionized and analysed by the mass spectrometer. The laser energy facilitates efficient and controlled desorption of the molecules, resulting in a higher concentration of ions for detection²²⁵.

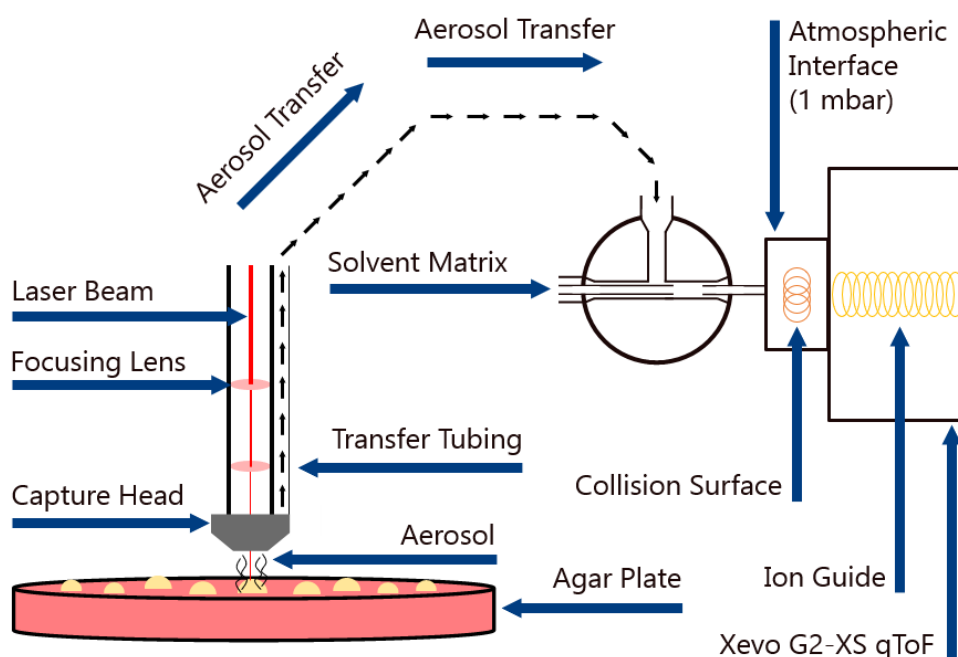


Figure 1.6 Overview of simplified high-throughput LA-REIMS setup for microbial classification analysis. Adapted with permission from Cameron et al. 2019, see section Permission to reprint.

The integration of lasers in REIMS brings several advantages. First, it enables more precise and selective sampling of specific areas or features within a sample, enhancing spatial resolution^{225,229}. This capability is especially useful in applications such as microbial identification, where detailed molecular information is required from specific regions of interest. Second, the increased sensitivity allows for the detection of low-abundance molecules, enabling the analysis of complex samples with higher accuracy and depth²³⁰.

The development of an automated high-throughput laser assisted REIMS (LA-REIMS) system has also been targeted in order to make the REIMS platform more suitable for a microbiology laboratory²³¹. In comparison to the monopolar and bipolar method, this simplified automated, and laser assisted procedure eliminates user input, resulting in technique variability and inconsistency. As such, it is well suited to microbiology laboratory workflows and analysis, and real-time identification taking only 15 seconds would allow analysis of around 3000 to 4000 colonies over a 24-hour period, representing a significant improvement over the other current identification systems^{227,230,232}.

This research projects main focus was the adaptation of the high-throughput LA-REIMS platform for the use of microbial identification for industrial pharmaceutical. The main advantage of using LA-REIMS in the pharmaceutical industry is its ability to provide near-instantaneous identification of microorganisms. This is particularly valuable in quality control processes, where the rapid detection of microbial contaminants or the identification of specific strains is essential to ensure product safety and adherence to regulatory standards^{226,233,234}.

Furthermore, automated LA-REIMS provides rapid results, significantly reducing the time required for microbial identification compared to traditional methods²²⁵. This allows for more efficient decision-making and timely implementation of appropriate measures in the production process. The technique also requires only a small amount of sample material, minimizing the impact on the sample integrity. This is particularly important when working with limited or valuable samples.

By implementing LA-REIMS for microbial identification in the pharmaceutical industry, companies can enhance their quality control procedures, streamline production processes, and ensure the safety and efficacy of their products. It provides a valuable tool for the rapid and accurate identification of microbial contaminants, enabling proactive measures to prevent product recalls and ensure consumer safety^{231,234–237}.

1.3.4 Spectroscopy based methods for microbial identification

1.3.4.1 FT-IR Spectroscopy

Fourier transform infrared (FTIR) spectroscopy, which was formerly employed to validate the identity of materials or chemicals, is now establishing itself as a tool for the quick separation and identification of microorganisms²³⁸. Spectroscopy is the study of molecular vibrations, in this case those stimulated by an infrared beam. The absorption of different molecules can be used as “fingerprint” and hence the spectrum indicates the sample's composition. Since the 1950s, IR spectroscopy has been used to analyse biomolecules such as proteins, nucleic acids, carbohydrates, and lipids^{239–241}. The spectral fingerprint of complex biological materials, such as whole cells, is the superposition of all vibrational modes of all molecules in the sample. The physiological status of the cells influences this pattern.

The FT-IR spectral analysis is straightforward: cells are suspended in solution, dried on a carrier, and spectra are obtained^{242,243}. However, the data analysis is complicated mainly due

to spectral overlap. To examine strain level differentiation, the fingerprint can be utilised as a comparison tool. Furthermore, similar to numerous other fingerprint-based systems, identification is accomplished by comparing an unknown microbe to a reference database. As a result, the reference spectral database is critical in the identification procedure²⁴⁴.

1.3.4.2 Raman Spectroscopy

Raman spectroscopy is another technique used for whole-organism characterisation with a spectroscopic fingerprinting²⁴⁵. Light can be transmitted, absorbed, reflected, or scattered when it interacts with materials. Scattered photons either have the same energy or frequency as the incident light (Rayleigh or elastic scattering) or have a different energy or frequency (Raman, Inelastic scattering or Brillouin). Photons exchange some of their energy with molecular vibrations in the substance, causing the shift. The energy amount correlates to specific molecular vibrations and hence the type of bonding.

The Raman spectra are collected and compared with reference spectra in a database after they are brought into the system^{246–248}. Combining Raman spectroscopy with optical microscopy improves sensitivity and resolution, allowing spectra to be recorded from a single cell. The interpretation of the spectra remains complicated, and reliable interpretation requires sophisticated analysis and huge databases. The identification process is influenced by sample treatment, physiological state of the cells, and measurement settings, just as it is by phenotypic approaches^{249,250}. As well as many other techniques, Raman is also affected by the underlying database's quality and the statistical algorithms utilised in the study. Development of this approach is still ongoing, limited by intrinsic technological challenges such as poor signal, long data acquisition times, strong background signals, high instrument costs, and a lack of automation in sample processing²⁵¹.

1.3.5 Summary of MS based microbial identification approaches in comparison to LA-REIMS

LA-REIMS (Laser Ablation Rapid Evaporative Ionization Mass Spectrometry) offers several advantages and unique features compared to other analytical techniques for microbial identification. Here is a comparison of LA-REIMS with methods discussed in this introduction:

LA-REIMS vs. MALDI-TOF MS (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry):

- Sample Preparation: LA-REIMS requires minimal sample preparation, while MALDI-TOF MS requires the deposition of a matrix on the sample.
- Sensitivity and Resolution: Both techniques offer high sensitivity and resolution for microbial analysis.
- Analysis Speed: LA-REIMS provides rapid analysis, comparable to MALDI-TOF MS.
- Coverage and Versatility: LA-REIMS can analyse a broad range of microorganisms, similar to MALDI-TOF MS.
- Identification Accuracy: Both techniques have high accuracy, but LA-REIMS has the advantage of non-destructive analysis, while MALDI-TOF MS requires sample transfer and matrix application.

LA-REIMS vs. DESI MS (Desorption Electrospray Ionization):

- Sample Preparation: LA-REIMS requires minimal sample preparation, while DESI may involve additional steps such as extraction and purification.
- Sensitivity and Resolution: Both techniques offer high sensitivity and resolution for microbial analysis.
- Analysis Speed: LA-REIMS provides rapid analysis, while DESI may require more time due to sample preparation steps.
- Coverage and Versatility: LA-REIMS can analyse a broad range of microorganisms, similar to DESI.
- Non-Destructive Analysis: LA-REIMS offers non-destructive analysis, whereas DESI involves direct sample transfer and sample contact with the ionization source.

LA-REIMS vs. DART MS (Direct Analysis in Real Time):

- Sample Preparation: LA-REIMS requires minimal sample preparation, while DART may involve direct sample contact with the ionization source.
- Sensitivity and Resolution: Both techniques offer high sensitivity and resolution for microbial analysis.
- Analysis Speed: LA-REIMS provides rapid analysis, comparable to DART.
- Coverage and Versatility: LA-REIMS can analyse a broad range of microorganisms, similar to DART.

- Non-Destructive Analysis: LA-REIMS offers non-destructive analysis, while DART involves direct sample transfer and sample contact with the ionization source.

LA-REIMS vs. PAPER SPRAY MS:

- Sample Preparation: LA-REIMS requires minimal sample preparation, while Paper Spray may involve additional steps such as extraction and purification.
- Sensitivity and Resolution: Both techniques offer high sensitivity and resolution for microbial analysis.
- Analysis Speed: LA-REIMS provides rapid analysis, comparable to Paper Spray.
- Coverage and Versatility: LA-REIMS can analyse a broad range of microorganisms, similar to Paper Spray.
- Non-Destructive Analysis: LA-REIMS offers non-destructive analysis, whereas Paper Spray involves direct sample transfer and sample contact with the ionization source

LA-REIMS vs. Spectroscopy Methods:

- Sample Preparation: LA-REIMS requires minimal sample preparation, while spectroscopy methods may involve additional steps such as extraction and purification.
- Sensitivity and Resolution: Both techniques offer high sensitivity and resolution for microbial analysis.
- Analysis Speed: LA-REIMS provides rapid analysis, comparable to some spectroscopy methods.
- Coverage and Versatility: LA-REIMS can analyse a broad range of microorganisms, similar to spectroscopy methods.
- Non-Destructive Analysis: LA-REIMS offers non-destructive analysis, while some spectroscopy methods may involve destructive sample preparation.

All of the aforementioned MS techniques share a common advantage, which is their reliance on the availability and comprehensiveness of reference spectral databases that are crucial for achieving precise microbial identification. Overall, LA-REIMS stands out due to its ability to rapidly analyse intact microbial samples without the need for any sample preparation, real-time analysis and automation making it a promising tool for microbial identification in various applications.

Chapter 2

Hypothesis and Aims

Chapter 2 Hypothesis and Aims

Hypothesis

The LA-REIMS platform, in conjunction with machine learning, provides a rapid and non-invasive approach of microbial classification that is performed directly on an agar plate without any sample preparation.

Aims

1. To develop and optimise protocols for the spectral data collection of various microbial species from various microbial plates and pharmaceutical assays using automated LA-REIMS platform.
2. To compile and validate a large-scale microbial species LA-REIMS spectral data base that includes numerous genus and species isolates.
3. To develop and validate stable, repeatable, and reproducible multivariate classification models capable of accurately identifying unknown bacteria, yeast and fungi species with high precision for three key microbial applications in the industrial pharmaceutical sector: environmental monitoring, detection of water pollutants, and identification of microbial presence in drugs or excipients.

Chapter 3

Materials and Methods

Chapter 3 Materials and Methods

3.1 Health and Safety

The Imperial College's Health and Safety requirements were followed during the LA-REIMS analysis (Bio1 form ID: IC-4987). All microorganisms and project samples were treated as Hazard Group II materials and handled within the containment of Class 2 facilities. LA-REIMS analysis was carried out in a custom-made Class 2 biological safety cabinet with an integrated laser safety interlock system, using necessary protocols to reduce the exposure risk of pathogens and containment of the Class 4 laser. The laboratory had decontamination and emergency procedures in place, and they were always followed in the event of spills or exposure. All hazardous substances were handled by following the Control of Substances Hazardous to Health (COSHH) regulations.

3.2 Microbiological sample storage

After bacteria, yeast and fungi purity isolates' biomass had reached a sufficient level of growth, it was removed from the petri plate using a sterile 10 μ L loop (VWR, Leicestershire, UK) and placed in cryovials glycerol broth (MicrobankTM, Pro-Lab Diagnostics, Wirral, UK) that were stored at -80°C. When required, isolates cryovials were removed from the freezer and cultured on a required agar plate by dipping a sterile loop and streaking.

3.3 Development of a custom LA-REIMS database

As part of a collaboration with AstraZeneca, samples used in this project came from routine quality control testing conducted by AstraZeneca. Isolate identification was provided by AstraZeneca prior to the shipment and validated on arrival using Sanger sequencing (see 3.4). AstraZeneca's identification methods were a combination of VITEK, MicroSEQ, and MALDI-TOF techniques, depending on the gram stain result. Identification results from sequencing and AstraZeneca's in-house results were compared and evaluated to select isolates to be used for further experiments and storage as species reference. Over 4000 samples were collected and analysed, with ~1500 chosen for the reference database. The final species reference database includes 206 bacteria, five yeast and 12 fungi species with multiple biological replicates. All final reference database species were analysed using the LA-REIMS platform to construct a custom REIMS spectral database, which was then used to build multivariate prediction models for microbial classification.

3.4 Culturing

Isolate culturing was completed depending on the type of the species and experimental design.

3.4.1 Environmental sample monitoring culturing

3.4.1.1 Small-scale model building

The small-scale environmental bacteria and yeast sample monitoring model was built at the start of the project as POC using in-house available media. Bacteria samples were cultured and incubated for 24 hours at 37°C, O₂ on 90 mm Columbian blood agar (CBA) (Scientific Laboratory Supplies Ltd) plates and yeast - at 35°C, for 48 hours, O₂ on 90 mm Sabouraud agar (SAB) (Scientific Laboratory Supplies Ltd) plates.

After initial incubation, samples were analysed within 8 hours using the LA-REIMS platform.

3.4.1.2 Large-scale model building

The large-scale environmental bacteria and yeast sample monitoring model was built using the finalised species reference library. All bacteria and yeast samples were cultured and incubated following AstraZeneca's in-house method for 24 hours at 37°C, O₂ on 90 mm Trypticase soy agar (TSA) (Scientific Laboratory Supplies Ltd) plates.

After initial incubation, samples were analysed within 8 hours using the LA-REIMS platform.

3.4.2 Water analysis culturing

Samples used for water analysis experiments were cultured using filter membranes and Reasoner's 2A (R2A) media (Scientific Laboratory Supplies Ltd). First, 200 mL of sterile water containing a water contaminant spike was filtered through a 47 mm diameter membrane (0.45 µm pore size) filter funnel (ThermoFisher Scientific Nalgene Analytical Test Filter Funnels). Secondly, filters were then washed with 100 mL of 150 mM Ammonium acetate to avoid salt contamination; the membrane was then removed, placed on a 90 mm R2A agar plate (Scientific Laboratory Supplies Ltd) and incubated for 24 hours at 37°C, O₂.

After initial incubation, samples were analysed within 8 hours using the LA-REIMS platform.

3.4.3 Specified organism culturing for absence or presence experiment

Three microorganisms were prepared following the step below for analysis on the LA-REIMS platform for microbial identification in the presence of a drug or excipient investigation.

Escherichia coli

Sample preparation and Pre-Incubation - 1 in 10 dilution of 1 g of the tablet to be examined was placed into Soybean–Casein Digest Broth (Scientific Laboratory Supplies Ltd) and incubated at 35°C for 24 hours.

Selection and Subculture - 1 mL of Soybean–Casein Digest Broth (Scientific Laboratory Supplies Ltd) and tablet mixture from the previous step was then transferred to 100 mL of MacConkey Broth (Scientific Laboratory Supplies Ltd) and incubated at 42°C for 24 hours and subcultured on a plate of MacConkey Agar (Scientific Laboratory Supplies Ltd) at 35°C for 24 hours.

Staphylococcus aureus

Sample preparation and Pre-Incubation - 1 in 10 dilution of 1 g of the nasal spray to be examined was placed into Soybean–Casein Digest Broth (Scientific Laboratory Supplies Ltd) and incubated at 35°C for 24 hours.

Selection and Subculture - Soybean–Casein Digest Broth (Scientific Laboratory Supplies Ltd) and spray mixture from the previous step was then subcultured on a plate of Mannitol Salt Agar (Scientific Laboratory Supplies Ltd) at 35°C for 24 hours.

Candida albicans

Sample preparation and Pre-Incubation - 1 in 10 dilution of 1 g of the ointment to be examined was placed into Sabouraud Dextrose Broth (Scientific Laboratory Supplies Ltd) and incubated at 35°C for 72 hours.

Selection and Subculture - Sabouraud Dextrose Broth (Scientific Laboratory Supplies Ltd) and ointment mixture from the previous step were then subcultured on a plate of Sabouraud Dextrose Agar (Scientific Laboratory Supplies Ltd) at 35°C for 48 hours.

After final incubation, samples were analysed within 8 hours using the LA-REIMS platform.

3.4.4 Fungi culturing

The fungi classification model required fungal species were cultured and incubated for 72 hours at 35°C, O₂ on 90 mm Sabouraud Dextrose Broth (SAB) (Scientific Laboratory Supplies Ltd) agar plates.

Fungal samples for the classification model were streaked out in the agar plate, where fungal samples (10 µL) for the imaging experiment were placed in the middle of the plate, allowing single colony growth. Fungi culturing time for the imaging experiment varied from 24 hours to 5 days for multiple time points in the spectral reference model due to the complexity of the species.

After initial incubation, samples were analysed within 8 hours using the LA-REIMS platform.

3.5 Fridge incubation

Since many existing analytical techniques are time and temperature sensitive, the effect of using refrigerated plates on LA-REIMS speciation was investigated. A small-scale environmental sample monitoring model was replicated four times, and incubated at the samples were incubated at the same conditions as before (3.4 for bacteria and yeast). After the initial incubation, one set was immediately analysed using the LA-REIMS platform, while the other three sets were stored in the fridge at a temperature of 4 °C for a period of 24, 48, and 72 hours, respectively, before being analysed according to their time in the refrigerator.

3.6 Isolate identification and gene sequencing

3.6.1 Sanger sequencing

All the isolates underwent Sanger sequencing targeting either the 16S rRNA gene for bacteria or the ITS region for yeast/filamentous fungi.

PCR products were prepared by an in-house developed DNA amplification workflow. First, isolates are sub-cultured in 96 deep-well plates (Greiner Bio-One Ltd) overnight at 37°C using 10 µL of isolate stock and 1 mL of either TSB (Fisher Scientific Lt) media for bacteria or SAB (Fisher Scientific Lt) media for yeast and fungus.

The supernatant is then removed after centrifuging the culture plate at maximum speed for 10 minutes at 4°C. Next, an obtained pellet is re-suspended in 100 µL PrepMan Ultra Sample

Preparation Reagent (Life Technologies Limited) and transferred to a non-skirted PCR plate before placing it into the Thermocycler for 15 minutes at 100°C. Non-skirted PCR plate is then centrifuged at maximum speed for 10 minutes at 4°C and 5 µL of obtained product is transferred to a new non-skirted plate. 45 µL of the mixture containing 2940 µL of MyTaq 2x (Meridian Bioscience), 235 µL reverse and 235 µL forward primer (Eurofins Genomics) and 1489.6 µL PCR grade water (Life Technologies Limited) were then added, and the plate is sealed. Table 3.1 summarises the primers used.

Thermocycler was programmed to a standard 3-step cycling profile for template denaturation, primer annealing and primer extension. 35 cycles of 94°C for 1 min for DNA denaturing, 55°C for 1 min for primers annealing with denatured target DNA and DNA synthesis at 72°C for 1 min were completed for each sample plate. LGC Genomics, a Berlin-based external provider, was used to sequence the obtained PCR products.

The obtained sequence was crossmatched using the EZBioCloud database for bacterial genomes and the NCBI-BLAST database for fungal genomes. The results were considered valid if the database sequence similarity rate was above 99%.

Table 3.1 PCR primers list for bacteria, yeast and fungi DNA amplification.

PCR primers list		
Bacteria	Forward, 27F	AGAGTTTGATCMTGGCTCAG
	Reverse, 1492R	GGTTACCTTGTTACGACTT
Yeast, Fungi	Forward, ITS1	TCCGTAGGTGAACCTGCGG
	Reverse, NL4	GGTCCGTGTTTCAAGACGG

3.6.2 Agarose gel electrophoresis

The final product was analysed on a 1% agarose gel and SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) to check if DNA amplification was successful and if enough product was made for further sequencing.

To prepare agarose gel, three tablets of agarose (Meridian Bioscience) were dissolved in 150 mL of TBE buffer (Thermo Scientific) by heating in the microwave and poured into the 96-well agarose mould. A mixture of 5 µL of final product and 1 µL of blue loading buffer was loaded to solidified agarose wells and run for 90 min at 80mA. HyperLadder™ 1kb (Meridian Bioscience) was used as a molecular weight marker for the final product.

3.7 LA-REIMS platform

3.7.1 Mass Spectrometry

Analyses of mass spectrometry were carried out using a Waters Xevo G2-XS QToF instrument (Waters Corporation, USA), which was connected to a REIMS atmospheric interface ion source (Waters Corporation, USA). Negative ion mode was selected, and data collecting was carried out via sensitivity scan mode and MS mode, respectively, across a mass range of 50-1200 m/z . Table 3.2 provides a summary of the MS instrument settings used throughout.

Table 3.2. MS instrument parameters used during LA-REIMS analysis.

Xevo G2-XS QToF Settings	
Ion mode	Negative
Scan mode	Sensitivity
Mass analyser	Time of Flight
Mass range	50-1200 m/z
Scan time	1000 ms
Resolution	23,000

3.7.2 High-Throughput LA-REIMS Analysis

A carbon dioxide laser (wavelength = 10.6 μm ,) (FELS-25A, OmniGuide) and a modified liquid handling robot (Freedom Evo 75, TECAN) equipped with a Pickolo platform (SciRobotics) for assisted colony selection were combined to perform high-throughput Laser-Assisted REIMS profiling on cultures.

The agar plates are placed on the rack, and a custom-written TECAN Freedom Evo 75 script (written by a previous post-doctoral research assistant) executes a command that permits the user to determine the number of culture plates to be analysed. The user can then view microbial colonies (using the Pickolo platform) and choose those of interest, or the application can automatically detect individual colonies for study. This project used only the manual selection mode for colony identification, and three colonies were chosen for each sample plate to serve as technical replicates.

Following the selection of colonies of interest, the CO₂ laser beam from an OmniGuide FELS-25® passes through a flexible hollow-core fibre guide system and focuses in a modified Aesculight® lens cell into a spot size of $\approx 500 \mu\text{m}$.

The rapid heating of the microbial biomass results in the formation of aerosol, which is aspirated through a 3D printed PTFE capture head and travels via a $\sim 1.5 \text{ M}$ long (1.5 mm inner diameter, 3.2 mm outer diameter) PTFE tube (Sigma-Aldrich) to the Xevo G2-XS Q-ToF instrument (Waters Corporation).

The analyte aerosol is then co-aspirated with a 10 ng/L solution of Leu-enkephalin (Sigma-Aldrich, Dorset, UK) in propan-2-ol (ROMIL Cambridge, UK) to be used as an internal standard for mass drift. The aerosol was aspirated using the mass spectrometer's vacuum, and a T-piece was used to introduce the aerosol and propan-2-ol/Leu-enkephalin solution at a flow rate of 0.15 mL/min via binary UPLC pump (ACQUITY, Waters Corp.).

Before entering the ion guide of the MS, the combined mixture reaches the REIMS interface of the Xevo G2-XS QToF® instrument and collides with a ribbon heated (700°C) collision surface for post-ionisation. A schematic of the automated high-throughput LA-REIMS platform is shown in Figure 3.1.

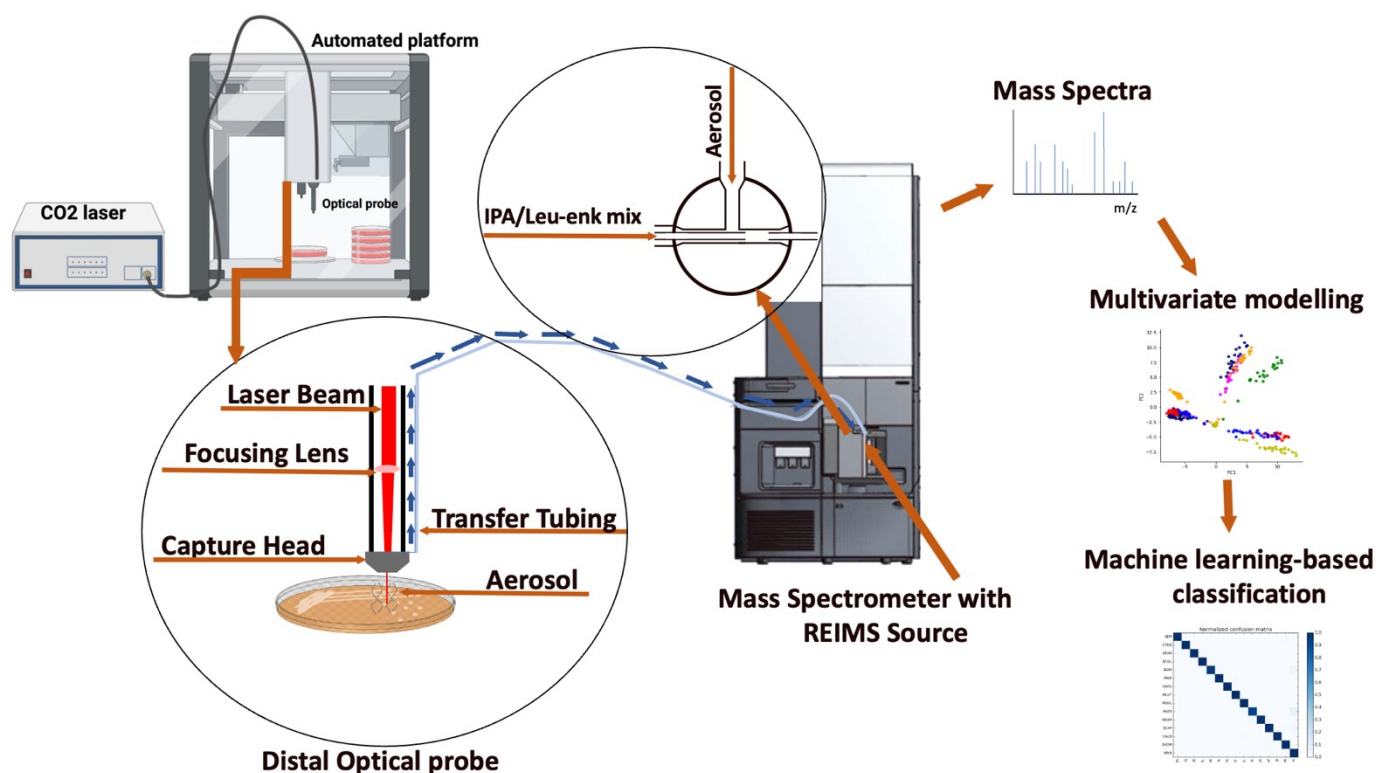


Figure 3.1 Automated Laser-Assisted REIMS platform (culture plate setup). Tecan liquid handling instrument was integrated with a CO₂ fibre optic laser system. Laser beam ablates the biomass, the generated aerosol travels into the MS via the transfer tubing. The resulting mass spectra are used for data analysis and multivariate modelling for machine learning-based classification.

The raw mass spectrometry data was subsequently pre-processed and exploited for the development of multivariate classification models.

3.7.3 Fungi low-resolution imaging data collection

Samples were run using Automated High-Throughput LA-REIMS platform with Xevo G2-XS qTOF mass spectrometer acquiring data in negative mode from 1mm spacing repeats over the whole colony area and using repeat laser pulsing on 40ms with super pulse, 30 PSI helium gas flow and 2 Watts power.

3.7.4 Laser power settings

Based on heuristics obtained from previous studies^{225,226,228}, Table 3.3 presents the laser power levels applied across all of the experiments. In super pulse mode, the laser pulses are delivered in a series of short bursts with high peak power, followed by a period of inactivity.

Table 3.3. OmniGuide laser settings used in LA-REIMS analysis.

OmniGuide laser power settings	
Bacteria samples	2W repeat laser pulsing on 40ms with a super pulse, 30 PSI helium gas flow
Yeast samples	2W repeat laser pulsing on 40ms with a super pulse, 30 PSI helium gas flow
Fungi samples	3W repeat laser pulsing on 40ms with a super pulse, 30 PSI helium gas flow

3.7.5 Calibration and Cleaning

The mass spectrometer was calibrated daily prior to analysis using 0.1 mM sodium formate (Sigma-Aldrich, UK) in negative ionisation mode at a flow rate of 0.25 mL/min for up to 5 seconds using a syringe pump (Elite Nanomite Programmable Syringe Pump, Harvard Apparatus) and a 5 mL Hamilton syringe. Calibration was performed across the 50-2400 m/z range using the manual Intellistart method through MassLynx, with pass parameters of <3 ppm RMS residual mass and >34 reference peaks detected. Following the calibration, propan-2-ol was injected into the system for approximately 30 minutes at a flow rate of 0.1 mL/min using a binary UPLC pump (ACQUITY, Waters Corp.). This was done to remove any leftover contamination from the calibration solution.

As part of regular maintenance, the instrument was vented once a week and the REIMS interface was removed to be cleaned. Capillary and the StepWave ion guide were also removed and sonicated in the ultrasonic bath using a 1:1 ratio HPLC grade water (VWR, Leicestershire, UK) and propan-2-ol (ROMIL Cambridge, UK). After the StepWave and interface were put back in place, the instrument was pumped to a vacuum again so it could be used the next day.

3.7.6 Leu- enkephalin preparation

The procedure below covers the preparation of a leucine enkephalin (LeuEnk) lockmass solution in 2-propanol solvent. The working concentration was 0.1 ng/μL and was used daily in data acquisition.

250 mL of Romil 2-propanol solvent was poured into a 250 mL measuring cylinder. The stock solution of LeuEnk was then removed from the -20°C freezer, and using a 10 μL pipette, 2.5 μL of stock solution was aspirated into the previously measured 2-propanol solvent and mixed well. From the measuring cylinder, lockmass preparation was transferred to a clean 250 mL Duran bottle, which was then connected to the mass spectrometer for data acquisition.

3.7.7 Quality control

Two isolates (*Escherichia coli* and *Micrococcus luteus*) were used as data and instrumentation quality controls. These samples were analysed using the LA-REIMS platform at the start and end of each experiment.

3.8 Data processing and statistical analysis

3.8.1 Peak picking pipeline

An in-house peak picking based workflow developed (by Dr. Alvaro Perdones-Montero, Dr. Yuchen Xiang, and Dr. James McKenzie, research associates at Takats' group at Imperial College London), (python and R-based) was used for the pre-processing of the mass spectrometric data. Pipeline was developed in-house by adapting Waters (MS manufacturers) peak picking algorithms that employs advanced centroiding algorithms that take into account factors such as peak shape, width, and intensity distribution to accurately pick peaks.

In summary, the raw data were first background subtracted and lock mass corrected by using Leu-enkephalin (negative ionisation = m/z 554.2516) as a standard. Denoising was then carried out using FFT (Fast Fourier Transform) operations, with which the high-frequency noises are removed in the Fourier domain.

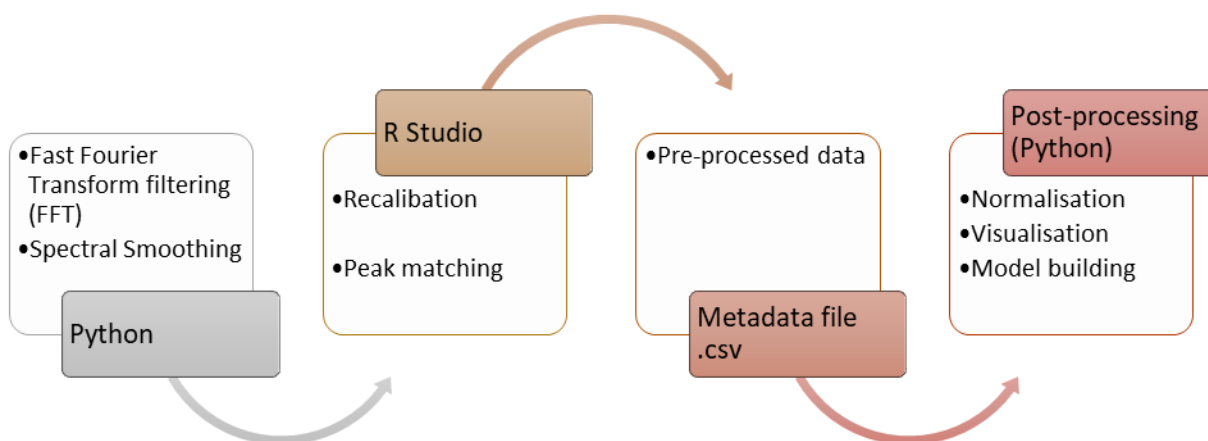


Figure 3.2. MS data pre-processing pipeline

The enhanced spectra were then pick peaked and mass recalibrated using functions available within the R package MALDIquant²⁵². Denoising, peak smoothing, peak detection, peak alignment, and recalibration were the key pre-processing processes in the pipeline (Figure 3.2). The resulting peak-picked data were lastly aligned between classes and projects and run to be used for post-processing after saving to .csv format. Here, data is normalised prior to the visualisation or further analysis using total ion count (TIC), log- correction, or both.

When dealing with mass spectrometry data, encountering a lot of zero values during raw data pre-processing is a common challenge that can affect further data analysis. Algorithms used in this pipeline thresholds signals below a certain intensity level to be filtered out resulting in zero values. Method of filtering out peaks based on their association with specific classes was used to reduce zeros within final dataset. Filter threshold used was 75%. During data normalisation, especially when log-correcting the data, algorithms with median log transformation, also known as the log-ratio transformation that involves adding a small constant to all values in the dataset before taking the logarithm was also used to handle the zero values.

3.8.2 1mm LA-REIMS imaging data analysis

Data analysis was performed using a modified in-house peak-picking script for area mapping in R studio part. The colony image was mapped by separating agar, edge, and the middle areas within the colony. The edge and middle of the colony area were analysed to obtain an intensity map/image using 100 selected features in regions 100-500 m/z and 600-900 m/z .

3.8.3 Multivariate data analysis

Multivariate analysis approaches are used in mass spectrometric data to investigate similarities and differences across multiple characteristics in large dimensionality datasets and are beneficial in datasets when discrimination based on a single variable is insufficient. Many features in the bacterial analysis are shared across species due to a very high degree of biological similarity between cell walls. Multivariate analysis is ideal for detecting relative changes between samples and the entire spectral range with high discriminatory power. Multivariate techniques are categorized as either supervised or unsupervised, with supervised tasks requiring user-defined inputs (e.g., class labels) and unsupervised tasks looking for data variances without taking into account any pre-defined class labels.

3.8.3.1 Post-processing

The .csv result from the pre-processing pipeline was used as input for statistical analysis. Data were analysed with an in-house, python-based pipeline (developed by Dr. Yuchen Xiang, research associate at Takats' group at Imperial College London) for data normalisation and unsupervised & supervised multivariate analysis. Here, Principal Component analysis (PCA), Support Vector Machine (SVM), Receiver Operating Characteristic curve (ROC), Recursive Feature Elimination (RFE), and univariate analysis functions were used and compared throughout for data analysis and interpretation, and their performance scores (see 3.8.3.7) will be examined, in order to reach a robust conclusion.

3.8.3.2 Principal Component Analysis

Principal component was used to reduce the data dimensionality and summarise the large datasets for easier visualisation and pattern recognition of multiple species analysed, preserving the information contained in the original dataset.

PCA is an unsupervised technique for visualising class differences that maps data onto a 2D plot by transforming to an orthogonal set of basis. As the basis vectors are known as principal components and are correlated to the variances in the data, with the greatest variance allocated to principal component one, the second greatest to component two, and so on. The number of components computed is determined by the amount of features, although later components are often mostly noise dominated; thus, only the top 3 components are considered in this work. PCA decreases the dimensionality of a dataset by removing low-relevance variables and grouping variables with high covariance, making it perfect for visualising complex large data

sets. Results from PCA are graphically displayed as a scatter plot (see example in Figure 3.3), where the scores of any two components can be visualised two-dimensionally, allowing observation of data trends and groups. Two points that appear close together on a PCA score plot imply that the data has a high degree of similarity or a low degree of variation. A loading plot is also provided, effectively showing the displayed components' spectral characteristics. The score and loading plots can already be used to identify some highly distinguishing features (biomolecules) that drive the separation of different bacterial/fungal classes in the principal components.

Because of its unsupervised nature, PCA is a non-biased approach, which is advantageous when looking for general REIMS data trends and possible adverse effects (e.g., batch effects, growth media differences, and experimental factors variations between samples). Because of this, PCA is employed as the first stage of the multivariate analysis.

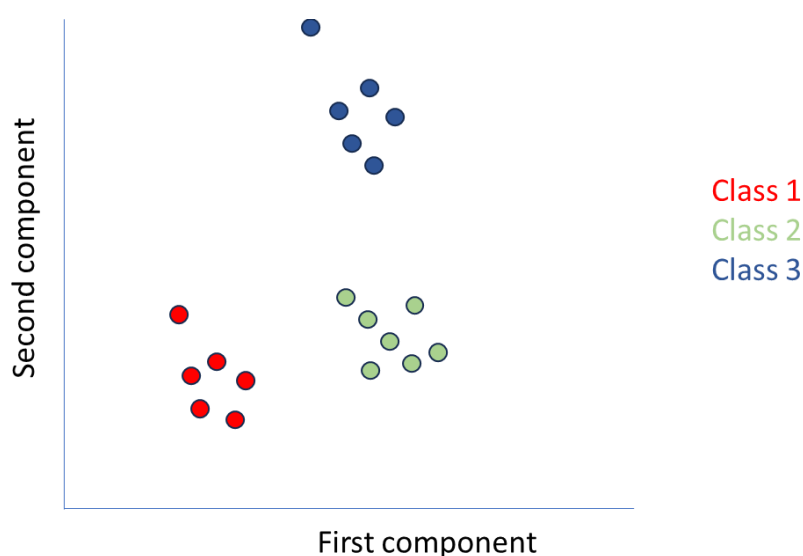


Figure 3.3 Example of PCA displayed as a scatter plot.

3.8.3.3 Support Vector Machine analysis

SVM is a non-linear supervised machine learning technique commonly used for spectral classification due to its high performance²⁵³. SVM performs complex data transformations

based on the kernel function specified and then seeks to optimise the maximum separation boundaries between data points based on the labels or classes defined.

In simple terms, SVM aims to create a decision boundary that maximizes the margin between classes. The data points closest to the decision boundary are called support vectors, hence the name "Support Vector Machine." These support vectors are crucial in defining the decision boundary and determining the class labels of new, unseen data points.

One of the key advantages of SVM is its ability to generalize well to unseen data and handle high-dimensional feature spaces. It is effective in dealing with small to medium-sized datasets. SVM is also known for its robustness against overfitting, as it aims to find the hyperplane with the maximum margin.

3.8.3.4 Data repeatability evaluation

Data variability evaluation was assessed by means of pairwise (Pearson) correlation between spectra from different species and the same modalities. MATLAB-based script written by Dr. Yuchen Xiang was used. Four sample sets were analysed during a four-day period was analysed, and the mean correlation coefficients were taken as a metric of similarity. The higher the correlation coefficient, the more repeatable the data between the biological and technical replicates of the same sample, analysed by the same modality. To visualise the results, clustered column plots of different species (X-axis) and the mean correlation coefficient (Y-axis) were generated for easy comparison.

3.8.3.5 Recursive Feature Elimination

The RFE algorithm was employed to optimise the model through feature refinement in a training dataset, with a focus on identifying the most relevant features for predicting the target variable. RFE identifies a subset of features significant for classification by starting with all of the features in the training dataset and removing features until the selected number is reached. Because this method needs manual input on the number of selected features, which can be biased, Recursive Feature Elimination with Cross-Validation (RFECV) can be used to determine both the number of and identity of most distinguishing features in a classification model. RFECV achieves this by employing normal recursive feature elimination to eliminate anywhere from zero to N features (where N is the total number of features), and then choosing the optimal subset of features based on the model's score by cross-validation. In this study SMV

based RFECV with leave one out cross validation was used. One feature was eliminated at each step of the feature selection process. The performance of the model is evaluated using the F1 score weighted by class frequencies. A new classification model was then created using RFECV's selected features and viewed to be the optimised model for a classification problem.

3.8.4 Performance evaluation

3.8.4.1 Stratified K-fold cross validation

Class size sensitive k-fold, also known as stratified k-fold, is a variant of the k-fold cross-validation technique that takes into account the imbalance in class distribution within a dataset.

In standard k-fold cross-validation, the dataset is divided into k subsets or folds, where each fold is used as a validation set while the remaining folds are used for training. However, this approach may lead to biased results when the class distribution is imbalanced, as certain classes may be underrepresented or even excluded in some folds. Class size sensitive k-fold addresses this issue by ensuring that each fold maintains the same class distribution as the original dataset.

By stratifying the dataset based on class proportions, class size sensitive k-fold provides a more reliable estimate of the model's performance, particularly for imbalanced datasets. It helps prevent overfitting or underestimation of the model's performance on minority classes, as each fold contains a representative sample from all classes²⁵⁴.

Because of the large sample quantity and class repeats, k-fold was chosen to be applied for small- and large-scale machine learning models in *Chapter 4*. K-fold that was used for small model building was 10 and 5 for the large model building due to the minimum sample count for the class being 9.

3.8.4.2 Leave one out (loo) cross validation

Another widely used approach for evaluating the performance of machine learning classification models is LOO. It is particularly useful when working with small datasets or when computational resources are limited. LOO cross-validation provides an unbiased assessment of the model's performance by iteratively leaving out a single data point as the validation set while using the remaining data points for training. LOO cross-validation ensures that every data point is used as a validation set exactly once, which provides a robust estimate of how the model will perform on unseen data. Since each iteration involves training a model

on a slightly different dataset, LOO can be computationally expensive for large datasets. However, it is less biased compared to other cross-validation techniques and can yield reliable results, especially in scenarios where the dataset is limited^{255,256}. LOO cross validation was used in Chapter 5 and 6 due to limited dataset sizes.

3.8.4.3 Independent data set validation

Understanding the importance of model validation and hence their robustness & generalisability, independent test sets are essential. During the course of this research, classification model validation was performed utilising a two-level validation approach. First, leave one out cross-validation (LOOCV) or k-fold was used to evaluate training model performance. Secondly, the model with a refined list of features was then validated using an independent validation set of 10 isolates for each species that were not included in the initial model training and generated at a different time point.

3.8.4.4 Confusion matrix

Results from supervised classification are generally assessed in the form of a confusion matrix which consists of important metrics for measuring the success of the classification:

- True positives, TP: microbial species predicted as their true label
- True negatives, TN: microbial species correctly not classified as a different microbial species
- False positives, FP: microbial species incorrectly classified as a different microbial species
- False negatives, FN: microbial species incorrectly misclassified as a different microbial species

Figure 3.4 illustrates a schematic of the TP, TN, FP, and FN meanings in a confusion matrix

Classes		Predicted label			
True label		TN	FP	TN	TN
		FN	TP	FN	FN
		TN	FP	TN	TN
		TN	RP	TN	TN

Figure 3.4. A multi-class confusion matrix describing the meanings of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

3.8.4.4.1 Performance scores

The performance of each model can, then, be assessed based on its²⁵⁷:

➤ **Accuracy:** $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

Accuracy measures the overall correctness of the model's predictions by calculating the ratio of correctly classified instances to the total number of instances. It provides an overall assessment of the model's performance.

➤ **Precision:** $precision = \frac{TP}{TP+FP}$

Precision calculates the ratio of true positive predictions to the total number of positive predictions made by the model. It assesses the model's ability to accurately predict positive instances.

➤ **Sensitivity:** $sensitivity = \frac{TP}{TP+FN}$

Sensitivity, also known as recall or true positive rate, measures the proportion of true positive predictions out of all actual positive instances. It indicates the model's ability to correctly identify positive cases.

➤ **Specificity:** $specificity = \frac{TN}{TN+FP}$

Specificity measures the proportion of true negative predictions out of all actual negative instances. It indicates the model's ability to correctly identify negative cases.

➤ **F1 Score:** $F1\ Score = \frac{2 \times (sensitivity \times precision)}{sensitivity + precision}$

The F1 score combines precision and sensitivity into a single metric by calculating their harmonic mean. It provides a balanced measure of the model's performance, mainly when dealing with imbalanced datasets.

The F1 score, which combines precision and sensitivity, was selected as the metric for the overall evaluation of classification models in order to account for the potential difference in class sizes.

3.8.4.5 ROC analysis

Receiver-Operating Characteristic (ROC) analysis was also carried out to evaluate the predictive models. Sensitivity (which relates to true positive rate) and specificity (which relates to true negative rate) are two key indicators of classification accuracy that demonstrate how well the model can distinguish between classes at a given operating point. An area under the ROC curve (AUC) indicates better model performance, with a value of 1 representing a perfect classifier and a value of 0.5 indicating random classifier²⁵⁸. Figure 3.5 illustrates schematics of ROC curve plot.

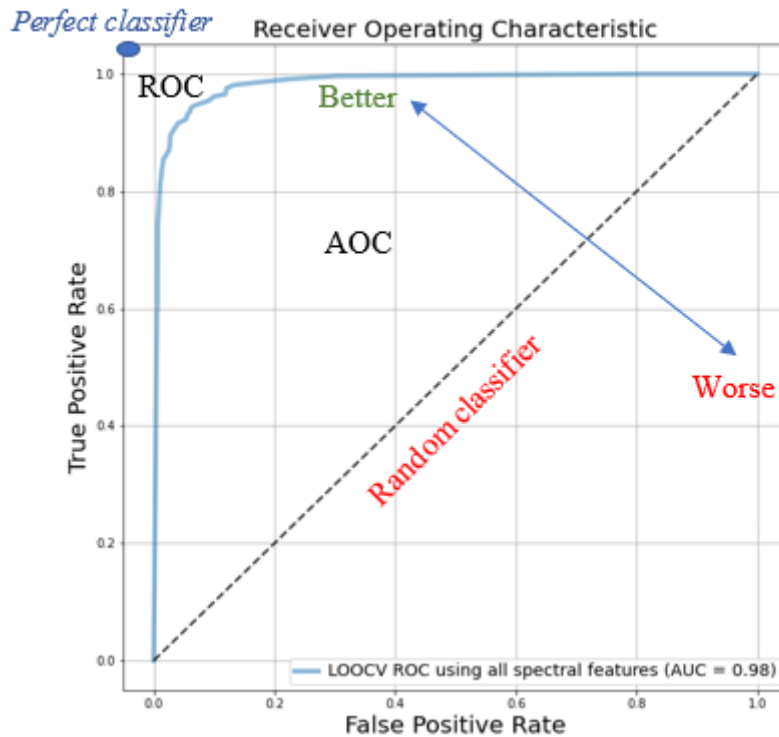


Figure 3.5. A schematics of receiver operating characteristic curve (ROC). AUC, area under the ROC curve.

3.8.4.5.1 Comparing ROC Curve and Confusion Matrix

The ROC curve offers a visual representation of the balance between the true positive rate and false positive rate, which is especially advantageous when adjusting the classification threshold. The selection of an optimal threshold can be facilitated by considering the desired balance between sensitivity and specificity.

On the other hand, the confusion matrix provides a more detailed breakdown of the classification results, including TN, TP, FP, and FN counts. From the confusion matrix, various performance metrics can be calculated, giving insights into the model's precision, recall, accuracy, and F1 score.

Both ROC curve and confusion matrix are important evaluation metrics in machine learning. The ROC curve is useful for understanding the model's overall discriminatory power and selecting the appropriate threshold, while the confusion matrix provides a more granular view of the model's performance across different classes and allows for the calculation of various performance metrics.

3.9 LA-REIMS Lipid and Metabolite annotation

Annotation of lipids and metabolites was carried out with the assistance of both in-house and online databases (SciFinder, LIPIDMAPS and METLIN), taking into consideration factors such as the accurate mass, elemental composition, isotopic ratio analysis, and chemical/biological importance. The REFCV specific feature list was used as a reference list for potentially biologically important features to be annotated.

First, ion annotation was carried out, giving special consideration to isotopes, adducts, and neutral loss fragments that may have originated from the same molecule of interest. The most common adducts in LA-REIMS negative mode ionisation are M-H, M+Cl, and M-H₂O-H. Furthermore, visual evaluation of the data is also performed in LA-REIMS analysis, with special emphasis devoted to isotope patterns and possible adducts apparent in the raw spectrum. For the M+Cl adduct, for example, special consideration was given to the natural isotope distribution pattern of chlorine, ³⁵Cl/³⁷Cl in a 3/1 ratio must be present.

Following ion annotation, the monoisotopic masses were calculated based on adducts and inputted to an online tool (Chemcalc) to generate a list of potential chemical formulas. Using the retrieved chemical formula list, SciFinder was used to create metabolite matches, which were then cross-checked in the LipidMaps database with the best ranking matches, based on ppm error value (5ppm). The mass accuracy error was determined in ppm using the equation below.

$$\Delta m^i \text{ in ppm} = \frac{(m^i - m^a)}{m^a} \times 10^6$$

m^i = Experimental mass, m^a = Theoretical exact mass, Δm^i = Mass deviation in ppm

3.9.1 Tentative ID confirmation using MS/MS

In order to verify the findings, REIMS MS/MS was performed on the metabolite matches that were uncovered throughout the annotation process.

The target biomarker m/z was entered into the quadrupole mass filter parameters, and collision energy of 35 eV was chosen. The optimal collision energy was determined by adjusting the value until the relative intensity of the precursor ion was approximately 1/3 of the highest intensity product ion. The collision chamber was supplied with argon gas.

3.10 Untargeted LC-MS assay

LC-MS analysis was performed alongside REIMS for the goal of validating specific feature findings.

3.10.1 Extraction method

For the lipid and metabolite extraction of the microbial species biomass, 500 µL of chilled 1:4 (Water:Propan-2-ol, Optima™ LC/MS Grade, Fisher Chemical™, Fisher Scientific) mix was added to the 140 mg of microorganism in the pre-chilled Eppendorf tube (Eppendorf UK Limited). Beads were then added to the sample and water:propan-2-ol mix and homogenised at 6000 rpm, 2x45 seconds, placed on ice for 5 minutes and homogenised again for 2x45 seconds. Samples were then vortexed for 1 minute and placed at 4°C for protein precipitation overnight. The samples were then centrifuged at 10,300 rpm for 10 minutes at 4°. The supernatant was transferred into a Waters Polypropylene Screw Neck Vial, 12 x 32 mm, 300 µL HPLC vial (Waters Corporation, UK) for LC-MS analysis. Quality control and blank samples were prepared to monitor the performance of the instrument throughout the analysis. The microbial species extracts were analysed by the National Phenome Centre (NPC) reverse phase lipid profiling method with a C8 column (Waters Acquity UPLC BEH C8, 1.7 µm, 2.1 x 100 mm. For the method used, please refer to the 3.10.2 paragraph.

3.10.2 RPC UPLC-QTOF Analysis of Lipids

LC-MS analysis was performed using an ACQUITY UPLC system (Waters Corp.) coupled to a Xevo G2-XS QToF mass analyser (Waters Corp.) using an LC-MS lipidomics assay developed and validated by the National Phenome Centre (NPC, Imperial College London)²⁵⁹. Mobile phase A consisted of H₂O/IPA/ACN in a ratio of 2:1:1 (v/v/v), with a buffer system of 5 mM ammonium acetate, 0.05% acetic acid and 20 µM phosphoric acid (Sigma-Aldrich). Mobile B consisted of IPA:ACN in ratio of 1:1 (v/v), 5 mM ammonium acetate (Sigma-Aldrich) and 0.5% acetic acid (Sigma-Aldrich). Please see Tables 3.4, 3.5, 3.6 and 3.7 below for detailed instrument parameters, MS source parameters, MS instrument parameters and gradient used during the analysis.

Table 3.4 NPC LC instrument parameters.

Variable	Description
Mobile Phase A	Water:isopropanol:acetonitrile 2:1:1, 5 mM ammonium acetate, 0.05% acetic acid, 20 µM phosphoric acid
Mobile Phase B	Isopropanol:acetonitrile 1:1, 5 mM ammonium acetate, 0.05% acetic acid
Seal Wash	Isopropanol:water 1:9, set to run every 0.5 minutes
Weak Wash	Water:isopropanol 1:4
Strong Wash	Isopropanol
Lockspray	Leucine enkephalin 600 pg/µL (prepared in water:acetonitrile 1:1)
Column	Waters Acquity UPLC BEH C8, 1.7 µm, 2.1 x 100 mm
PEEK tubing	Column to sample probe: ID = 0.004 inch, length = 38 cm
Column Temp.	55 °C
Sample Loop	Positive: 1 µL, negative: 2 µL
Injection Volume	15 µL
Run Time	13.25 minutes (N.B. including auto gain each injection will be ~15 minutes)
Autosampler Settings	Syringe: 25 µL Air volume, pre-cleans, filling strokes = 0, filling & injection speed = 5 µL/sec, inject to = LC Vlv1, pre-injection delay = 500 ms, post-injection delay = 500 ms, inject 10 µL to injector, followed by 5 µL to waste Needle wash sequence = 1x weak, 1x strong, 1x weak, wait time = 705 sec, Valve wash sequence = 1x weak, 2x strong, 1x weak
Sample Cooler	4 °C with low nitrogen flow to prevent condensation

Table 3.5 NPC MS source parameters.

Variable	Description
MS Function	ToF MS
Run time	13.15 minutes
Analyser Mode	Sensitivity
Dynamic Range	Extended
Mass Range	50 – 2000 Da
Scan Time	0.1 sec
Data Format	Centroid
Lockspray function Settings	Acquire lockspray & apply correction Select appropriate Lock Spray file Scan time = 0.15 sec, interval = 60 sec, scans to average = 4, Mass window = +/- 0.5
Method Events (enabled)	At initial conditions lockspray is infused, at 12.05 minutes lockspray is refilled, at 12.45 minutes lockspray starts infusing again

Table 3.6. NPC MS instrument parameters.

Variable	Description
Sample Probe Angle	7 mm
Capillary Voltage (kV)	Positive: 2.0 / Negative: 1.5
Sampling Cone	25
Source Offset	80
Source Temperature (°C)	120
Desolvation Temperature (°C)	600
Cone Gas Flow (L/hr)	150
Desolvation Flow (L/hr)	1000
Collision Energy (V)	Off but with default set to 6
Gas	API = On, Collision = On
Detector Auto Gain	On
Auto Gain Controls	Custom auto gain check = ticked, fluidics = tune page, detector check time = 70
Stepwave 2 Offset	15
Quad Profile	Auto
Lockspray Flow (µL/min)	15
Lockspray Capillary Voltage (kV)	Positive: 3.0 / Negative: 2.0

Table 3.7 NPC LC gradient list.

Gradient					
#	Time (min)	Flow (mL/min)	% A	% B	Curve
1	Initial	0.6	99.0	1.0	Initial
2	0.10	0.6	99.0	1.0	6
3	2.00	0.6	70.0	30.0	6
4	11.50	0.6	10.0	90.0	6
5	12.00	1.0	0.1	99.9	6
6	12.50	1.0	0.1	99.9	6
7	12.55	0.9	35.0	65.0	6
8	12.65	0.8	70.0	30.0	6
9	12.75	0.7	99.0	1.0	6
10	12.95	0.6	99.0	1.0	6
11	13.25	0.6	99.0	1.0	6

The National Phenome Centre (NPC) carried out the LC-MS analysis and provided the raw data files as well as the global profiling (XCMS) data of all the acquired spectra. Data analysis was then carried out by the candidate by using MassLynx v4.1 (Waters, Wilmslow, UK) software. The MS mode data was manually cross-checked with LA-REIMS annotated features of interest list. Features that were presented in both LA-REIMS and LC-MS were validated using LC-MS/MS and Data Dependent acquisition (DDA) mode.

Chapter 4

Environmental Sample Analysis (bacteria and yeast)

Chapter 4 Environmental Sample Analysis (bacteria and yeast)

4.1 Introduction

Pharmaceutical environmental monitoring refers to microbiological tests performed to detect changing trends in microbial counts and microflora growth within clean rooms or controlled settings performed to understand the environmental control within manufacturing settings, including clean rooms and critical zones. The compiled findings offer insights into the efficacy of the heating, ventilation, and air-conditioning infrastructure, the hygienic practises of the staff, the equipment's state of cleanliness, and the sanitation protocols employed. Strict microbial control with rapid and accurate detection of microorganisms is an essential aspect of pharmaceutical analysis to maintain high pharmaceutical standards¹⁵.

The identification of microorganisms presently involves the use of a number of different phenotypic, genotypic, and prototypic approaches, each of which requires some type of sample preparation. Techniques for quick high throughput direct no sample preparation analysis in pharmaceutical environment is one of the biggest gaps in the market (please refer to *Chapter 1* for existing methods and techniques). The introduction of mass spectrometry- based microbial identification has been trying to revolutionize microbiology laboratories for years now. The introduction of MALDI-ToF/MS has reduced identification times to minutes, although still requiring the preparation of a matrix to aid ionisation. Furthermore, additional extraction processes are required for some microbial groups, such as yeasts, in order to get precise species level identifications. Other current identification techniques (e.g. VITEK, MicroSEQ) all require purity plates, this step not only presents a burden on the efficiency of activities in the pharmaceutical micro lab, but also impacts important considerations such as data integrity & sample traceability as well as a significant impact to sustainability considerations – such as excessive single use plastic burden and the reliance of a natural resource (the burden on seaweed used to produce agar could be reduced if we significantly reduce the number of plates required).

The work demonstrated here using REIMS (Rapid Evaporative Ionization Mass Spectrometry) has the potential to completely change microbial identification for the better as it was performed in real time without any sample pre-processing or need of purity plates. Therefore,

this chapter's objective was to determine the applicability of LA-REIMS for environmental monitoring as a sample preparation free, direct-from-petri-dish microorganism identification platform. Due to the complexity of fungi (their analysis is discussed in detail in *Chapter 5*), this chapter will solely focus on bacteria and yeast analysis.

4.2 Aims of the chapter

The aim of this chapter was to develop and optimise a robust and reliable predictive classification model using an in-house generated industrial, pharmaceutical bacteria and yeast microbial contamination spectral database, which may be exploited as a machine learning-based microbial identification pipeline in combination with the LA-REIMS platform. Utilising a small-scale database, the optimum parameters and sample handling conditions will be researched and applied to develop a large-scale classification model for unknown microbial species.

4.3 Current industrial pharmaceutical environmental monitoring pipeline

For full identification of microorganisms, AstraZeneca's current environmental sample monitoring timelines can take 8-18 days to obtain identification from the point of sample collection. After collecting the samples, the plates are first incubated for five days. After which the number of colonies is manually entered into a computer, and if growth is observed above alert or action levels, the colony is sub-cultured onto a fresh purity plate and incubated for another 24 hours (in the case of moulds, this may take another five days). Gram staining is done after incubation. Depending on the Gram stain result, the sample is prepared for identification on VITEK or Micro-SEQ. The result is read and, if acceptable, manually entered into an electronic system; if the result is poor, the entire process is repeated.

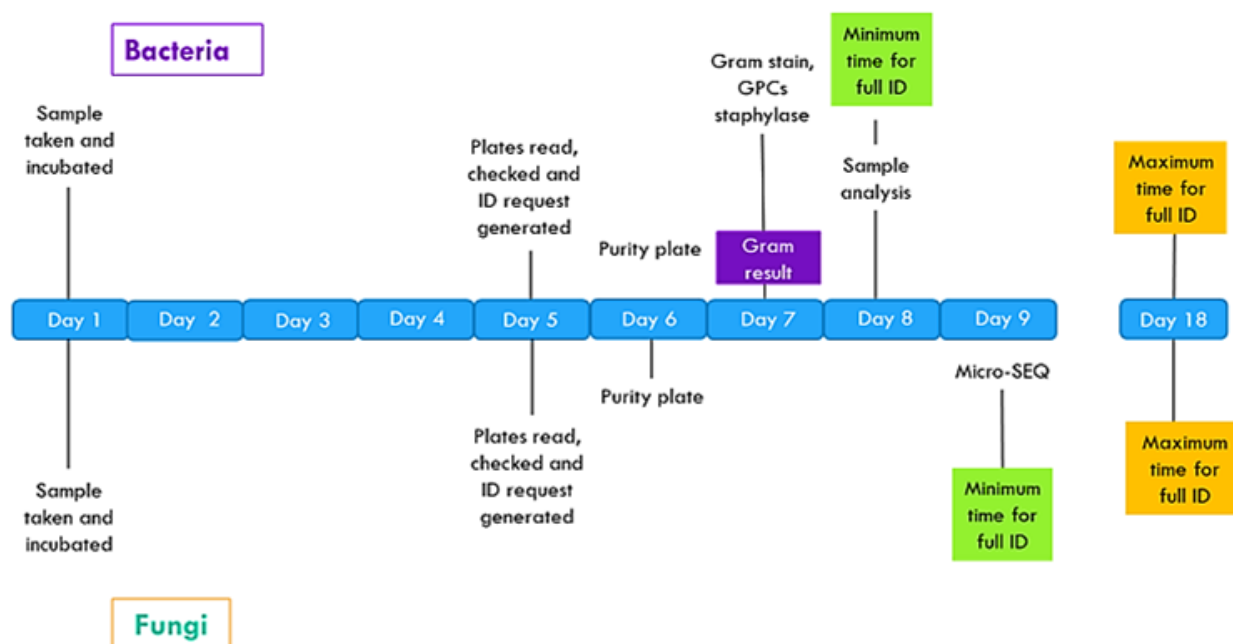


Figure 4.1 Current steps and timeline for identifying bacteria and fungus in the pharmaceutical industry's environmental monitoring.

4.3.1 Potential LA-REIMS pipeline for environmental sample monitoring

In the proposed pipeline for environmental monitoring with LA-REIMS platform, samples from important locations and personnel would be collected on various-sized agar plates. On the base of the plates are labels with bar codes that identify the operator, location, date, and time of the sampling. Following sample collection, one to five days of incubation on plates. Plates would be then fed into the LA-REIMS platform, and a technician would read them on a computer using a Piccolo platform before choosing all the colonies for identification; picture of plate image is captured for colony forming unit (CFU) count. Colonies are then identified using a predictive modelling-based approach using our in-house microbial species spectral database. Results are then linked to the bar code to provide identification results for the CFU and specific sample.

By analysing samples directly from the initial sampling plate using LA-REIMS after the initial incubation, this would be able to significantly cut down on the amount of time and cost needed for this process. The proposed time for the process would be one to five days only, in comparison to the current 18 days' timeline.

4.4 Results and Discussion

4.4.1 Small-scale multivariate model

A total of 524 isolates consisting of 419 bacterial and 105 yeast isolates representing 12 different bacterial and three yeast species were selected for analysis. The samples included in the study were a mix of widely seen pharmaceutical, environmental bacteria and yeast contaminants. For each of the 15 organisms, 35 isolates were tested, with 25 being randomly selected for predictive model building and 10 serving as an independent validation dataset. A 15 species small spectral database (Table 4.1) was then used to investigate:

- pre- and post-processing parameters for multivariate model building
- sample refrigeration effect on multivariate model building
- multivariate model reproducibility
- multivariate model stability
- multivariate model repeatability

The conclusions derived from the results obtained using the small-scale database will be used to scale up the multivariate model.

Table 4.1 Microbial species list for small scale multivariate model building. B-bacteria, Y-yeast.

Small scale multivariate model species list			
<i>Micrococcus luteus (MLUT)</i>	B	<i>E.coli (ECOL)</i>	B
<i>Staphylococcus capitis (SCAP)</i>	B	<i>Citrobacter koseri (CKOS)</i>	B
<i>Staphylococcus epidermidis(SEPI)</i>	B	<i>Proteus mirrabilis (PMIR)</i>	B
<i>Staphylococcus hominis (SHOM)</i>	B	<i>Pseudomonas aeruginosa (PAEU)</i>	B
<i>Staphylococcus warneri (SWAR)</i>	B	<i>Candida albicans (CALB)</i>	Y
<i>Bacillus simplex (BSIM)</i>	B	<i>Candida parapsilosis (CPAR)</i>	Y
<i>Kucuria rhizophila (KRHI)</i>	B	<i>Candida tropicallis (CTRO)</i>	Y
<i>Moraxella osloensis (MOSL)</i>	B		

4.4.1.1 Optimisation of Pre-processing parameters

In both the pre- and post-processing of MS data, there are numerous stages in which different parameters may be altered for optimisation to achieve the multivariate model's maximum accuracy and reliability.

The main stages, such as data filtering, peak matching and data normalisation from the peak-picking pipeline (Section 3.8.1), were selected for further investigation. Table 4.2 explains the role of each variable that was changed and the order in which they were combined. The models were built using support vector machine classification (SVM). Accuracy was measured using the F1 score and the ROC for cross-validation using a stratified K-fold (to be referred as K-fold from now on) and independent validation set. Please refer to “Chapter 3, Materials and Methods” for details on methods used to evaluate the models.

Table 4.2. MS data pre- and post-processing variables and parameters

MS data pre- and post-processing variables and parameters		
<u>Python: pre-processing</u>	<u>R Studio: peak picking/filtering</u>	<u>Python: data normalisation</u>
‘Pre-processing 1’: average scans and filter noise using FFT with more levels of noise detection.	‘75% peak filter’: 75% of samples in a class has to detect the peak to be included.	‘TIC’: Total ion count. All ions from metabolites is divided by the total number of ions observed in the sample
‘Pre-processing 2’: average scan but doesn’t filter noise	‘No peak filter’: no peaks will be excluded in this step.	‘Log’: median logarithmic transformation of the values, which transforms them onto a scale that approximates
		“TIC+Log”: TIC followed by log transformation

Multivariate model

```

graph TD
    A[Multivariate model] --> B[Pre-processing 1]
    A --> C[Pre-processing 2]
    B --> D[No peak filter]
    B --> E[75% peak filter]
    C --> F[No peak filter]
    C --> G[75% peak filter]
    D --> H[TIC]
    D --> I[Log]
    E --> J[TIC]
    E --> K[Log]
    F --> L[TIC]
    F --> M[Log]
    G --> N[TIC]
    G --> O[Log]
    H --> P[TIC+Log]
    I --> P
    J --> Q[TIC+Log]
    K --> Q
    L --> R[TIC+Log]
    M --> R
    N --> S[TIC+Log]
    O --> S
    
```

First, tables 3 and 4 below compare when ‘Pre-processing 1’ and ‘Pre-processing 2’ were used. When “‘Pre-processing 1’” was applied, both ‘No peak filter’ and ‘75% peak filter’ with the ‘TIC’, ‘Log’, and ‘TIC+Log’ normalisation indicated the decreased performance of the training model and independent validation.

Secondly, after comparing ‘No peak filter’ and ‘75% peak filter’ variables with ‘Pre-processing 2’ using three different normalisation combinations, the top two combinations with the highest overall F1 performance were chosen; ‘No peak filter’ with ‘Pre-processing 2’ and “Log” normalisation (Table 4.3) with an overall F1 accuracy of 98.02% for the independent validation set, as well as ‘75% peak filter’ with ‘Pre-processing 2’ and ‘TIC+Log’(Table 4.4) with an overall F1 accuracy of 100%.

Table 4.3. Pre- and post-processing two-step validation results using “No peak filter”, ‘Pre-processing 1’, ‘Pre-processing 2’, and three different normalisation variations. The best combinations with an F1 accuracy of 98.02% and ROC=1 was ‘No peak filter’ with ‘Pre-processing 2’ and “Log” normalisation parameters.

	Pre-processing 2’			Pre-processing 1’		
	“No peak filter”			“No peak filter”		
	TIC	LOG	TIC+LOG	TIC	LOG	TIC+LOG
Model Kfold validation_F1 score	27.52%	96.27%	96.54%	31.90%	94.97%	96.29%
Model ROC-AUC average	0.88	1	1	0.92	1	0.98
Independent validation_F1 score	51.18%	98.02%	98.00%	54.49%	96.03%	97.33%
Indep validation ROC-AUC average	0.87	1	1	0.93	1	0.99

Table 4.4. Pre- and post-processing two-step validation results using ‘75% peak filter’, ‘Pre-processing 1’, ‘Pre-processing 2’ and three different normalisation variations. The best combinations with an F1 accuracy of 100% and ROC=1 was ‘75% peak filter’ with ‘Pre-processing 2’ and ‘TIC+Log’ normalisation parameters.

	Pre-processing 2’			Pre-processing 1’		
	‘75% peak filter’			‘75% peak filter’		
	TIC	LOG	TIC+LOG	TIC	LOG	TIC+LOG
Model Kfold validation_F1 score	35.21%	98.39%	99.73%	35.94%	97.84%	99.73%
Model ROC-AUC average	0.92	0.98	1	0.76	0.99	1
Independent validation_F1 score	68.09%	100%	100%	70.86%	100%	100%
Indep validation ROC-AUC average	0.74	1	1	0.92	1	1

4.4.1.1.1 Recursive Feature Elimination cross-validation selection (RFECV)

The top two best-performing pre-processing parameters were then compared using a third-stage validation step. Recursive feature elimination cross-validation (RFECV) was used to extract statistically significant features, which were then used to build features-only classification models to confirm classification model accuracy and performance and to select the best-performing pre-processing option.

To develop the model and validate its performance, data matrices containing only the RFECV-selected features were first generated, which consisted mainly of phospholipids and some fatty acids.

First, a comparison was carried out between the top two pre-processing options using the RFECV-derived feature number. Whereas the ‘75% peak filter’ based classification matrix produced 188 features, the ‘No peak filter’ based matrix only provided 84 RFECV features. Training models were created for both parameter choices using those features, and then those models were validated using independent validation sets.

Table 4.5 summarises the average accuracy achieved by the training model and independent validation set.

Table 4.5. Results from the top two pre-processing parameter combinations using third-stage validation of RFECV features. The model accuracy obtained using only specific features was greatest when ‘75% peak filter’ with ‘Pre-processing 2’ and ‘TIC+Log’ normalisation was used.

	“No peak filter”	‘75% peak filter’
	‘Pre-processing 2’	‘Pre-processing 2’
RFECV model normalisation	LOG	TIC+LOG
RFECV feature number	84	188
RFECV feature model Kfold validation_F1 score	97.91%	99.62%
RFECV feature independent validation_F1 score	100%	100%

The RFECV-optimised machine learning model produced using ‘75% peak filter’ yielded an overall F1 accuracy of 99.62% for K-fold cross-validation and 100% for the independent validation set validation (Figure 4.2), outperforming the RFECV feature model built using ‘No peak filter’ parameter.

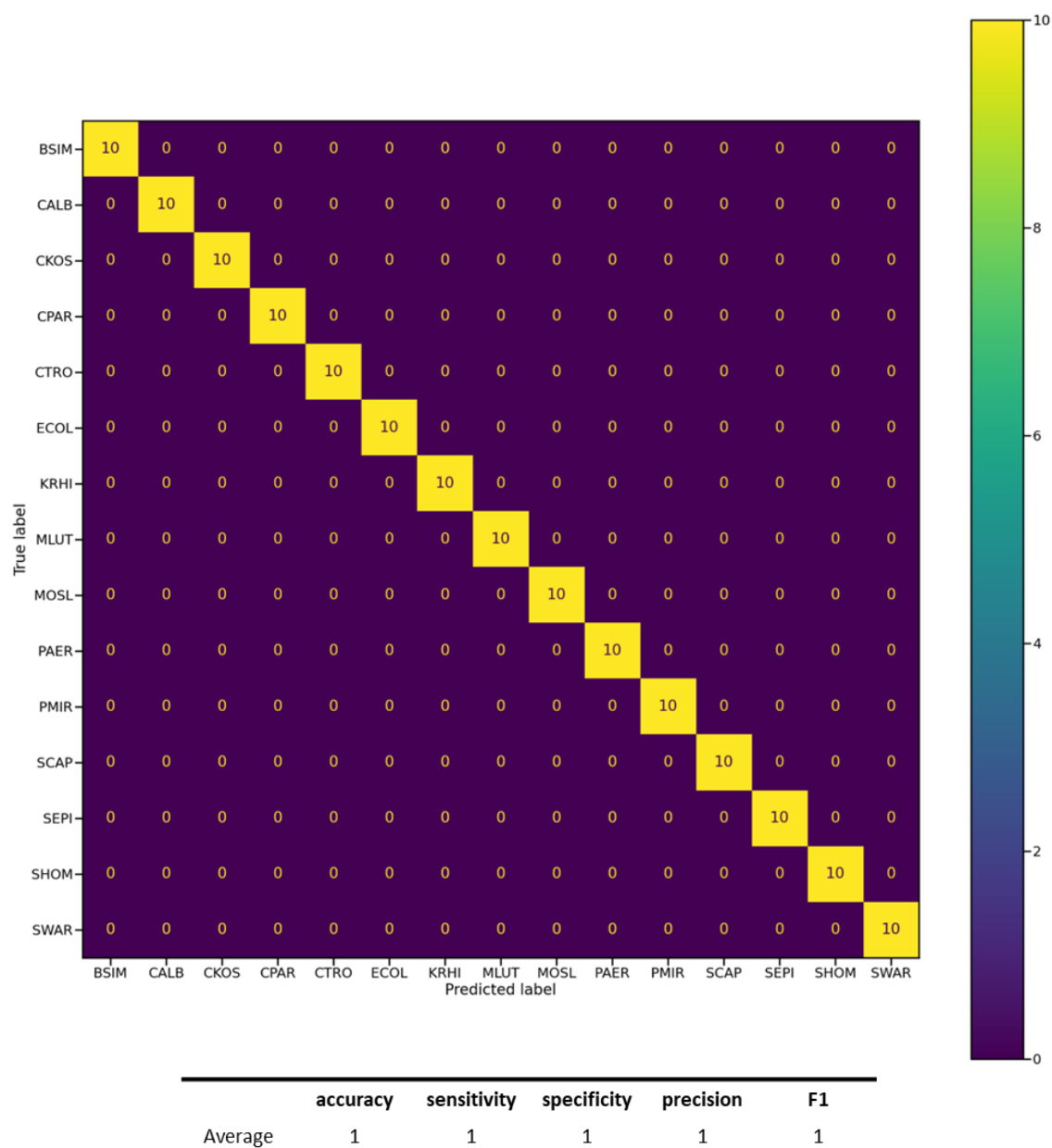


Figure 4.2. Confusion matrix and average performance based of ‘75% peak filter’ RFECV features only model validation using independent validation set, with an overall species-level F1 accuracy of 100%.

One of the RFECV selected features, 745.517 m/z (PG (34:2)) from an ‘75% peak filter’ generated matrix, is shown in Figure 4.3 below. This example demonstrates a significant difference in normalised intensity between *Pseudomonas aeruginosa* and the other species in the model. Therefore, the better model performance could be associated with a higher number (188 vs 84) of the species-specific features using ‘75% peak filter’.

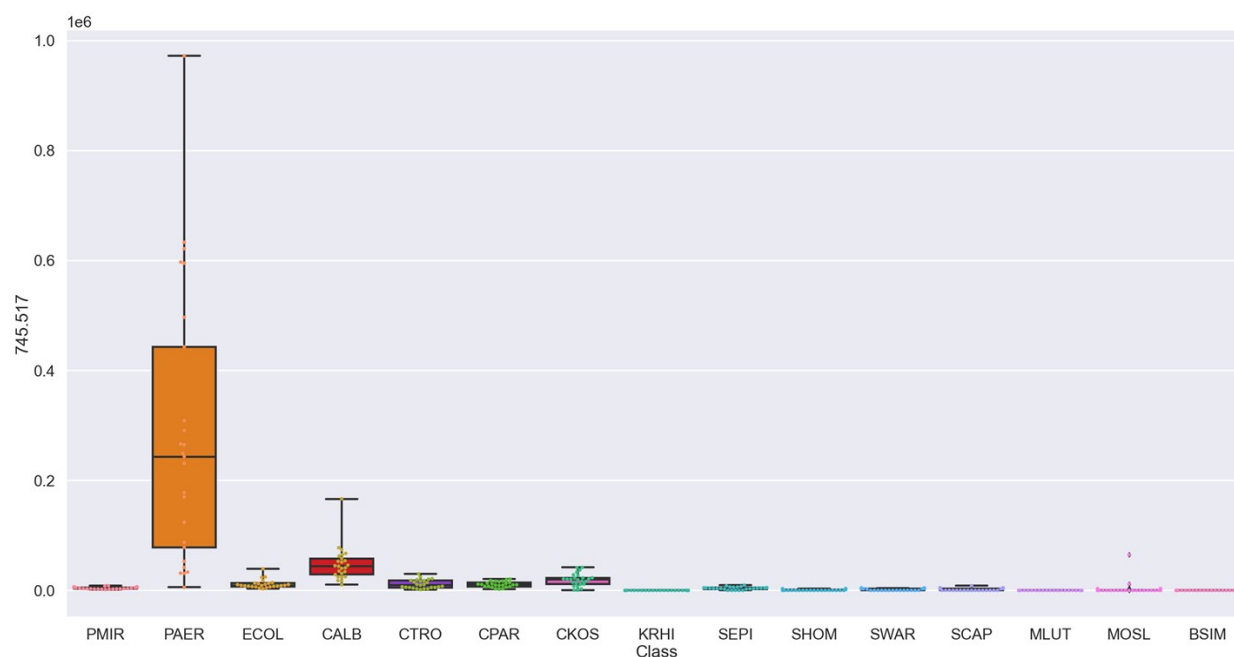


Figure 4.3 Box plot of 745.517 m/z, PG(34:2), (1/188) features used as a new matrix for the analysis. RFECV feature developed using ‘75% peak filter’, ‘Pre-processing 2’, and ‘TIC+Log’ normalisation pre-processing.

Through three stages of model validation, with the full data matrix, RFECV feature matrix and independent validation set, ‘Pre-processing 2’ with ‘75% peak filter’ and ‘TIC+Log’ the training model and validation consistently achieved the highest classification accuracy.

Furthermore, the separation of different species was most obvious when the final data pre-processing combination: ‘Pre-processing 2’ with ‘75% peak filter’ and ‘TIC+Log’, was used. The PCA below (Figure 4.4) shows a clear separation between yeast and bacteria species and between individual species.

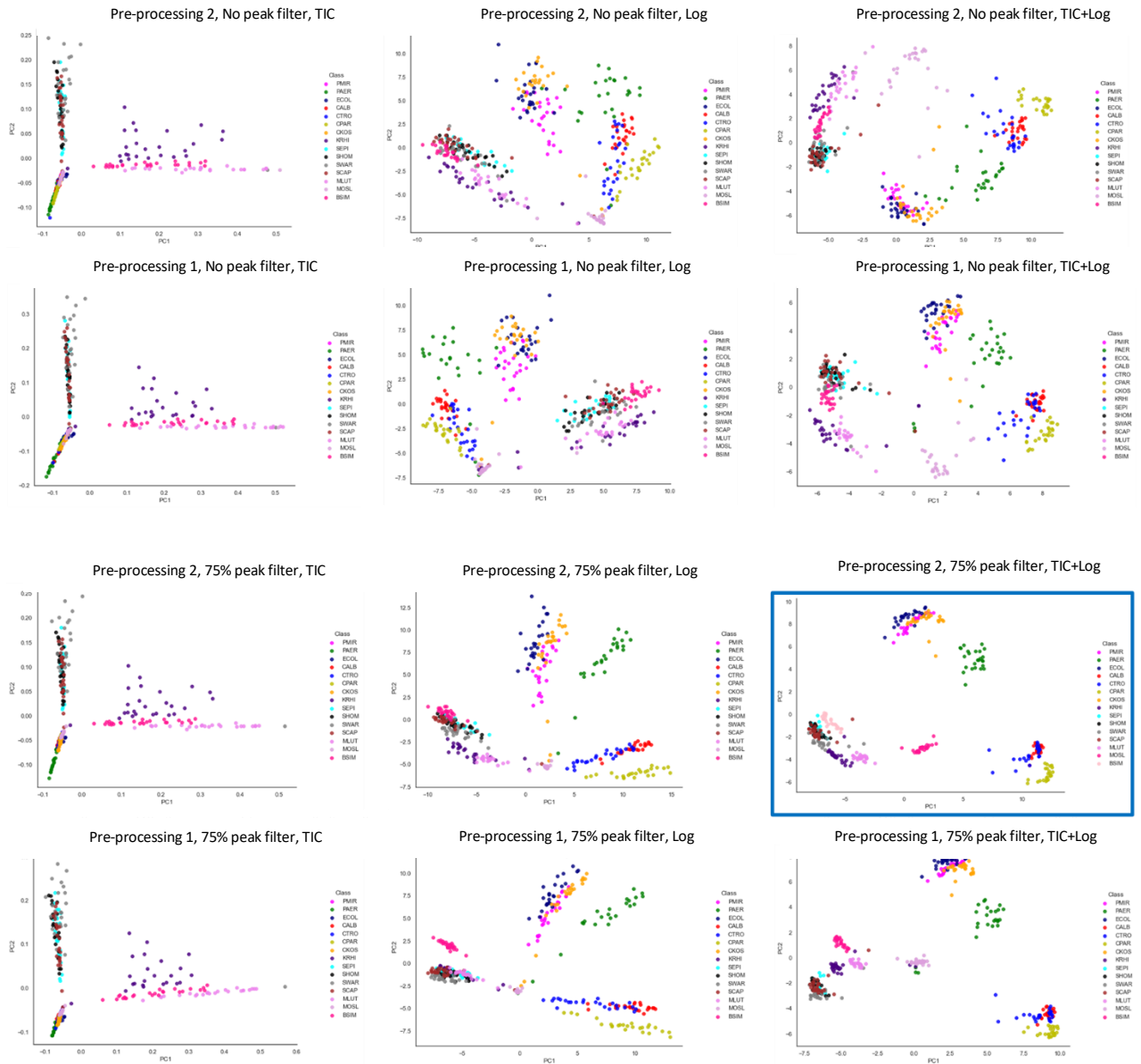


Figure 4.4 PCA scores plots of 12 different pre-processing parameter variations. ‘75% peak filter’ with ‘Pre-processing 2’and ‘TIC+Log’ shows the most separation between species and classes.

Continuing from the above outcomes, a small-scale multivariate classification model built using the selected top parameter combination of ‘75% peak filter’ with ‘Pre-processing 2’and ‘TIC+Log’ normalisation was tested for model stability, repeatability, and reproducibility.

4.4.1.2 Investigation of sample storage effect on LA-REIMS multivariate model stability

Since many existing analytical techniques are time and temperature sensitive that require samples to be processed and analysed within hours, the effect of using refrigerated plates with LA-REIMS speciation was investigated.

A training model was built using 15 organisms where 35 biological replicates for each organism were tested, with 25 being randomly selected for predictive model building and 10 serving as an independent validation dataset. An in-house peak-picking platform with an '75% peak filter' setup was used to generate a data matrix with class labels for post-processing, with 748 peaks identified across the spectrum. The training model was developed using '75% peak filter', and 'Pre-processing 2' with 'TIC+Log' parameters.

Here, to assess the effect of sample refrigeration on LA-REIMS data and the stability of the classification model, first, the training model built was validated using an independent validation set treated under the same conditions as the training model; samples were analysed shortly after the initial incubation for microbial growth (after 24 hours).

Secondly, the same training model was also then validated three more times using the original validation set samples that were replicated three times and stored in the fridge at a temperature of 4 °C for a period of 24, 48, and 72 hours, respectively, before being analysed according to their time in the refrigerator (Figure 4.5).

Figure 4.5 shows the summary of the incubation and refrigeration times prior to the analysis for samples used to create the training model, as well as all independent validation sets.

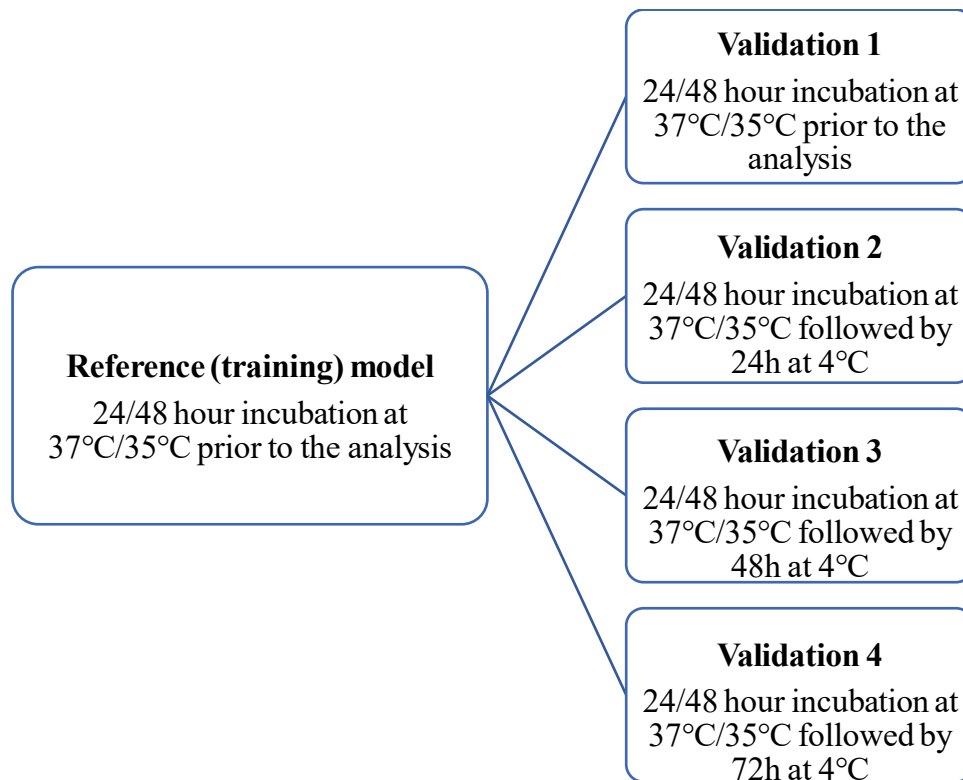


Figure 4.5 Experimental design for sample storage in the refrigeration investigation. 525 isolates from 15 species were used, which were then used for classification analysis. The reference model and Validation 1 set were treated under the same conditions prior to the analysis - 24/48-hour incubation at 37°C/35°C. Validation 2,3, and 4 were kept in the 4°C fridge for 24, 48, and 72 hours, respectively.

The following two figures demonstrate and compare average and individual classification accuracies for each species utilised in this experiment, resulting from the training model that was cross-verified using K-fold cross-validation, as well as the accuracy of the independent validation set.

In this instance, the training model produced an accuracy of F1 score of 99.73% by K-fold cross-validation (Figure 4.6); the only bacteria with lower than perfect 100% identification accuracy F1 score were *Micrococcus luteus* and *Staphylococcus warneri*, at 98% and 97.9%, respectively. The model was then validated using an independent validation set, resulting in an accuracy of F1 score of 100% for all 10 species (Figure 4.7).

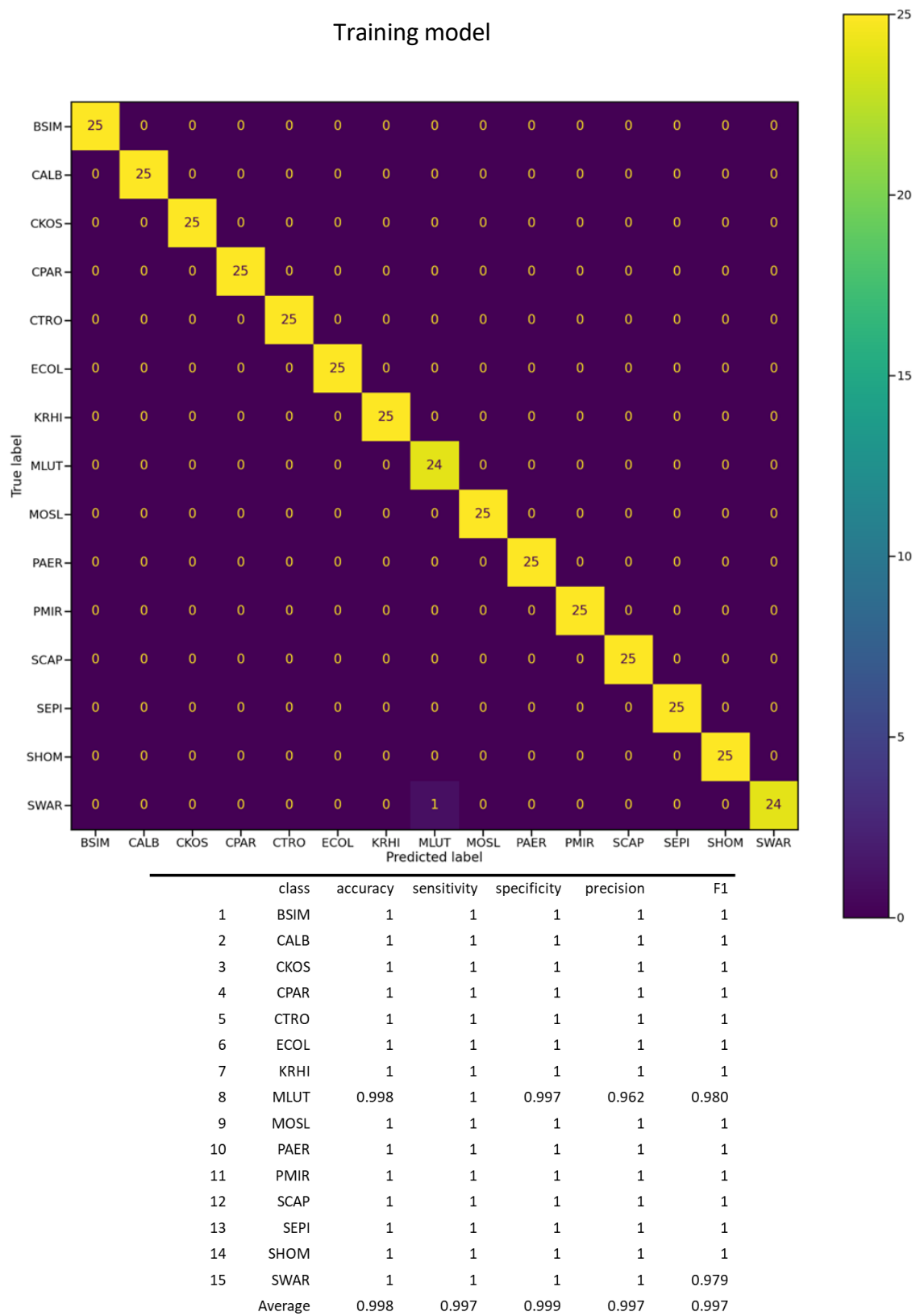


Figure 4.6 Confusion matrix and performance based on SVM classification model with K-fold validation built using 15 species small-scale training model with an overall species-level accuracy of F1 99.73 %.

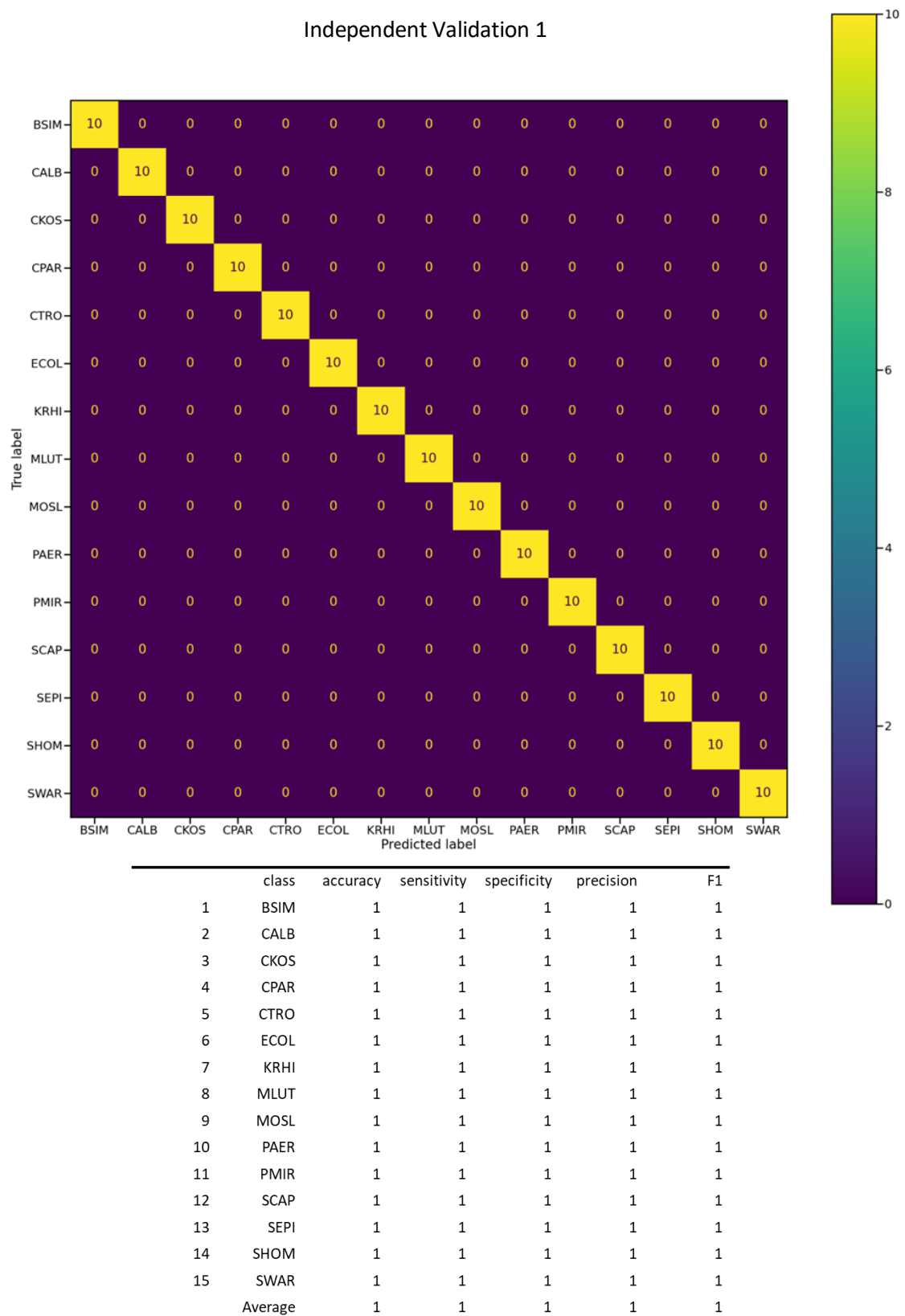


Figure 4.7 Confusion matrix and performance based on the independent validation set used to validate the reference model above (Figure 4.6) with an overall species-level accuracy of F1 100%.

Following this, three further validation sets, referred to as Validation 2, 3, and 4, were classified with the training model.

Validation 2 which was analysed after 24 hours in the 4°C fridge, yielded average classification accuracy F1 score of 96.10%. Although data collected from samples refrigerated for 24 hours prior to analysis resulted in reduced accuracy, the performance was affected mainly by *Staphylococcus warneri* (SWAR) species that were incorrectly classified at the species level while still correctly classified at the genus level (Figure 4.8).

Validation 3 was analysed after 48 hours in the fridge, resulting in an average classification accuracy F1 score of 97.45% (Figure 4.9). Similar to the results obtained from 24-hour refrigeration, a slight reduction in accuracy was primarily due to only one species (SWAR) being wrongly identified at the species level, albeit predicting the correct genus group.

Finally, for Validation 4, which was analysed after 72 hours in the fridge, the average accuracy F1 score obtained was 100% (Figure 4.10), demonstrating that possible accuracy increases are contingent on extended periods of refrigeration leading to biomass changes.

In conclusion, the overall average accuracy F1 score obtained at the species level for all three validation sets isolates was greater than 96%, indicating that sample analysis using the LA-REIMS platform is generally robust to sample variations, such as the change in refrigeration time. In addition, the only erroneous prediction was easily interpretable and still placed in the right genus group, providing sufficient confidence in the robustness of the machine learning model.

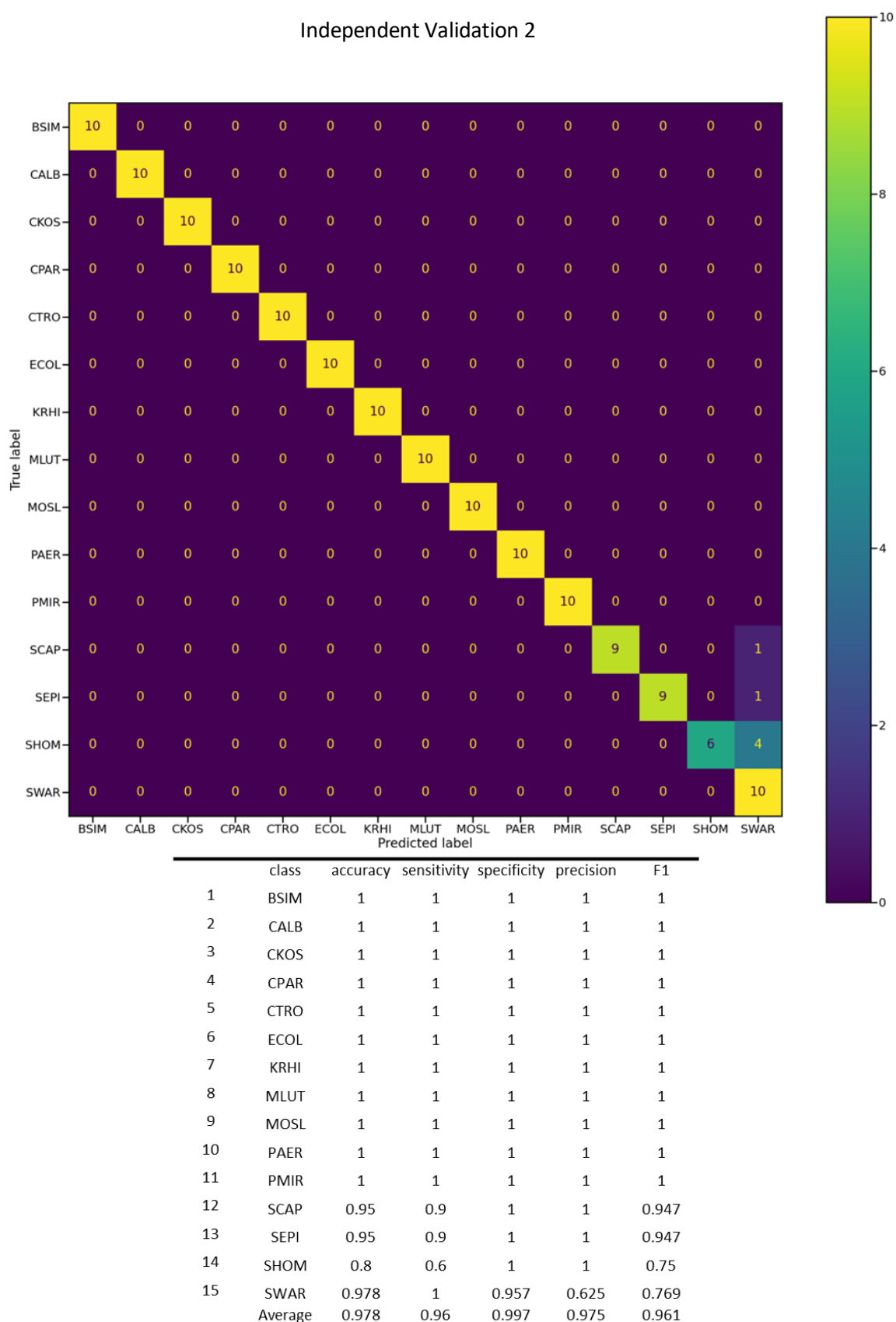


Figure 4.8 Confusion matrix and performance with the Validation 2, which was analysed after 24 hours in the 4°C fridge and yielded a species-level accuracy of F1 96.10%.

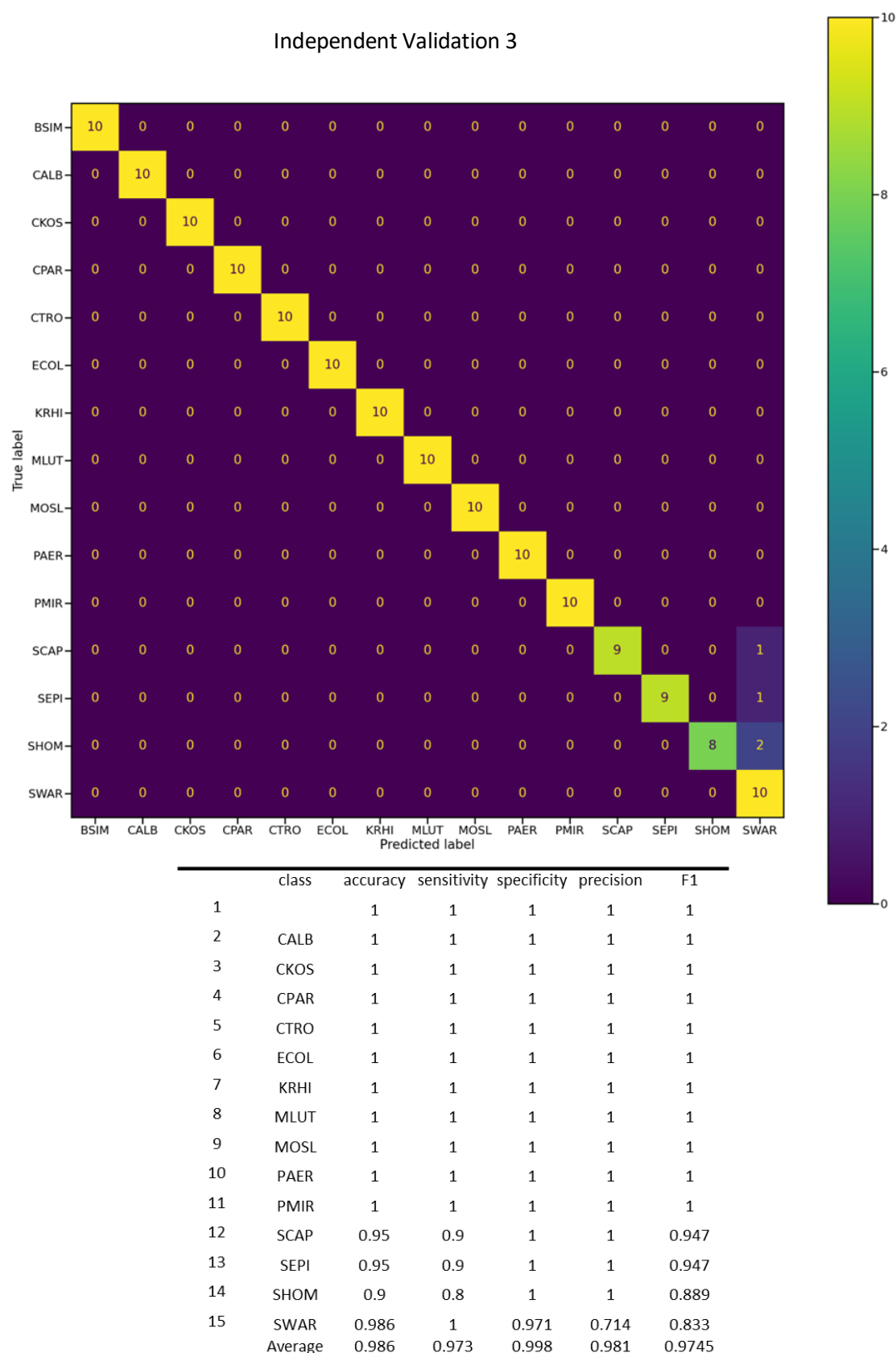


Figure 4.9 Confusion matrix and performance with the Validation 3, which was analysed after 48 hours in the 4°C fridge and yielded a species-level accuracy of F1 97.45%.

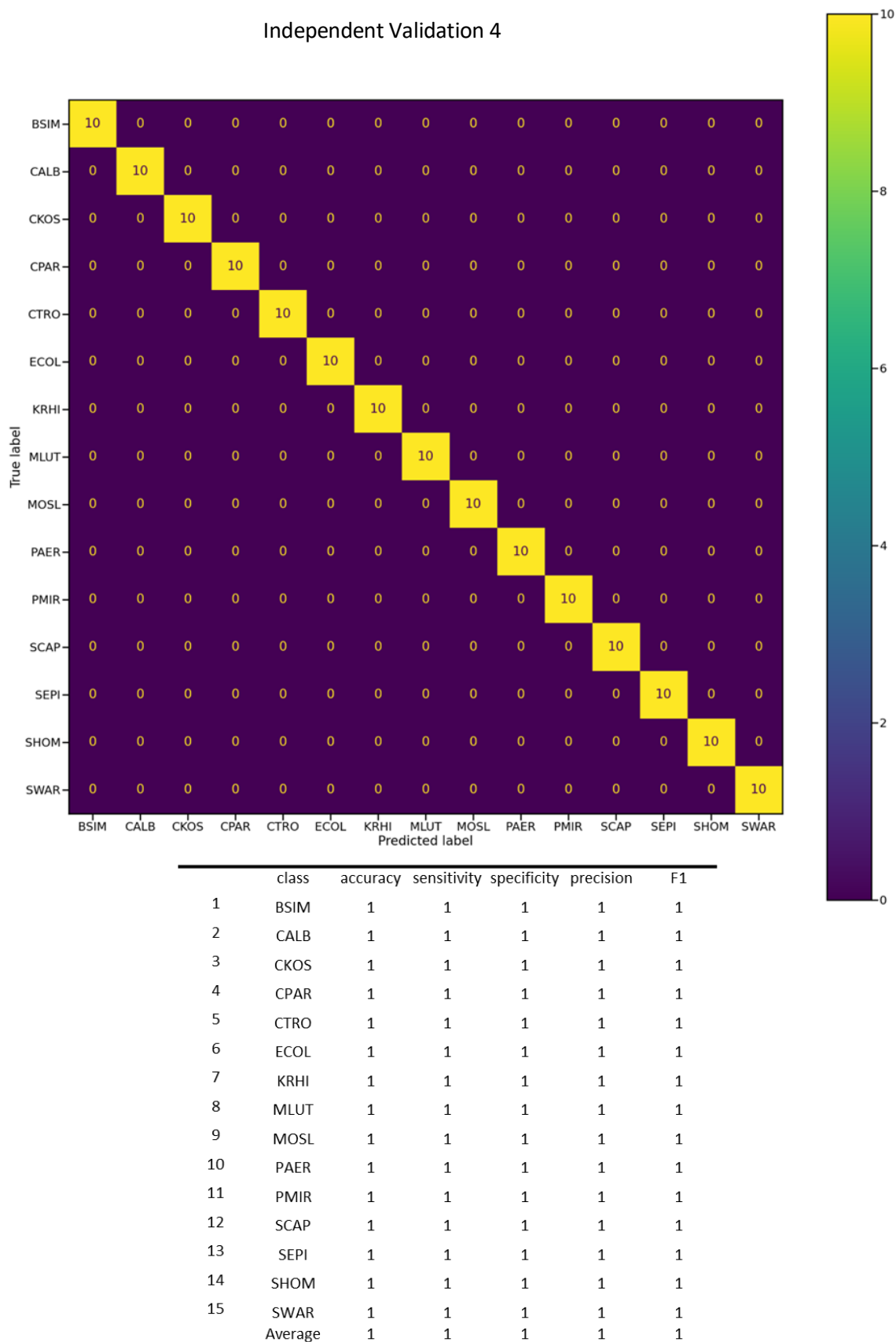


Figure 4.10 Confusion matrix and performance with the Validation 4, which was analysed after 72 hours in the 4°C fridge and yielded a species-level accuracy of F1 97.45%.

4.4.1.3 ROC curve analysis

Receiver-Operating Characteristic (ROC) curve analysis was performed on validation sets 1, 2, 3, and 4 to evaluate and confirm the performance and effectiveness of a classification model reported in the previous paragraph.

In this section, a probability table (Figure 4.11) for classification was generated from ROC analysis for each validation set used to validate the training model, and a ROC plot of the distribution of those probabilities was produced. The probabilities were acquired through the classification of the validation sets using an initial training model that had been built beforehand and had previously achieved an average classification accuracy of 99.73% on the F1 score. Figure 4.11 shows an example of accuracy probability scorings of 8 samples from Validation 1. The probabilities were used to evaluate each class's prediction confidence; the overall probability score was above 0.9, showing strong accuracy.

Figure 4.12 shows four ROC plots of four independent validation sets (validation 1, 2, 3 and 4) used to validate the training model. AUC values of 0.99 were obtained for the original Validation 1, in which samples were handled in the same conditions as the training model, and 1.00 were obtained for each refrigerated validation set 2, 3, and 4. This confirmed that the classification accuracy obtained was reliable and could distinguish between classes near perfect.

Sample	BSIM	CALB	CKOS	CPAR	CTRO	ECOL	KRHI	MLUT	MOSL	PAER	PMIR	SCAP	SEPI	SHOM	SWAR	Actual
AZ008_AZ00037_20190228	0	0	0	0	0	0	0	0	0.0075	0	0	0.005	0.9025	0.0225	0.0625	SEPI
AZ008_AZ00060_20190228	0.0025	0	0	0	0	0	0.0225	0.975	0	0	0	0	0	0	0	MLUT
AZ008_AZ00071_20190228	0.005	0	0	0	0	0	0.9775	0.0075	0	0	0	0	0.0025	0	0.0075	KRHI
AZ008_AZ00089_20190228	0	0	0	0	0.005	0	0	0	0.99	0.0025	0	0	0	0.0025	0	MOSL
AZ008_AZ00094_20190228	0.9925	0	0	0	0	0	0.005	0	0	0	0	0	0	0	0.0025	BSIM
AZ008_AZ00135_20190228	0	0	0	0	0	0	0.0025	0.9975	0	0	0	0	0	0	0	MLUT
AZ008_AZ00161_20190228	0	0	0	0	0	0	0	0	0	0	0	0.0075	0.015	0.97	0.0075	SHOM
AZ008_AZ00165_20190228	0.0025	0	0	0	0	0	0.0325	0.9525	0.005	0.0025	0	0	0	0.0025	0.0025	MLUT

Figure 4.11 Example of probability table for classicisation generated for each of the validation sets. The table shows the actual vs expected probability of scoring 1. Green cells represent the highest probability of classification for an individual sample.

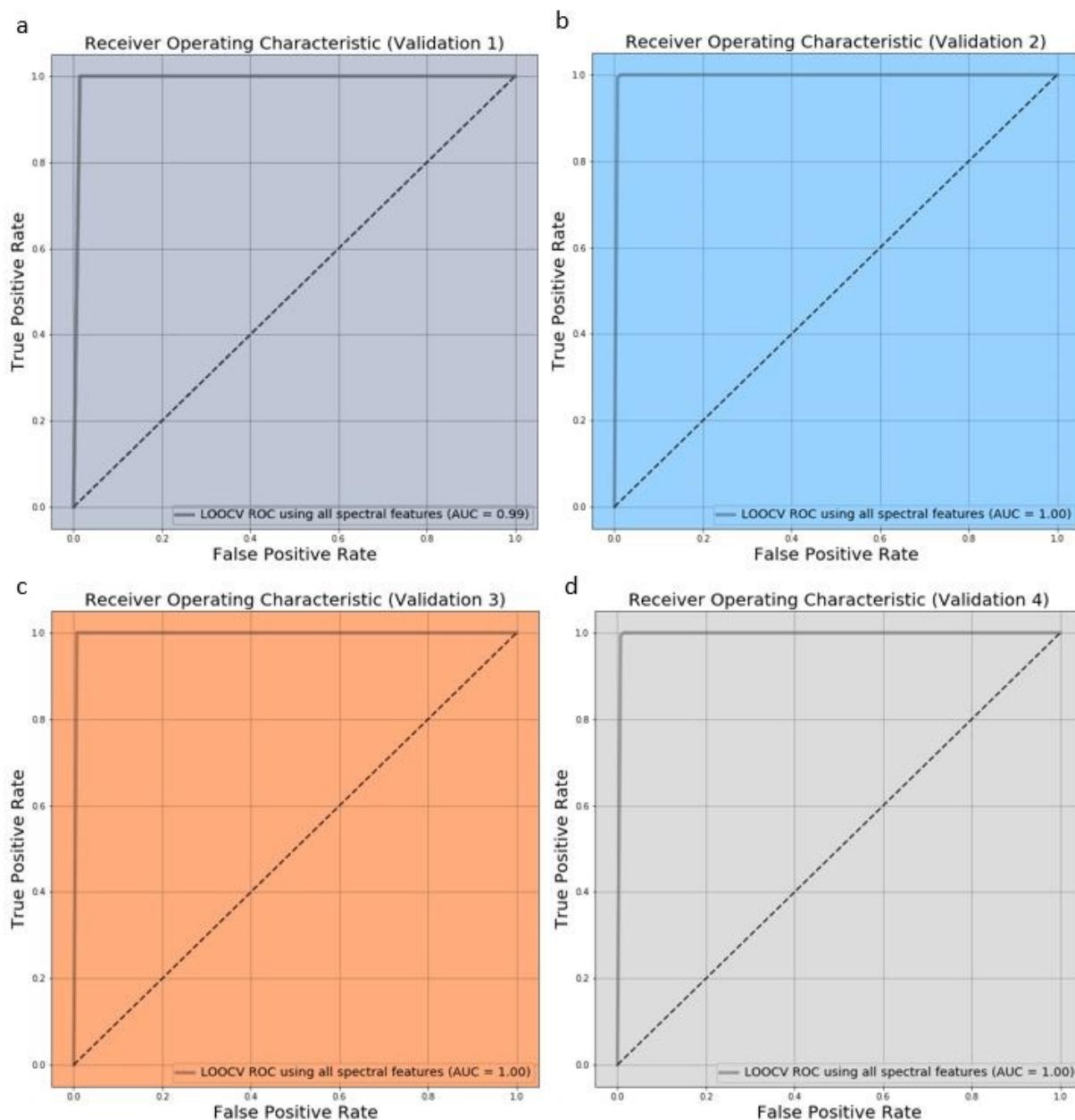


Figure 4.12 (a), (b), (c) and (d) shows ROC curves from four validation sets tested against the training/reference model. (a) displays the ROC curve for the original validation set (accuracy 98.66%), which was constructed under the same conditions as the reference model. (b) ROC curve of refrigerated validation set (accuracy 98.00%) analysed on LA-REIMS after 24 hours in the 4°C. (c) shows the ROC curve for a refrigerated validation set (accuracy 100%) analysed on LA-REIMS after 48 hours in the 4°C. (d) shows ROC curve which represents refrigerated validation set (accuracy 98.00%) analysed on LA-REIMS after 74 hours in the 4°C.

4.4.1.4 Repeatability evaluation

The variability of data obtained from LA-REIMS sampling was evaluated by examining pairwise correlations among spectra from the same species. This assessment served to establish the average correlation as a measure of similarity and, consequently, repeatability across four validation sets conducted over a four-day period (see Figure 4.5). In this context, pairwise correlation refers to the calculation of correlation coefficients between spectra acquired from the same species. This approach helps gauge the similarity between spectra, serving as a measure of repeatability across multiple data acquisition sessions. Pairwise correlation essentially quantifies how closely related two sets of data are to each other. A correlation coefficient of +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. In the context of LA-REIMS, a higher average correlation between spectra from the same species suggests greater consistency and reliability in the acquired data.

The mean pairwise correlation among 15 species is shown in Figure 4.13. Notably, the LA-REIMS platform exhibited robust stability across multiple days of data acquisition and demonstrated consistent data comparability within the same day. The analysis reveals that *Moraxella osloensis* (MOSL) exhibits relatively large variability in its spectra due to its lower biomass growth rate, resulting in lower correlation and repeatability compared to other species. However, despite this variability, the spectral fingerprint of MOSL showcases a distinct phospholipid region containing unique and more stable biomarkers, contributing to its overall high-accuracy classification.

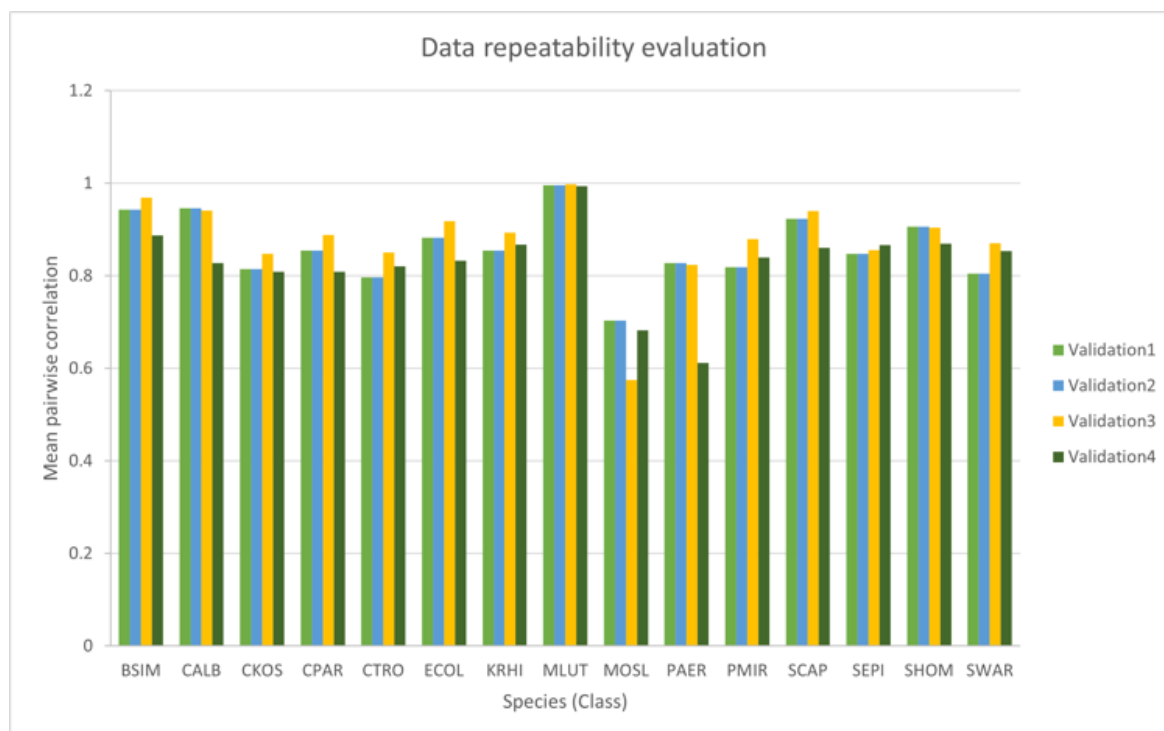


Figure 4.13 Data repeatability evaluation between data from same species. Graph indicates that the mean pairwise correlation factor is stable across days for the same species. The LA-REIMS platform demonstrated great stability and data comparability throughout multiple days of data collection.

4.4.1.5 Reproducibility and stability testing

The reproducibility of the mass spectrometer and the stability of microbiological samples were studied after a three-year period. In 2019, a small-scale model was built and validated with an independent validation set using a Waters Xevo G2-XS QToF instrument (Charring Cross campus, ICL). Three years later, a second mass spectrometer of the same make and manufacturer was used to generate another independent validation set (Burlington Danes campus, ICL), this data was then used to validate the model generated using data from 2019.

Due to the highly different peak intensities, the pre-processing parameters were re-optimised. The hypothesis was that due to differences in mass spectrometer backgrounds, classification parameter peak matching would yield too much variance in peak matching from two separate instruments. If it is proven that ‘75% peak filter’ parameters result in higher performance, it

will provide further validation on the final multivariate model building parameters as well as sample and instrumentation stability.

In order to perform the analysis, new data comprising an independent validation set and old data containing training model data were combined into a new metadata file and pre-processed together to obtain a new matrix file to be used for classification model building and validation. The validation process was carried out using an independent validation set (new data) and K-fold cross-validation on the training model (old data).

In this section, the ‘No peak filter’ parameters and the ‘75% peak filter’ parameters both resulted in high-performance training models with a high level of accuracy: the training model for the ‘No peak filter’ parameters had an average accuracy of F1 score of 97.33%, while the training model for the ‘75% peak filter’ parameters had an F1 score of 99.73% on average for its accuracy (Figure 4.14).

However, based on the average accuracy summary in Table 4.6 below, when validated with an independent data set run three years after the training data set and on different Xevo G2-XS QTOF mass spectrometers, it was clear that the ‘75% peak filter’ parameter during pre-processing was the best option due to the ability only to include peaks matching in 75% of the classes.

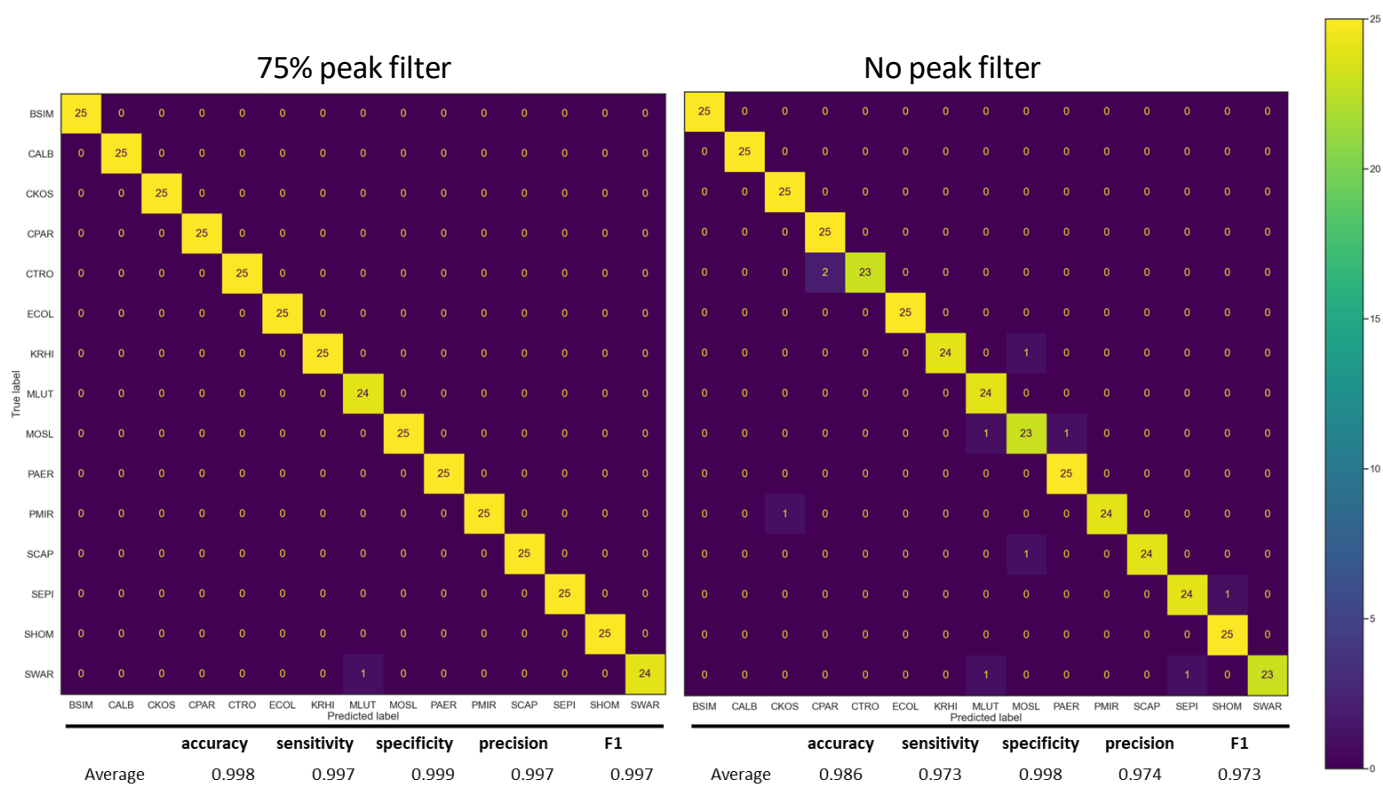


Figure 4.14 Confusion matrixes and average classification accuracy representing two training models build using ‘75% peak filter’ and ‘No peak filter’ pre-processing parameters.

Table 4.6 The training model and independent validation classification accuracy using ‘No peak filter’ and ‘75% peak filter’ pre-processing along with ‘Pre-processing 2’ and “TIC+Log”.

	‘No peak filter’	‘75% peak filter’
	‘Pre-processing 2’	‘Pre-processing 2’
Model normalisation	TIC+LOG	TIC+LOG
Training model (old data) Kfold validation_F1 score	97.33%	99.73%
Independent validation (new data) F1 score	29.71%	95.43%

Here, the average classification accuracy of the F1 score of an independent validation set applied to validate the ‘No peak filter’ parameters-based model was only F1 29.71% (Figure 4.15), but the ‘75% peak filter’ based model had an average accuracy of 95.43%. (Figure 14). Only old vs new common peaks were used in the analysis by applying the ‘75% peak filter’ parameter rather than all peaks, including background peaks that vary. A stable multivariate model with high accuracy was produced by eliminating these variable peaks.

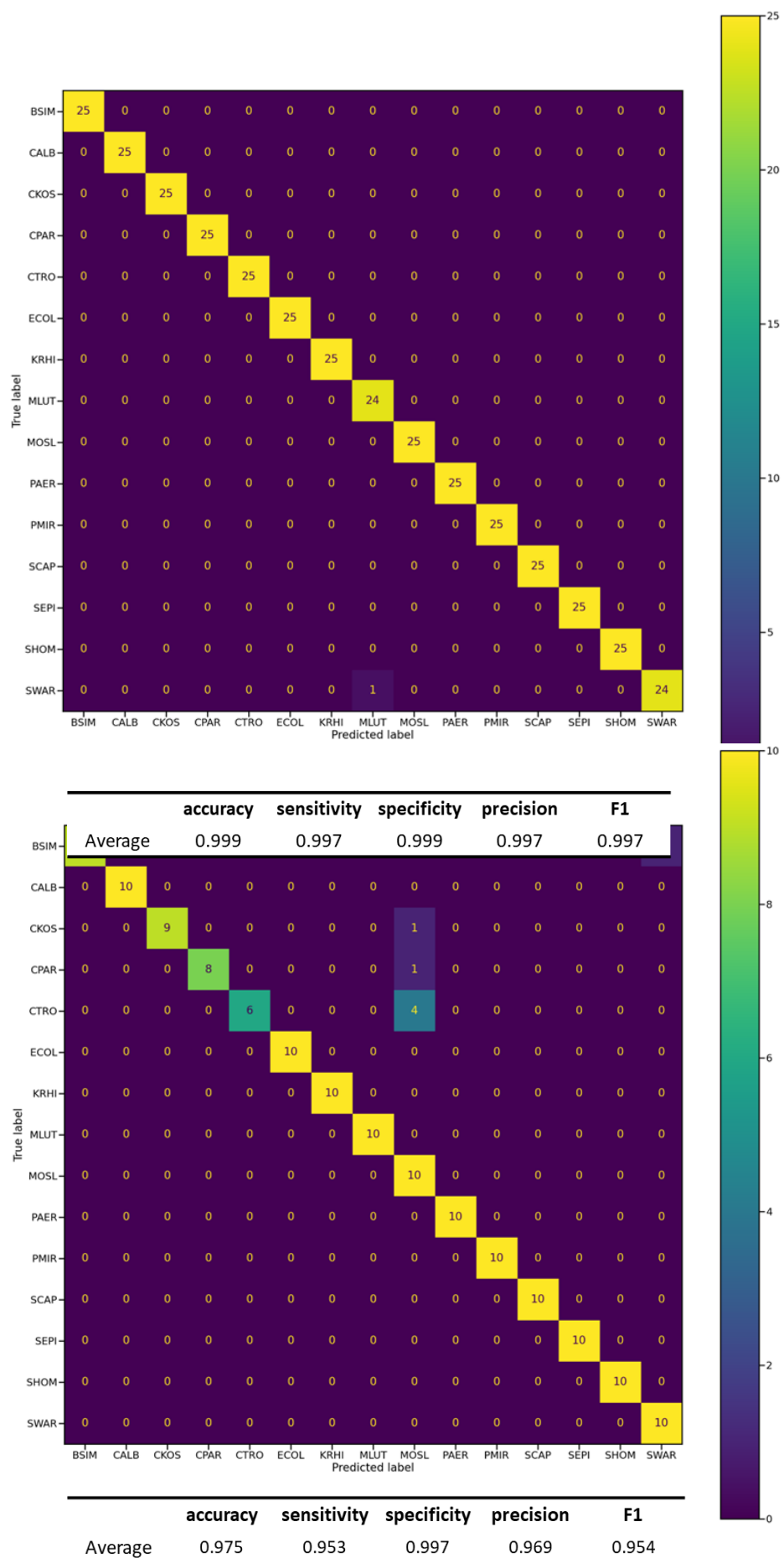


Figure 4.15 Confusion matrix and performance of training model and independent validation, based on '75% peak filter', with an overall accuracy for independent validation of F1 95.43%.

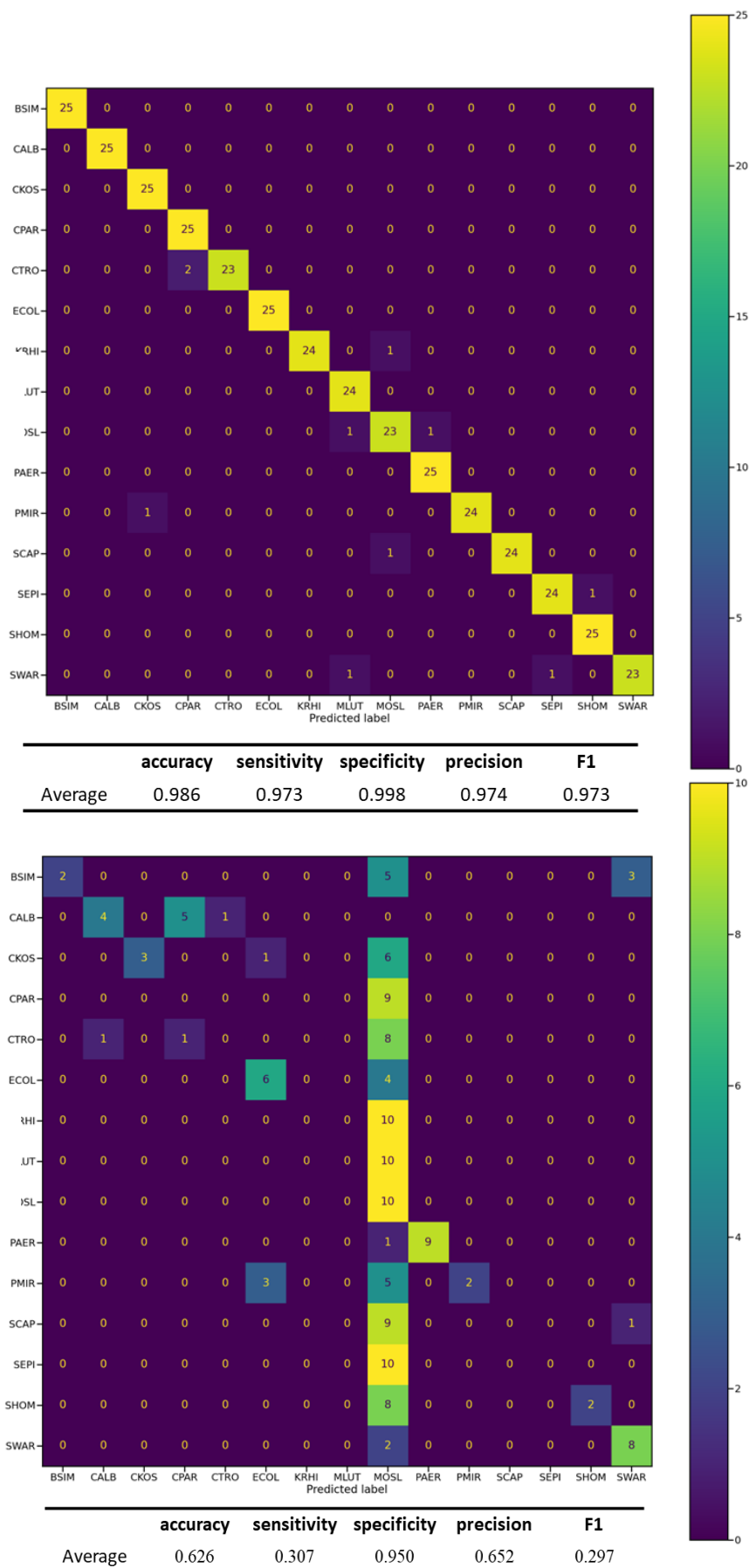


Figure 4.16 Confusion matrix and performance of training model and independent validation, based on 'No peak filter', with an overall accuracy for independent validation of F1 25.71%.

4.4.2 Large-scale multivariate model

Final data pre-processing combination that was concluded in small scale model workflow above: ‘75% peak filter’ with ‘Pre-processing 2’ and “TIC+Log”, was used for large scale model data pre-processing. SVM modelling approach with K-fold validation was used. A total of 1327 isolates consisting of a mixture of bacterial and yeast isolates representing 211 different bacterial and yeast species (Table 4.7) were selected for analysis. The samples included in the study were a mix of widely seen pharmaceutical environmental bacteria and yeast contaminants.

For a final list of the isolates with assigned classes that were used to create a large classification model for environmental monitoring, please refer to Table 4.10 in the *Appendix*. For each of the 162 organisms, a mixture of 4264 biological and analytical replicates were combined, with 3599 being randomly selected for predictive model building and 665 (15.5%) serving as an independent validation dataset. Biological replicates refer to separate samples or cultures of the same bacterial strain grown under identical conditions. These replicates are essential for assessing the reproducibility and reliability of experimental results. Analytical instrumental replicates of individual colony analysis involve using the same analytical instrument to measure the properties or characteristics of multiple individual bacterial colonies in same agar plate (five per plate). These replicates help account for any variability introduced by the measurement process itself, such as instrumental noise or fluctuations.

2265 peaks were identified via pre-processing. Building on the results from method development and validation from training a small-scale model, a large-scale model was built accordingly. The trained model then evaluated using K-fold cross validation, an independent validation set both before and after it has been optimised by RFE.

In addition, LC-MS was performed on all of the species that were used in the large classification model building study. The subsequent LC-MS results were then used to identify the features that were derived using RFE, which were finally validated by using LC-MS/MS.

Table 4.7 Bacterial and yeast strains representing 211 different species used in large-scale classification model building for industrial pharmaceutical environmental monitoring

<i>Species</i>	<i>Species</i>	<i>Species</i>	<i>Species</i>
<i>Acinetobacter baumannii</i>	<i>Brevibacterium frigoritolerans</i>	<i>Mesobacillus subterraneus</i>	<i>Pseudomonas hydrolytica</i>
<i>Acinetobacter lwoffii</i>	<i>Brevibacterium iodinum</i>	<i>Mesobacillus thioparans</i>	<i>Pseudomonas lactis</i>
<i>Acinetobacter radioresistens</i>	<i>Brevibacterium pityocampae</i>	<i>Metabacillus halosaccharovorans</i>	<i>Pseudomonas nosocomialis</i>
<i>Acinetobacter schindleri</i>	<i>Candida albicans</i>	<i>Microbacterium aurum</i>	<i>Pseudomonas rhizosphaerae</i>
<i>Acinetobacter townneri</i>	<i>Candida guilliermondii</i>	<i>Microbacterium enclense</i>	<i>Pseudomonas stutzeri</i>
<i>Acinetobacter ursingii</i>	<i>Candida parapsilosis</i>	<i>Microbacterium lacticum</i>	<i>Ralstonia pickettii</i>
<i>Aerococcus urinaequi</i>	<i>Citricoccus zhacaiensis</i>	<i>Microbacterium oxydans</i>	<i>Raoultella planticola</i>
<i>Aerococcus viridans</i>	<i>Citrobacter freundii</i>	<i>Microbacterium paraoxydans</i>	<i>Rhizobium radiobacter</i>
<i>Agrococcus jenensis</i>	<i>Corynebacterium accolens</i>	<i>Microbacterium proteolyticum</i>	<i>Rhodotorula mucilaginosa</i>
<i>Arthrobacter luteolus</i>	<i>Corynebacterium afermentans</i>	<i>Microbacterium testaceum</i>	<i>Roseomonas mucosa</i>
<i>Bacillus altitudinis</i>	<i>Corynebacterium aurimucosum</i>	<i>Micrococcus flavus</i>	<i>Rothia aerea</i>
<i>Bacillus amyloliquefaciens</i>	<i>Corynebacterium coyleae</i>	<i>Micrococcus luteus</i>	<i>Rothia dentocariosa</i>
<i>Bacillus aryabhattai</i>	<i>Corynebacterium propinquum</i>	<i>Micrococcus lylae</i>	<i>Rothia kristinae</i>
<i>Bacillus atrophaeus</i>	<i>Corynebacterium sanguinis</i>	<i>Micrococcus terreus</i>	<i>Rothia terrae</i>
<i>Bacillus cereus</i>	<i>Corynebacterium tuberculostearicum</i>	<i>Moraxella osloensis</i>	<i>Salmonella enterica</i>
<i>Bacillus circulans</i>	<i>Corynebacterium xerosis</i>	<i>Niallia circulans</i>	<i>Shigella flexneri</i>
<i>Bacillus dakarensis</i>	<i>Dermacoccus nishinomiyaensis</i>	<i>Niallia Species</i>	<i>Sphingobacterium multivorum</i>
<i>Bacillus flexus</i>	<i>Dietzia papillomatosis</i>	<i>OceanoBacillus chironomi</i>	<i>Sphingomonas carotinifaciens</i>
<i>Bacillus gibsonii</i>	<i>Enterobacter aerogenes</i>	<i>Oceanobacillus kimchii</i>	<i>Sphingomonas koreensis</i>
<i>Bacillus halosaccharovorans</i>	<i>Enterobacter kobei</i>	<i>OrnithiniBacillus bavariensis</i>	<i>Sphingomonas melonis</i>
<i>Bacillus indicus</i>	<i>Enterococcus faecalis</i>	<i>Paenibacillus amylolyticus</i>	<i>Sphingomonas olei</i>
<i>Bacillus infantis</i>	<i>Escherichia coli</i>	<i>Paenibacillus campinasensis</i>	<i>Sphingomonas panaciterrae</i>
<i>Bacillus kochii</i>	<i>Escherichia fergusonii</i>	<i>Paenibacillus cucumis</i>	<i>Sphingopyxis terrae</i>
<i>Bacillus kyonggiensis</i>	<i>Exiguobacterium artemiae</i>	<i>PaeniBacillus glucanolyticus</i>	<i>Staphylococcus argenteus</i>
<i>Bacillus licheniformis</i>	<i>Exiguobacterium profundum</i>	<i>Paenibacillus humicus</i>	<i>Staphylococcus arlettae</i>
<i>Bacillus licheniformis</i>	<i>Exophiala dermatitidis</i>	<i>Paenibacillus ihumii</i>	<i>Staphylococcus aureus</i>
<i>Bacillus massiliosenegalensis</i>	<i>Exophiala species</i>	<i>Paenibacillus illinoisensis</i>	<i>Staphylococcus auricularis</i>
<i>Bacillus megaterium</i>	<i>Fictibacillus barbaricus</i>	<i>PaeniBacillus lautus</i>	<i>Staphylococcus capitis</i>
<i>Bacillus megaterium</i>	<i>Fictibacillus phosphorivorans</i>	<i>Paenibacillus macerans</i>	<i>Staphylococcus caprae</i>
<i>Bacillus methylotrophicus</i>	<i>Fusobacterium necrophorum</i>	<i>Paenibacillus pabuli</i>	<i>Staphylococcus cohnii</i>
<i>Bacillus mobilis</i>	<i>Georgenia muralis</i>	<i>Paenibacillus peoriae</i>	<i>Staphylococcus condimenti</i>
<i>Bacillus mycoides</i>	<i>Gordonia aichiensis</i>	<i>PaeniBacillus provencensis</i>	<i>Staphylococcus edaphicus</i>
<i>Bacillus nealsonii</i>	<i>Gordonia bronchialis</i>	<i>Paenibacillus segetis</i>	<i>Staphylococcus epidermidis</i>

<i>Bacillus oceanisediminis</i>	<i>Gordonia didemni</i>	<i>Paenibacillus segetis</i>	<i>Staphylococcus haemolyticus</i>
<i>Bacillus onubensis</i>	<i>Gordonia jacobaea</i>	<i>Paenibacillus Species</i>	<i>Staphylococcus hominis</i>
<i>Bacillus paramycoides</i>	<i>Gordonia otitidis</i>	<i>Paenibacillus susongensis</i>	<i>Staphylococcus lentus</i>
<i>Bacillus pumilus</i>	<i>Gordonia polyisoprenivorans</i>	<i>Paenibacillus taichungensis</i>	<i>Staphylococcus lugdunensis</i>
<i>Bacillus safensis</i>	<i>Gordonia sputi</i>	<i>Paenibacillus urinalis</i>	<i>Staphylococcus pasteurii</i>
<i>Bacillus siamensis</i>	<i>Janibacter hoylei</i>	<i>PaeniBacillus xylanilyticus</i>	<i>Staphylococcus petrasii</i>
<i>Bacillus subtilis</i>	<i>Janibacter indicus</i>	<i>Pantoea agglomerans</i>	<i>Staphylococcus pettenkoferi</i>
<i>Bacillus tequilensis</i>	<i>Klebsiella pneumoniae</i>	<i>Pantoea anthophila</i>	<i>Staphylococcus saprophyticus</i>
<i>Bacillus thioparans</i>	<i>Kocuria arsenatis</i>	<i>Pantoea conspicua</i>	<i>Staphylococcus simulans</i>
<i>Bacillus thuringiensis</i>	<i>Kocuria marina</i>	<i>Pantoea septica</i>	<i>Staphylococcus succinus</i>
<i>Bacillus toyonensis</i>	<i>Kocuria palustris</i>	<i>Paracoccus yeei</i>	<i>Staphylococcus warneri</i>
<i>Bacillus vallismortis</i>	<i>Kocuria rhizophila</i>	<i>Peribacillus butanolivorans</i>	<i>Stenotrophomonas panacihumi</i>
<i>Bacillus velezensis</i>	<i>Kocuria salina</i>	<i>Peribacillus psychrosaccharolyticus</i>	<i>Streptococcus australis</i>
<i>Bhargavaea changchunensis</i>	<i>Kocuria salsicia</i>	<i>Planococcus okeanoikoites</i>	<i>Streptococcus dysgalactiae</i>
<i>Brachybacterium muris</i>	<i>Lactococcus lactis</i>	<i>Prolinoborus fasciculus</i>	<i>Streptococcus salivarius</i>
<i>Brachybacterium paraconglomeratum</i>	<i>Leifsonia soli</i>	<i>Proteus mirabilis</i>	<i>Streptococcus sanguinis</i>
<i>Brevibacillus aydinogluensis</i>	<i>Listeria monocytogenes</i>	<i>Pseudomonas aeruginosa</i>	<i>Streptococcus vestibularis</i>
<i>Brevibacillus parabrevis</i>	<i>Lysinibacillus composti</i>	<i>Pseudomonas azotoformans</i>	<i>Streptomyces griseoflavus</i>
<i>Brevibacterium casei</i>	<i>Lysinibacillus fusiformis</i>	<i>Pseudomonas chlororaphis</i>	<i>Xanthomonas species</i>
<i>Brevibacterium celere</i>	<i>Lysinibacillus odysseyi</i>	<i>Pseudomonas coleopterorum</i>	

4.4.2.1 Large-scale model building and validation

Firstly, PCA score plot revealed a strong separation of classes in the dataset (Figure 4.17), demonstrating that the data contains species-specific features that drive class separation in the classification model and allow for high-accuracy prediction of unknown species. When comparing PC1 to PC3 in Figure 4.17, there is a distinct clustering and separation of the *Staphylococcus* genus species from the rest of the species at the top right of the plot.

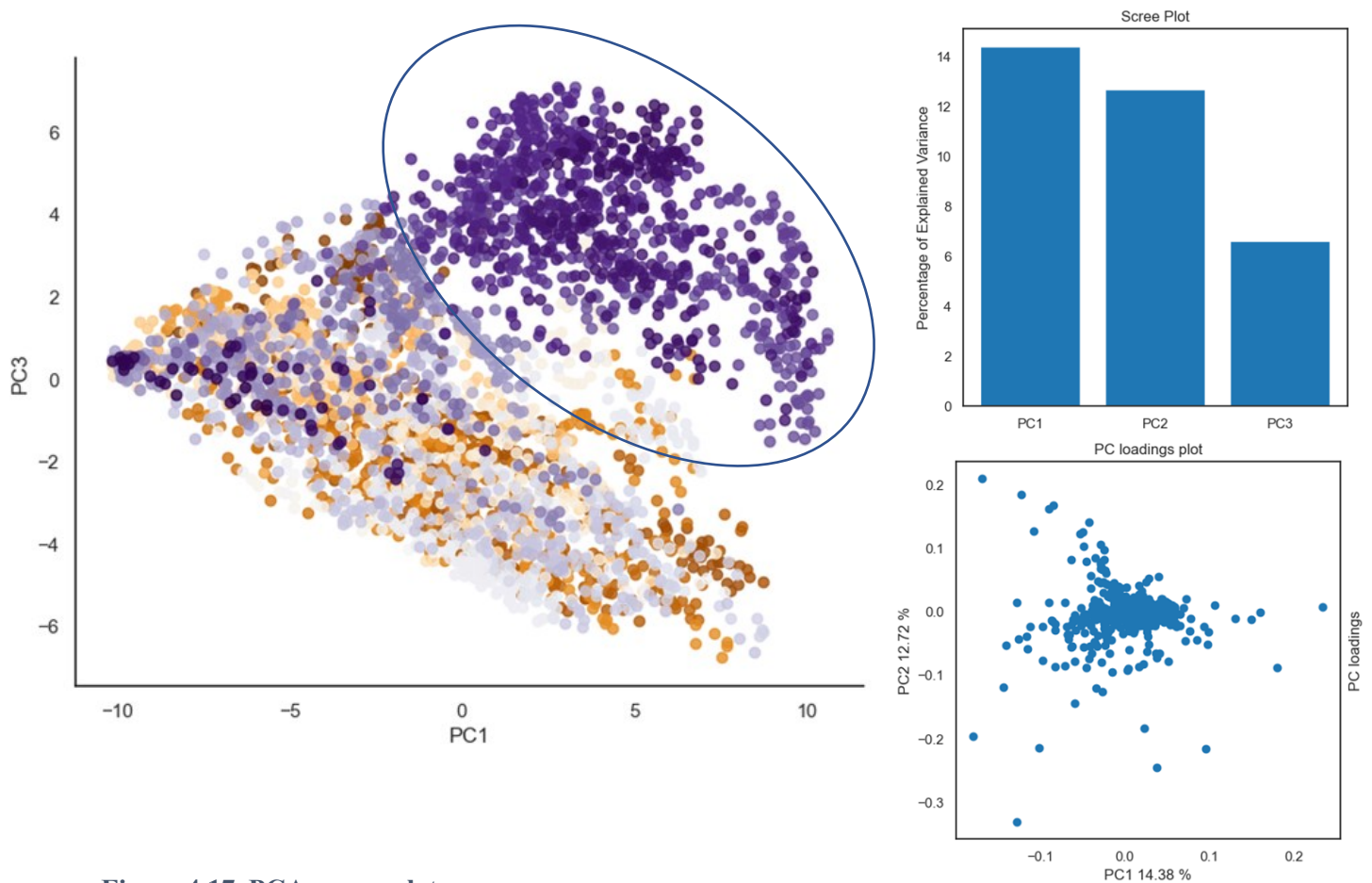


Figure 4.17. PCA scores plot displaying PC1 vs PC3, of the microbial species used for to develop large-scale classification training model.

Class			
● AcineBauma	● CitriZhaca	● MicroProte	● SphinPanac
● AcineWoff	● CitroFreun	● MicroSP	● SphinSP
● AcineRadio	● CorynAccol	● MicroTerre	● SphinTerra
● AcineSchin	● CorynAferm	● MoraxOsloe	● StaphArgen
● AcineTowne	● CorynAurim	● NiallCircu	● StaphAriet
● AcineUrsin	● CorynCoyle	● NiallSP	● StaphAureu
● AerocUrina	● CorynPropi	● OceanChiro	● StaphAuric
● AerocVirid	● CorynSangu	● OceanKimch	● StaphCapit
● AgrocJenen	● CorynTuber	● OrnitBavar	● StaphCapra
● ArthrLuteo	● CorynXeros	● PaeniAmylo	● StaphCohni
● BacilAltit	● DermaNishi	● PaeniCampi	● StaphCondi
● BacilAmylo	● DietzPapil	● PaeniGluca	● StaphEdaph
● BacilAtrop	● EnterAerog	● PaeniLautu	● StaphEpide
● BacilCereu	● EnterFaeca	● PaeniProve	● StaphHaemo
● BacilCircu	● EnterKobei	● PaeniSeget	● StaphHomin
● BacilFlexu	● EscheColi	● PaeniSP	● StaphLentu
● BacilHalos	● EscheFergu	● PaeniTaich	● StaphLugdu
● BacilInfan	● ExiguSP	● PantoSepti	● StaphPaste
● BacilLiche	● ExophDerma	● PantoSP	● StaphPetra
● BacilMassi	● ExophSP	● ParacYeei	● StaphPette
● BacilMegat	● FictiBarba	● PeribButan	● StaphSapul
● BacilMycoi	● FictiPhosp	● PeribPsych	● StaphSimul
● BacilOcean	● FusobNecro	● PlanoOkean	● StaphSucci
● BacilParam	● GeorgMural	● ProlifFasci	● StaphWarne
● BacilPumil	● GordoBronc	● ProteMirab	● StenoPanac
● BacilSafen	● GordoSP	● PseudAerug	● StrepSaliv
● BacilSP	● GordoSputi	● PseudAzoto	● StrepSangu
● BacilSubti	● JanibSP	● PseudChlor	● StreptocSP
● BacilTequi	● KlebsPneum	● PseudSP	● StreptomSP
● BacilThuri	● KocurArsen	● PseudStutz	● XanthSP
● BhargChang	● KocurPalus	● RalstPicke	
● BrachSP	● KocurRhizo	● RaoulPlant	
● BreviAydin	● KocurSP	● RhizoRadio	
● BreviCasei	● LactoLacti	● RhodoMucil	
● BreviCeler	● LeifsSoli	● RoseoMucos	
● BreviFrigo	● ListeMonoc	● RothiAeria	
● Brevilodin	● LysinCompo	● RothiDento	
● BreviParab	● LysinFusif	● RothiKrist	
● BreviPityo	● LysinOdyss	● RothiTerra	
● CandiAlbic	● MesobSP	● ShigeFlexn	
● CandiGuill	● MetabHalos	● ShingOlei	
● CandiParap	● MicroFlavu	● SphinCarot	
	● MicroLacti	● SphinMelon	
	● MicroLuteu	● SphinMulti	
	● MicroLylae		
	● MicroParao		

To establish the validity of the large-scale model and prevent overfitting, a multivariate classification model was built and validated by employing the same three-level validation process that was used in the previous paragraph for the small-scale model.

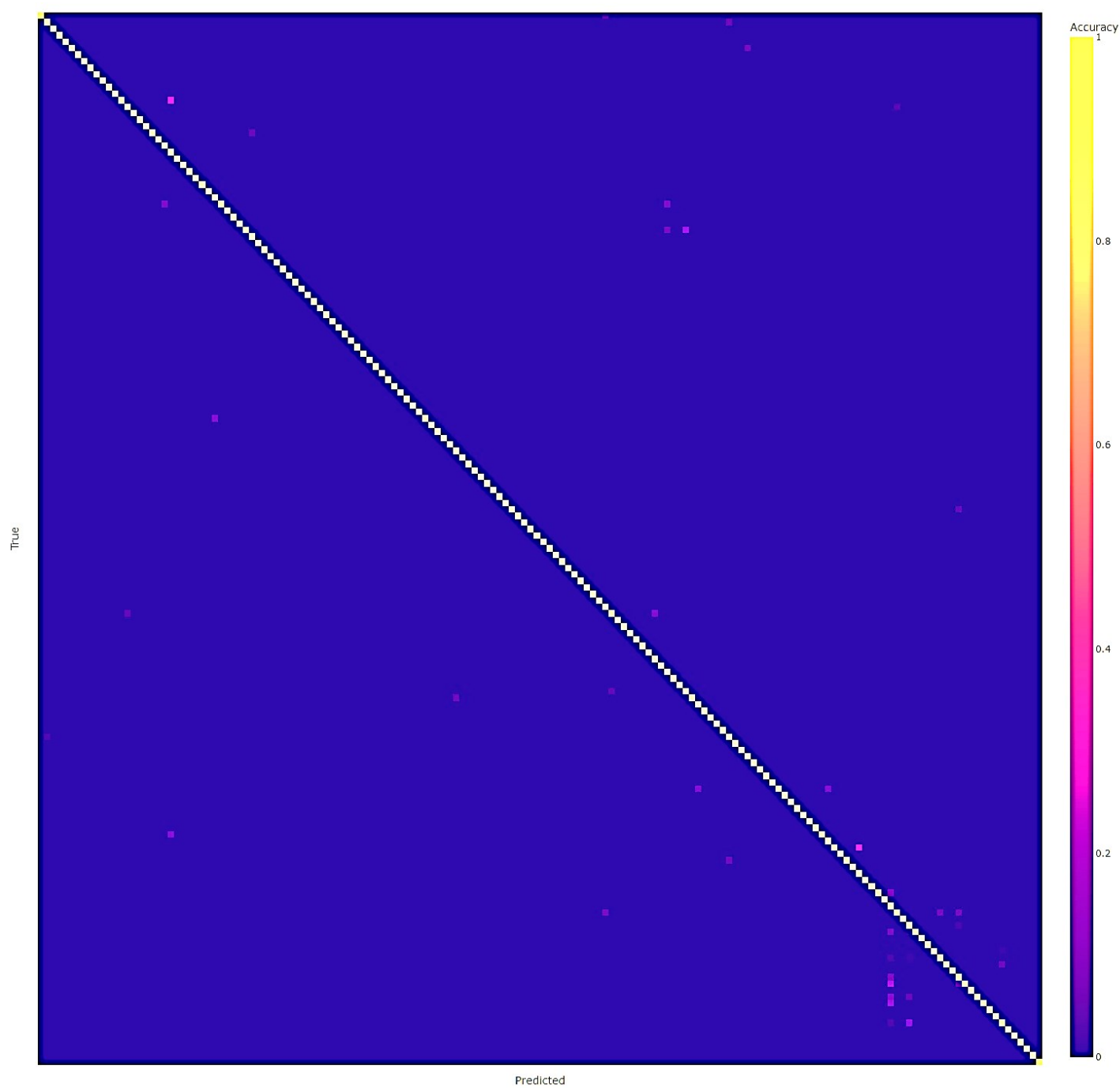
By inspection of Figure 4.18, the trained model achieved an average accuracy of F1 score of 98.1% after K-fold cross-validation.

The training model was then validated using an independent validation set that was designed by leaving 15.5% of the overall acquired data from the training model building and had an average classification accuracy of F1 score of 100% (Figure 4.19).

Please refer to Tables 4.11 and 4.12 in the *Appendix* for full list of classification results for K-fold cross validation and independent set validation.

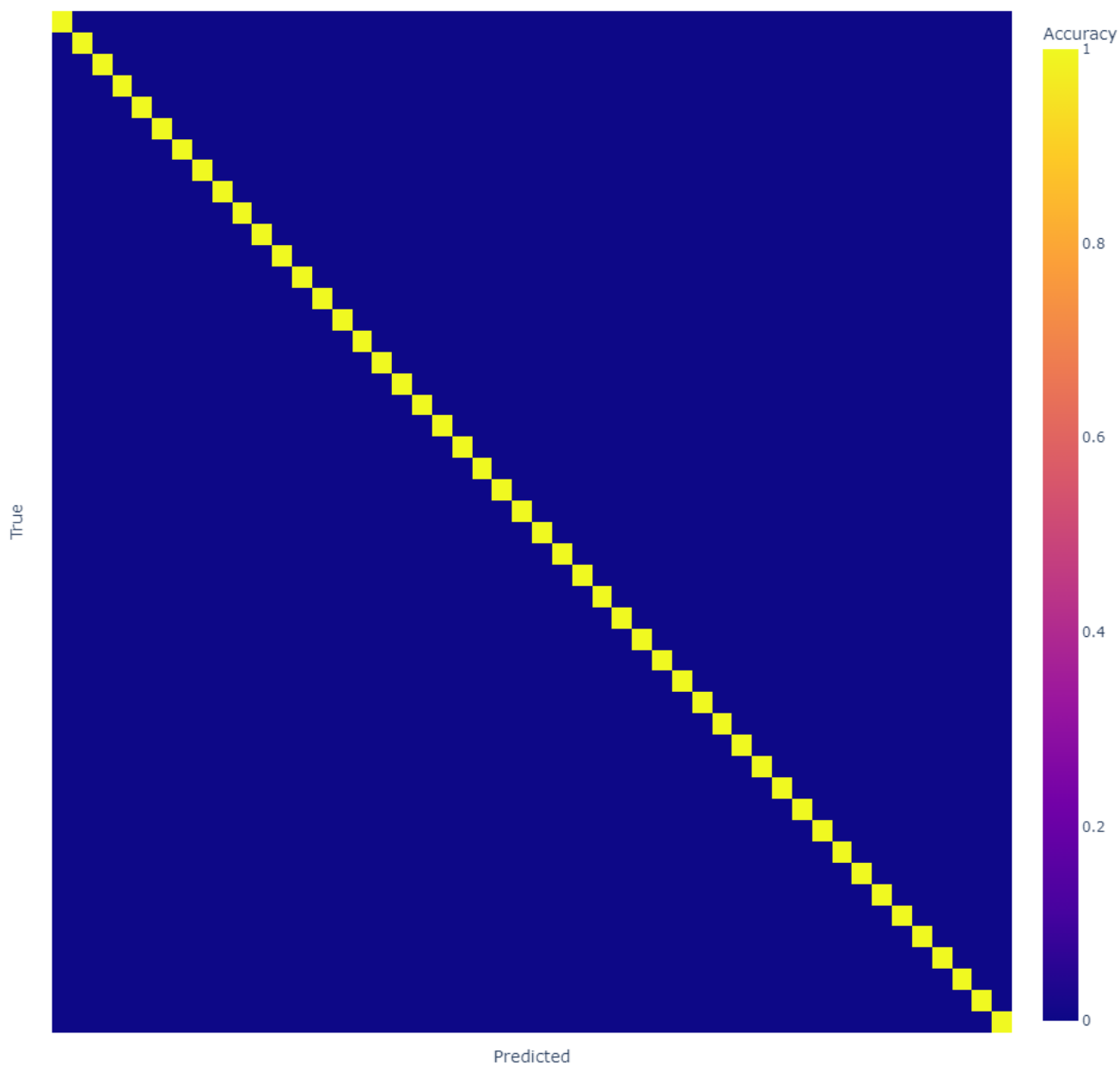
The whole list of individual class prediction performance can be seen in the appendix below, table 9, and it indicates both consistency in the accuracy of predicting unknown species and a low percentage of false negatives.

Furthermore, the ROC-AUC evaluation for the unknown species classification (independent validation set) in Figure 4.20, confirms the model's ability to generate correct predictions with a ROC-AUC score of 1.00.



	accuracy	sensitivity	specificity	precision	F1
Average	0.990	0.980	0.1	0.984	0.981

Figure 4.18 Confusion matrix and average performance of training model containing 162 microbial classes representing 211 species, with an overall species-level accuracy of F1 98.1%.



	accuracy	sensitivity	specificity	precision	F1
Average	1	1	1	1	1

Figure 4.19 Confusion matrix and average performance of independent validation set containing 48/162 microbial classes representing 211 species, with an overall species-level classification accuracy of F1 100%.

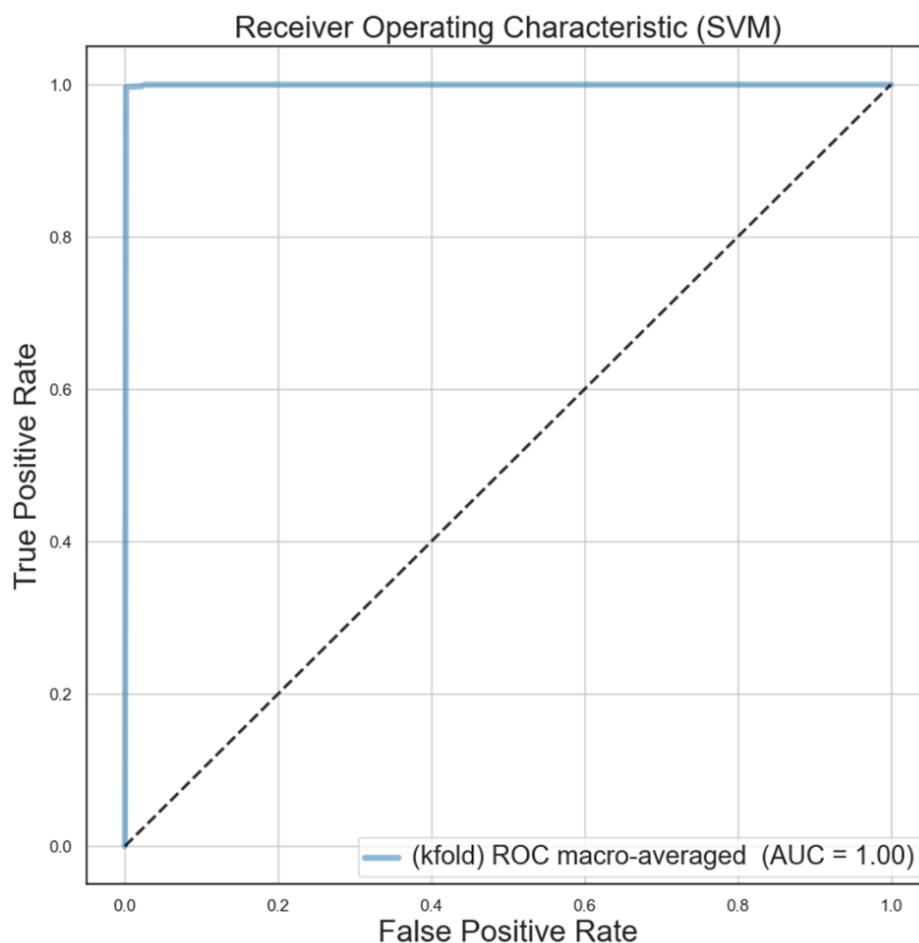


Figure 4.20 Macro-averaged ROC curve for independent validation set containing 48/162 microbial classes representing 211 species, with AUC=1.00

4.4.2.1.1 Recursive Feature Elimination selection (RFE)

Due to the large number of spectral features included in the model, only the top 100 RFE features were extracted to perform stage three of the large-scale model validation and to test the robustness of the model.

At this phase, LC-MS analysis were applied in order to annotate RFE features. 75 of the 100 features were discovered to be present in the LC-MS data and were subsequently annotated using the data received from the corresponding LC-MS/MS data, which can be seen as species-specific biomarkers (see 3.10 for LC-MS assay details).

Here, Table 4.8 below, displays the possible tentative annotations of these statistically significant features, which were mostly phospholipids with C13 isotopes and fatty acids. The feature annotations were aligned with LC-MS/MS data obtained using Data Dependent Acquisition (DDA) mode. Despite the annotations being labelled with expected names, there remains a possibility of inaccuracies or incomplete assignments. This is primarily due to limitations in database matching, particularly in the context of microbial lipidomics, where the comprehensive coverage of lipids in existing databases may be lacking.

The most common species was the deprotonated [M-H]⁻ phosphatidylglycerol (PG) and phosphatidylethanolamine (PE) lipids that are known to make up the biological membranes of the microorganisms. There were also other markers that were specific to individual species, as can be seen from the box plot in figure x, feature 679.4642 *m/z*, which was identified as PG (29:2). It is clearly more prevalent only in *Georgenia muralis*, indicating that it could be a biomarker for this species.

Table 4.8 List of statistically important biomarkers and their lipid and metabolite annotations.

m/z	Matched Mass	Name	Ion	Formula	Error ppm
199.1653	199.1704	FA:12:0	[M-H] ⁻	C12H24O2	25.60
215.0298	215.0319	Glucose	[M+Cl] ⁻	C6H12O6Cl	9.76
227.1959	227.2017	FA 14:0	[M-H] ⁻	C14H28O2	25.52
237.0891	237.0899	FA 10:1;O2	[M+Cl] ⁻	C10H18O4	3.37
240.1249	240.125	LPG 16:1	[M-H] ⁻	C16H19NO	0.41
241.2114	241.2173	FA 15:0	[M-H] ⁻	C15H30O2	24.45
253.2108	253.2168	FA 16:1	[M-H] ⁻	C16H29O2	23.69
255.226	255.233	FA 16:0	[M-H] ⁻	C16H32O2	27.42
256.2303	C13 Isotope of FA 16:0				
257.2074	C13 Isotope of FA 16:0				
265.2245	265.2173	FA 17:2	[M-H] ⁻	C17H30O2	-27.14
267.2268	267.233	FA 17:1	[M-H] ⁻	C17H32O2	23.20
269.2424	269.2486	FA 17:0	[M-H] ⁻	C17H34O2	23.02
279.2331	279.233	FA 18:2	[M-H] ⁻	C18H32O2	-0.35
281.2424	281.2486	FA 18:1	[M-H] ⁻	C18H34O2	22.04
282.2473	C13 Isotope of FA 18:1				
283.2582	283.2643	FA(18:0)	[M-H] ⁻	C18H36O2	21.53
297.2356	297.2435	FA 18:1;O	[M-H] ⁻	C18H34O3	26.57
391.2154	391.225	LPA(16:0)	[M-H2O-H] ⁻	C19H36O6P	24.53
411.3842	411.3844	FA 26:0;O	[M-H] ⁻	C26H52O3	0.48
417.2346	417.2406	LPA(18:1)	[M-H2O-H] ⁻	C21H38O6P	14.38
495.478	495.4783	FA 32:0;O	[M-H] ⁻	C32H64O3	0.60
496.484	C13 Isotope of FA 32:0;O				
515.23	515.2263	LPI 12:0	[M-H] ⁻	C21H41O12P	-7.18
517.2443	517.2339	LPG 16:1	[M+Cl] ⁻	C22H43O9P	-24.35
518.2364	C13 Isotope of LPG 16:1				

603.4908	603.4761	DG 32:0	[M+Cl]-	C35H68O5	-12.37
662.4848	662.4766	PE 30:0	[M-H]-	C35H70NO8P	-7.66
665.445	665.4399	PG 28:0	[M-H]-	C34H67O10P	-16.87
675.4721	675.4607	PG O-30:2	[M-H]-	C36H69O9P	
676.4901	C13 isotope of PG O-30:2				-12.65
679.4642	679.4556	PG 29:0	[M-H]-	C35H69O10P	
680.4672	C13 isotope of PG 29:0				-13.79
688.5018	688.4923	PE 32:1	[M-H]-	C37H72NO8P	
689.4987	C13 isotope of PE 32:1				-17.47
692.4993	692.4872	PS O-30:0	[M-H]-	C36H72NO9P	-11.34
693.4791	693.4712	PG 30:0	[M-H]-	C36H71O10P	
694.4829	C13 isotope of PG 30:0				
695.4863	C13 isotope of PG 30:0				-16.51
702.5195	702.5079	PE 33:1	[M-H]-	C38H74NO8P	
703.5182	C13 isotope of PE 31:1				-19.42
705.4849	705.4712	PG 31:1	[M-H]-	C37H71O10P	
706.4895	C13 isotope of PG 31:1				-13.85
707.4967	707.4869	PG 31:0	[M-H]-	C37H73O10P	
708.5022	C13 isotope of PG 31:1				
709.4972	C13 isotope of PG 31:1				-16.65
714.5198	714.5079	PE 34:2	[M-H]-	C39H74NO8P	-13.82
716.5335	716.5236	PE 34:1	[M-H]-	C39H76NO8P	-16.31
717.4829	717.4712	PG 32:2	[M-H]-	C38H71O10P	-14.04
719.497	719.4869	PG 32:1	[M-H]-	C38H73O10P	
720.502	C13 isotope of PG 32:1				-10.67
721.5102	721.5025	PG 32:0	[M-H]-	C38H75O10P	
722.5155	C13 isotope of PG 32:0				
723.5193	C13 isotope of PG 32:0				-13.18
728.5332	728.5236	PE 35:2	[M-H]-	C40H76NO8P	
729.5286	C13 isotope of PEE 35:2				-20.86
733.5178	733.5025	PG 33:1	[M-H]-	C39H75O10P	
734.5221	C13 isotope of PG 33:1				-14.55
735.5289	735.5182	PG 33:0	[M-H]-	C39H77O10P	
736.5343	C13 isotope of PG 33:0				-12.79
742.5482	742.5387	PE(36:2)	[M-H]-	C41H77NO8P	
743.5467	C13 isotope of PE 36:2				-25.35
745.5225	745.5036	PG(34:2)	[M-H]-	C40H74O10P	
746.517	C13 isotope of PE 34:2				-12.84
747.5278	747.5182	PG(34:1)	[M-H]-	C40H76O10P	
748.5316	C13 isotope of PE 34:1				-11.47
749.5424	749.5338	PG34:0	[M-H]-	C40H78O10P	
750.5502	C13 isotope of PG 34:0				-13.92
761.5439	761.5333	PG 35:1	[M-H]-	C41H79O10P	-8.91
763.5563	763.5495	PG 35:0	[M-H]-	C41H81O10P	
764.5645	C13 isotope of PA 41:5				-15.50
773.5453	773.53331	PG 36:2	[M-H]-	C42H79O10P	
774.5484	C13 isotope of PG 36:2				-1.60
899.5664	899.564957	PI 39:4	[M-H]-	C48H85O13P	-5.76
927.6016	927.596258	PI 41:4	[M-H]-	C50H89O13P	-5.75

A final optimised model was re-trained using the 75 biomarkers. As a result, an average classification accuracy F1 score of 92.8% was obtained through CV. By achieving model accuracy of more than 90% with only 75 statistically significant characteristics, the model shows great promise as a biomarker-based predictive model for unknown microbial species.

Table 4.9. Average classification results of training model developed using 75 statistically important biological biomarkers.

	accuracy	sensitivity	specificity	precision	F1
Average	0.964	0.928	1	0.937	0.928

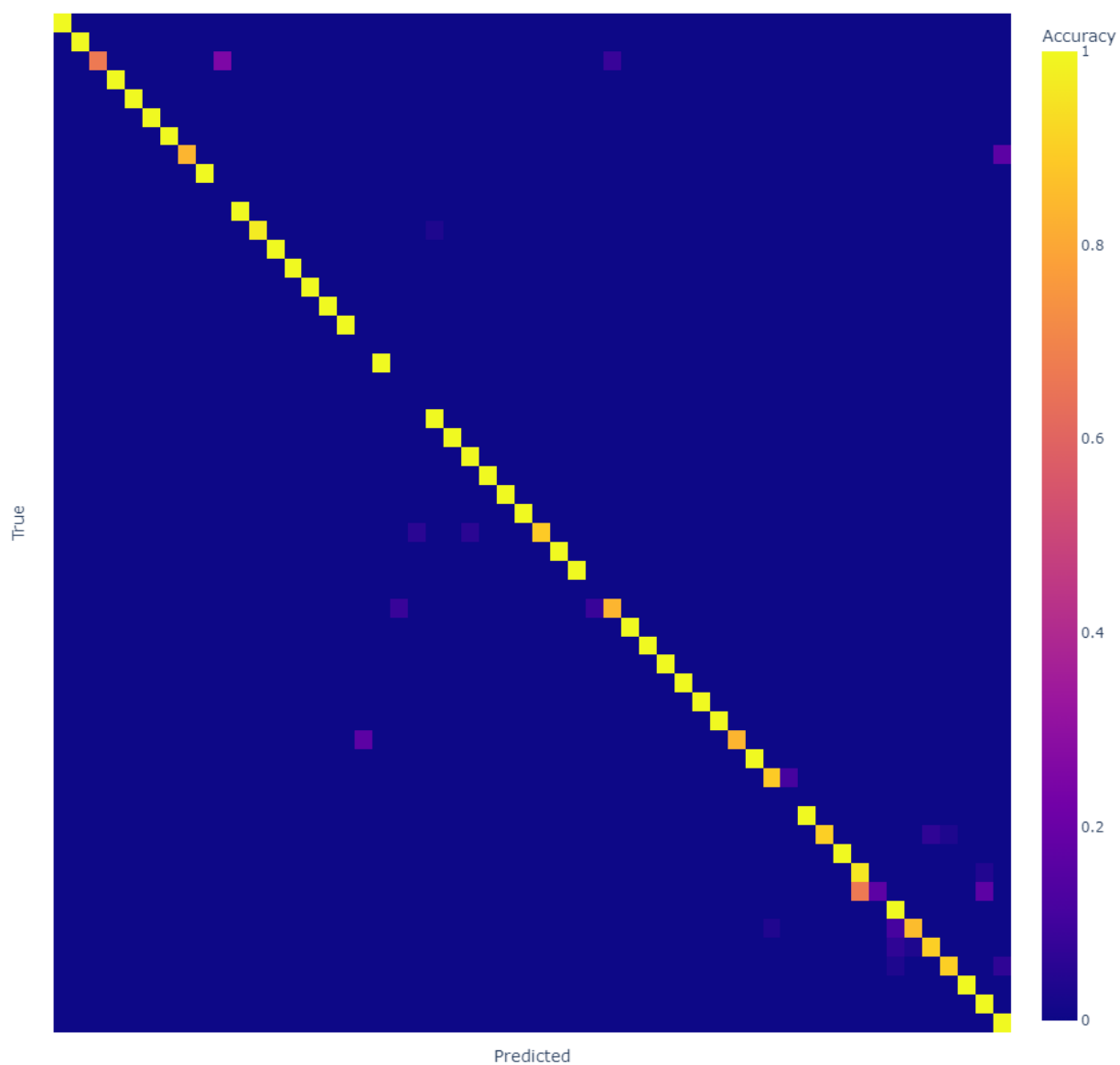
An independent validation set containing 48 out of 162 species was also used to test and verify the generalisability of the optimised model.

In this case, as is shown in Figure 4.21, the average accuracy of F1 score that was acquired from the independent validation set was 83.7%. Even though the classification accuracy was lower than 90%, this was primarily due to the fact that 6/48 species were completely misidentified, which means that 75 annotated features were insufficient to match the profile of every unknown sample. Nonetheless, this accuracy is expected to be easily enhanced to be higher than 90% by expanding the number of annotated biomarkers, which was arbitrarily set to minimise the computational intensity and to explore the limits of the approach.

The remaining 42/48 species were able to be identified with a classification accuracy of 90% or higher. Tables 4.13 and 4.14 in the *Appendix* contain the complete listing of individual classification results for the RFE training model and the independent validation set, respectively.

Figure 4.22 below provides an illustration of an example of one of the biomarkers, specifically 679.4642 *m/z* (PG 29:0). This example indicates very clearly that there is a significant difference in the normalised intensity of *Georgenia muralis* compared to the other species that were used for the training model.

When assessing the outcomes of all three stages of validation for the large-scale classification model, it became abundantly clear that the pre-processing parameters, in conjunction with the method of data collection (please refer to the *Chapter 3, Materials and Methods* for details on the LA-REIMS setting that was used), produced a stable, high-accuracy model that possessed encouraging potential for biomarker-specific multivariate predictive modelling.



	accuracy	sensitivity	specificity	precision	F1
Average	0.975	0.951	0.999	0.858	0.837

Figure 4.21 Confusion matrix and average performance of RFE feature independent validation set containing 48/162 microbial classes representing 211 species, with an overall species-level classification accuracy of 83.7%.

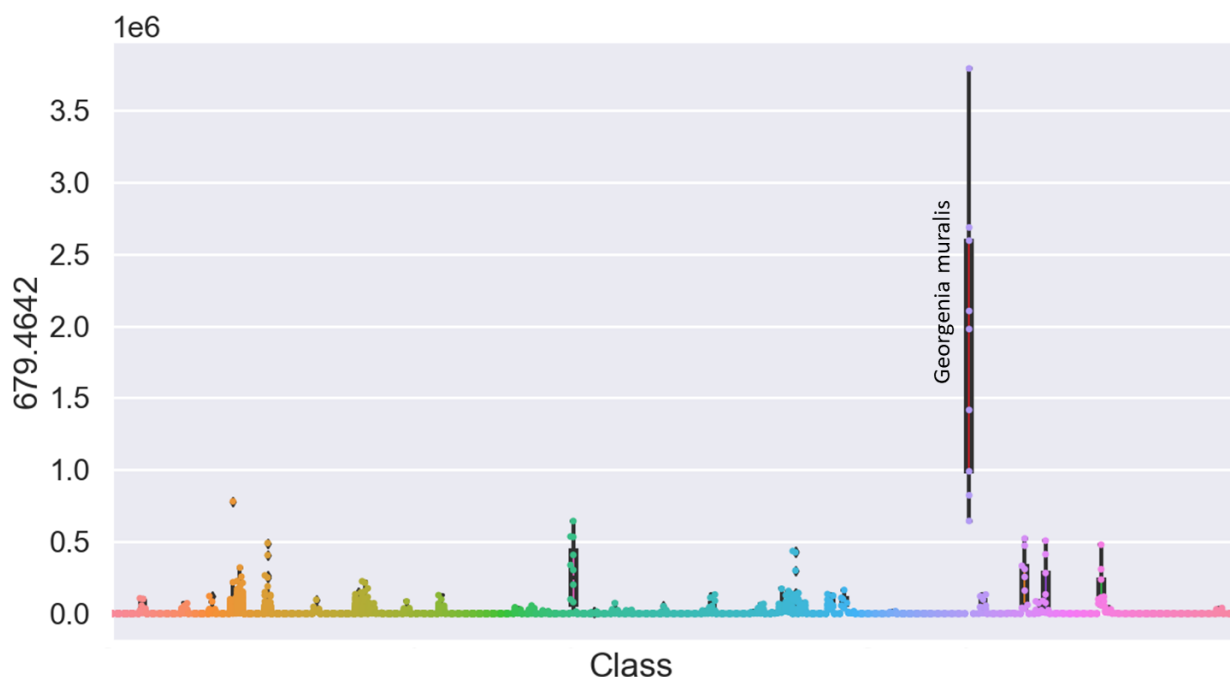


Figure 4.22 Box plot of RFE feature 679.4642 m/z (PG 29:0) with the highest abundance obtained in *Georgenia muralis* species in the training model.

4.5 Conclusion

In this chapter, the aim was to use a small-scale microbial species LA-REIMS acquired spectral data set in order to investigate all possible variations of data pre-processing hyperparameters in order to obtain a stable, repeatable, and reproducible multivariate classification model that can then be scaled up to a large-scale model.

First, by using three different levels of model validation, it was determined that the final parameters selected, which were ‘75% peak filter’, ‘Pre-processing 2’ with ‘TIC+Log’ normalisation, provided the highest accuracy. The training model had an average accuracy of F1 score of 99.73%, and the independent validation and RFECV validation both had an average accuracy of F1 score of 100%.

Second, the small-scale model was shown to be unaffected by the storage of samples in the refrigerator for up to 72 hours prior to the analysis and to have good repeatability between numerous days of data collection and analysis. In addition, a significant reproducibility study that utilised a time gap of three years between the collection of data for the training model and

the collection of data for an independent validation set, as well as the utilisation of different instruments produced an average classification accuracy of unknown species that was greater than 90%.

In the final results, a large-scale multivariate classification model that was developed using 162 microbial classes representing 211 species yielded an average accuracy of F1 score of 98.1% after Kfold cross validation and an average classification accuracy of F1 score of 100% for independent validation set. Furthermore, by employing the RFE features only approach, 75 biological features that were statistically significant were identified. These features were then utilised to re-validate the entire data matrix created model and explore the feasibility of biomarker-specific multivariate modelling. With an average accuracy of F1 score of 92.8% for the RFE feature training model and a F1 score of more than 90% for correctly identifying 46 species in the independent validation set, it showed an adequate amount of potential for biomarker-specific multivariate predictive modelling, which could be accomplished by increasing the number of statistically important biomarkers.

Overall, the conclusion that was reached was that the LA-REIMS platform, when combined with the peak picking platform for data analysis, had a significant potential to be incorporated into the industrial pharmaceutical environmental monitoring pipeline. This would result in faster identification results at high accuracy for unknown microbial contaminants and enormous cost savings for the company. By facilitating accurate ID results quickly, cost savings can be gained by expediting investigation and initiating corrective actions. The importance of an accurate ID in such circumstances is critical to leading the investigation direction. Furthermore, the streamlined process eliminates sample preparation processes, saving time and money.

Chapter 5

Environmental Sample Analysis (fungi)

Chapter 5 Environmental monitoring (fungi)

5.1 Introduction

Due to the complexity and difficulty of identifying fungal species, it was essential to develop efficient and effective filamentous fungi identification method for industrial pharmaceutical environmental monitoring using the LA-REIMS platform.

Many filamentous fungi can be identified at the species level using molecular techniques such as PCR, ribosomal DNA (rDNA) internal transcribed spacer (ITS) sequence analysis, or 18S ribosomal RNA (rRNA) gene sequencing. However, due to its multistep workflow, this technology is costly, labour-intensive, and poses a high risk of environmental contamination²⁶⁰.

The primary issue raised by filamentous fungi is the presence of significant metabolic changes derived based on culture age, even between different subcultures of the same strain²⁶¹. The LA-REIMS platform was used to find a solution, which included the development of a broader and more complicated database containing profiles of a greater number of strains and cultures of various ages.

5.2 The aims of the chapter

Due to the difference in size and growth rate of filamentous fungi, the first objective of this chapter was to use LA-REIMS low-resolution imaging to evaluate the distribution of small molecules and phospholipids within the fungal colony in order to determine the optimal sampling strategy for the development of a classification model.

The second aim of this chapter was to develop and optimise a robust and reliable predictive classification model using an in-house generated industrial pharmaceutical fungi spectral database, which may be exploited as a machine learning-based fungal identification pipeline in combination with the LA-REIMS platform. In addition, due to the known instability of filamentous fungi metabolic profiles, the objective was to validate the developed classification model by analysing the same fungal strains six months after the initial reference model has been trained following multiple subcultures to confirm the pipeline's ability to identify samples accurately.

5.2.1 Current industrial pharmaceutical environmental monitoring for fungi pipeline

The current industrial pharmaceutical environmental monitoring for fungi pipeline is akin to the one outlined in *Chapter 4*, section 4.3.

5.2.1.1 Potential LA-REIMS pipeline for environmental fungi monitoring

The proposed LA-REIMS pipeline for monitoring environmental fungi has the same structure as described in *Chapter 4*, section 4.3.1.

5.3 Results and Discussion

5.3.1 1mm Low-resolution fungi imaging

Fungi and bacteria come under very different categories. Fungi growth period is much longer than bacteria (up to 14 days or more), it could have different height or regions of the colony within the same sample and also different density, shape or size. In order to improve fungal speciation using LA-REIMS, 1mm resolution REIMS imaging has been done to find most optimal sampling area by identifying fungal metabolite distribution within the colony over five days growth period.

Five different features (e.g., size, density) fungi were selected for 1mm LA-REIMS imaging experiment:

1. *Fusarium spp*
2. *Aspergillus niger*
3. *Penicillium spp*
4. *Aspergillus fumigatus*
5. *Trichophyton tonsorans*

After five days of growth, the photographs of five different fungal colonies are presented in Figure 5.1.



Figure 5.1 1. *Fusarium* spp, 2 .*Aspergillus niger*, 3. *Penicillium* spp, 4. *Aspergillus fumigatus*, 5. *Trichophyton tonsurans* photographs after 5 days of growth at 30°C O₂ atmosphere, on SAB agar plates.

Samples were cultured, run and analysed using the following methods described in *Chapter 3 Materials and Methods*, sections 3.4.4, 3.7.3 and 3.8.2.

Only a portion of the fungal colony was imaged due to the size of the colonies after 48 hours of growth, but it still contained different regions of interest (e.g., the edge and the middle). After data were acquired, the image was mapped by separating agar, edge and the middle areas within the colony (Figure 5.2). The intensity map/image analysis for specific areas was performed using an in-house created R script. In the edge and middle, the 100 most intense metabolic features in ranges of 100-500 *m/z* and 600-900 *m/z* were picked and studied independently.

Figure 5.3 illustrates an example of *Aspergillus niger* from day one to day five growth and the image mapping of the edge and the middle of the colony.

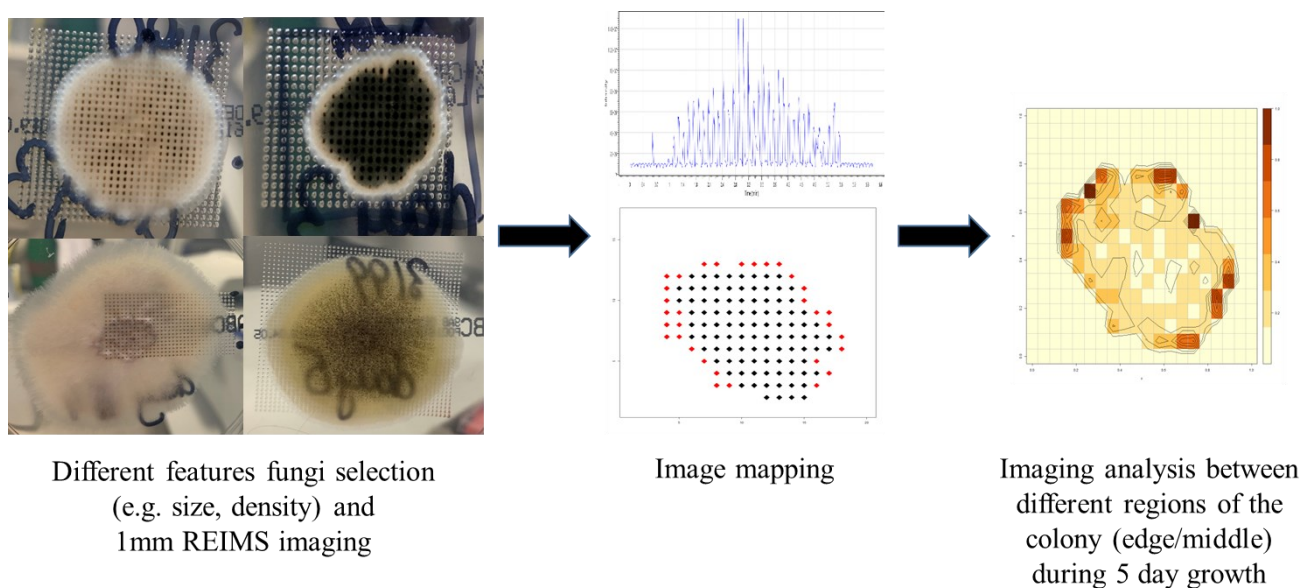


Figure 5.2 Workflow of 1mm LA-REIMS fungal imaging pipeline.

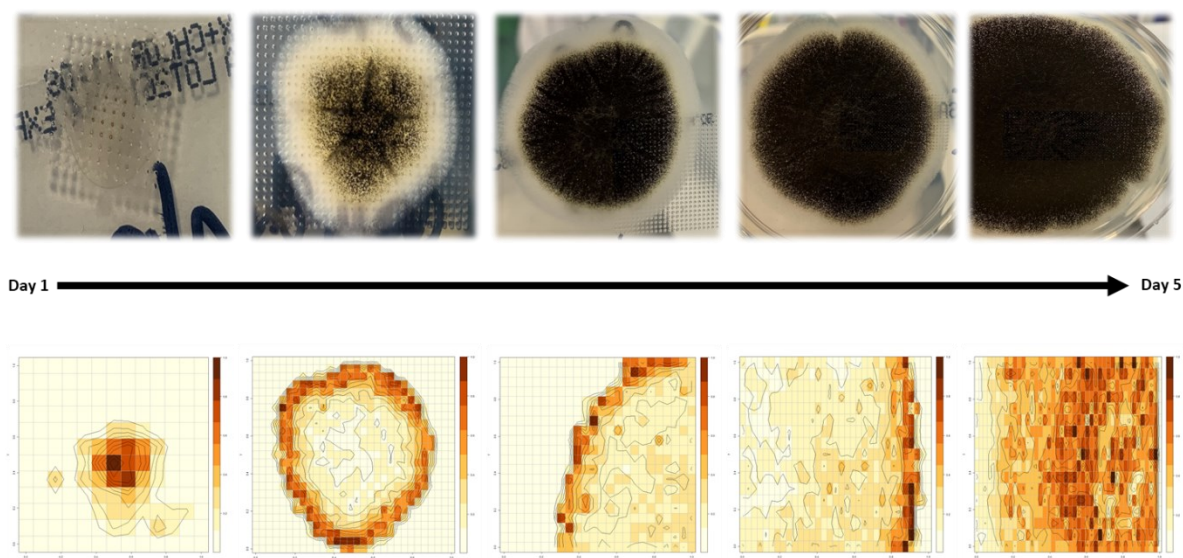


Figure 5.3 *Aspergillus niger* image maps of the growth from day one to day five.

Upon examination of various metabolic characteristics of fungal colonies, it has been determined that colonies of analysed fungi exhibit a heightened metabolic profile at the edge of the colony following 24 hours of growth.

As an example, Figure 5.4 shows that the fatty acid (FA 18:1) 281.25 m/z has a greater intensity in the middle region. However, in Figures 5.5 and 5.6, from day two to day five, the same fatty acid exhibits a higher intensity at the edge of the colony as compared to the middle region.

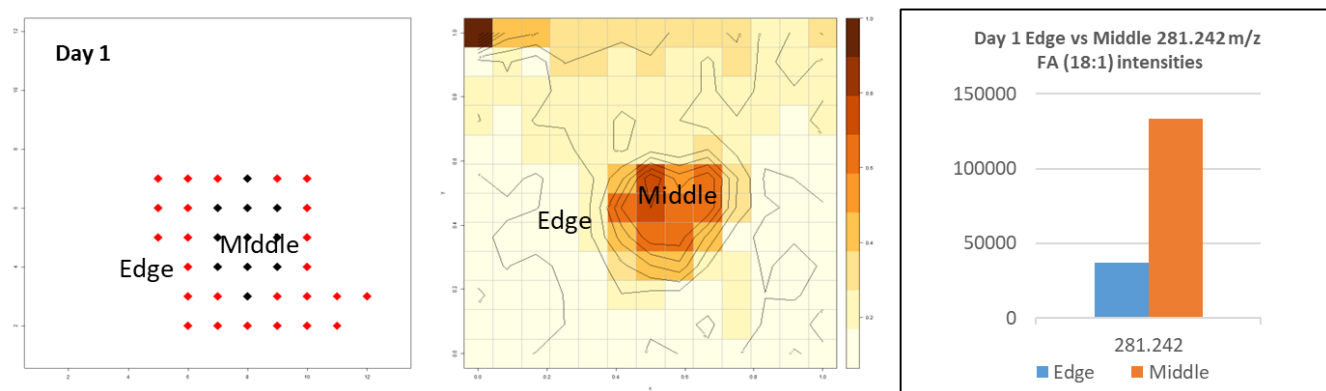


Figure 5.4 *Aspergillus niger*, 281.25 m/z (FA 18:1) middle vs edge intensity comparison after 24h growth.

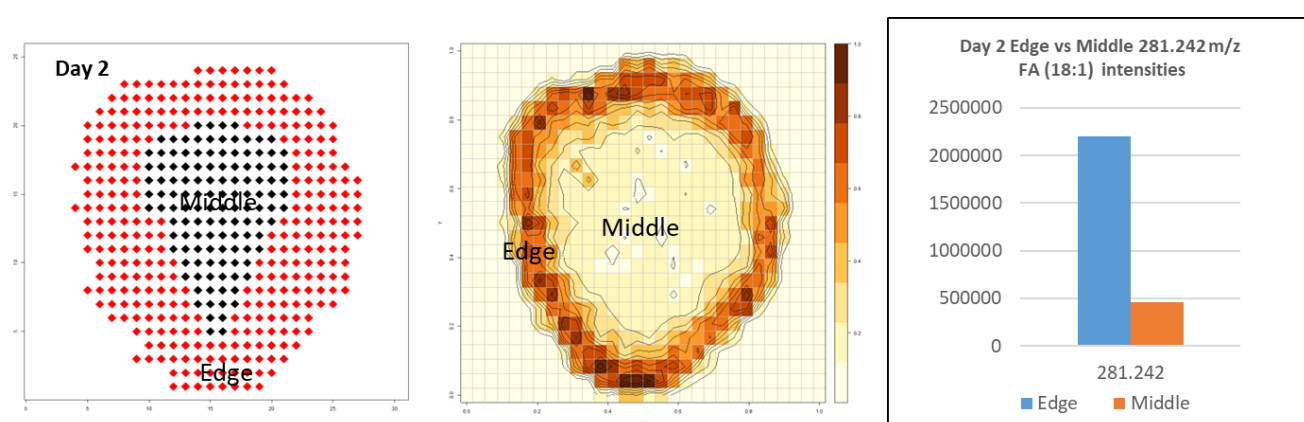


Figure 5.5 *Aspergillus niger*, 281.25 m/z (FA 18:1) middle vs edge intensity comparison after 48h growth.

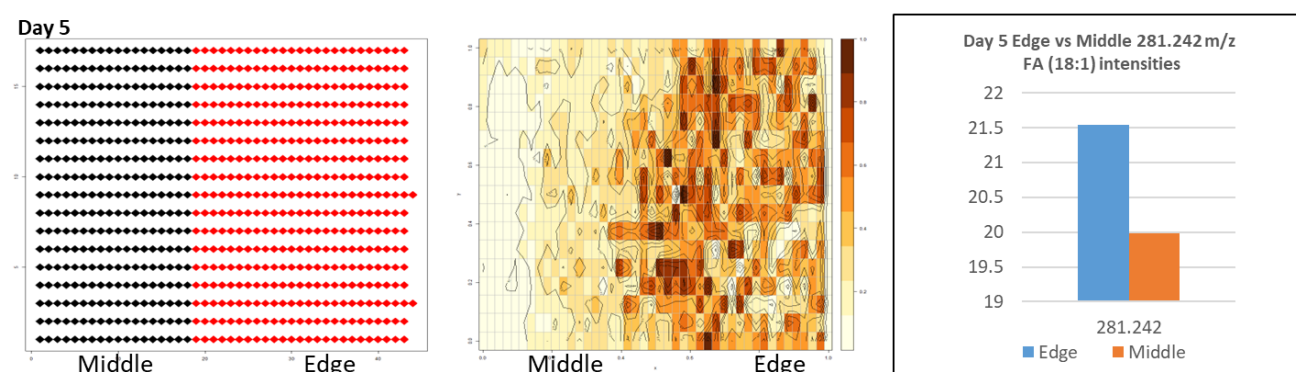


Figure 5.6 *Aspergillus niger*, 281.25 m/z (FA 18:1) middle vs edge intensity comparison after 5 day of growth.

Figure 5.7 shows the overall intensities averages of *Aspergillus niger* fungi from day one to day five growth.

Fungal hyphae can exhibit different growth patterns depending on their location within a colony. Hyphae at the edges may be actively growing and branching to explore new nutrient sources, while those in the middle may be more stationary due to resource depletion. These differences in growth dynamics can affect metabolic activity.

As result of this, it has been decided that the sampling for the classification model should consist of mixture of burns covering the edge and the middle to make sure that variations of different growth stages are captured in the model.

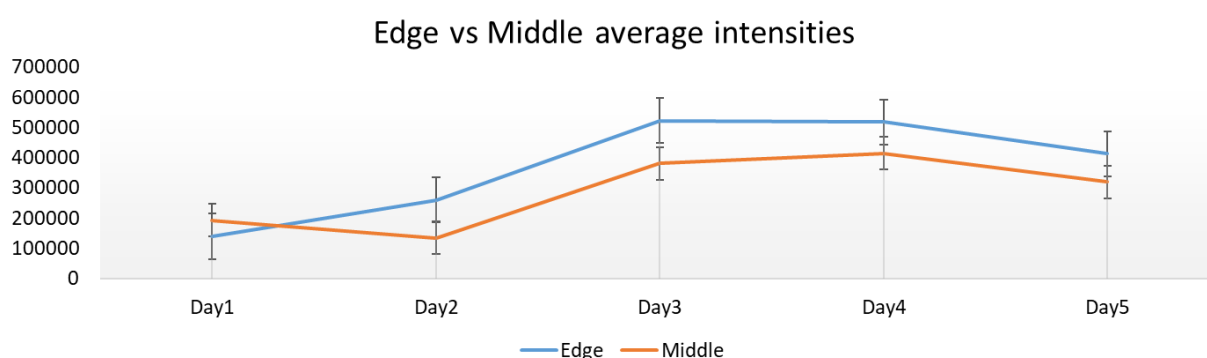


Figure 5.7 Comparison of *Aspergillus niger* edge and middle of the colony 100 most intense metabolites overall intensities averages from day one to day five growth.

5.3.2 Development and validation of classification model for fungal species

A total of 359 fungi test isolates were analysed using LA-REIMS platform. Data set contained a mixture of 12 different widely seen pharmaceutical environmental fungal contaminant species:

1. *Trichophyton interdigitale* (TINT)
2. *Trichophyton equinum* (TEQU)
3. *Aspergillus fumigatus* (AFUM)
4. *Aspergillus brasiliensis* (ABRA)
5. *Aspergillus niger* (ANIG)
6. *Aspergillus sydowii* (ASYD)
7. *Penicillium simplicissimum* (PSIM)
8. *Penicillium brasilianum* (PBRA)

9. *Penicillium meleagrinum* (PMEL)

10. *Penicillium adametzii* (PADA)

11. *Penicillium janthinellum* (PJAN)

12. *Purpureocillium lilacinum* (PLIL)

The data were acquired from cultures that were maintained for a period of three to five days. An in-house peak-picking platform with a ‘75% peak filter’ detection setup was used to generate a data matrix with class labels for post-processing, with 663 peaks identified across the spectral range. Following the results established in the development and validation of a robust and accurate bacteria and yeast environmental model, fungi classification model was built applying the same parameters with the exception of the normalisation strategy (‘Log’ instead of ‘TIC+Log’): SVM machine learning model with ‘75% peak filter’ and ‘Pre-processing 2’ with ‘Log’ normalisation parameters. Please refer to the *Chapter 4* for more details on model development parameter selection.

The training model was then cross validated using leave one sample out validation, as well as an independent validation set that was generated by analysing six fungal species that has been undergoing multiple subculturing and storage for 6 months after the initial reference model sampling.

‘Log’ normalisation was empirically chosen to enhance class separation on PCA score plot. Figure 5.8 demonstrates much better fungal class separation from 12 different species using ‘75% peak filter’ with ‘Log’ normalisation. When comparing PC1 to PC3 in Figure 5.8, there is a distinct clustering demonstrating that the data contains species-specific features that drive class separation.

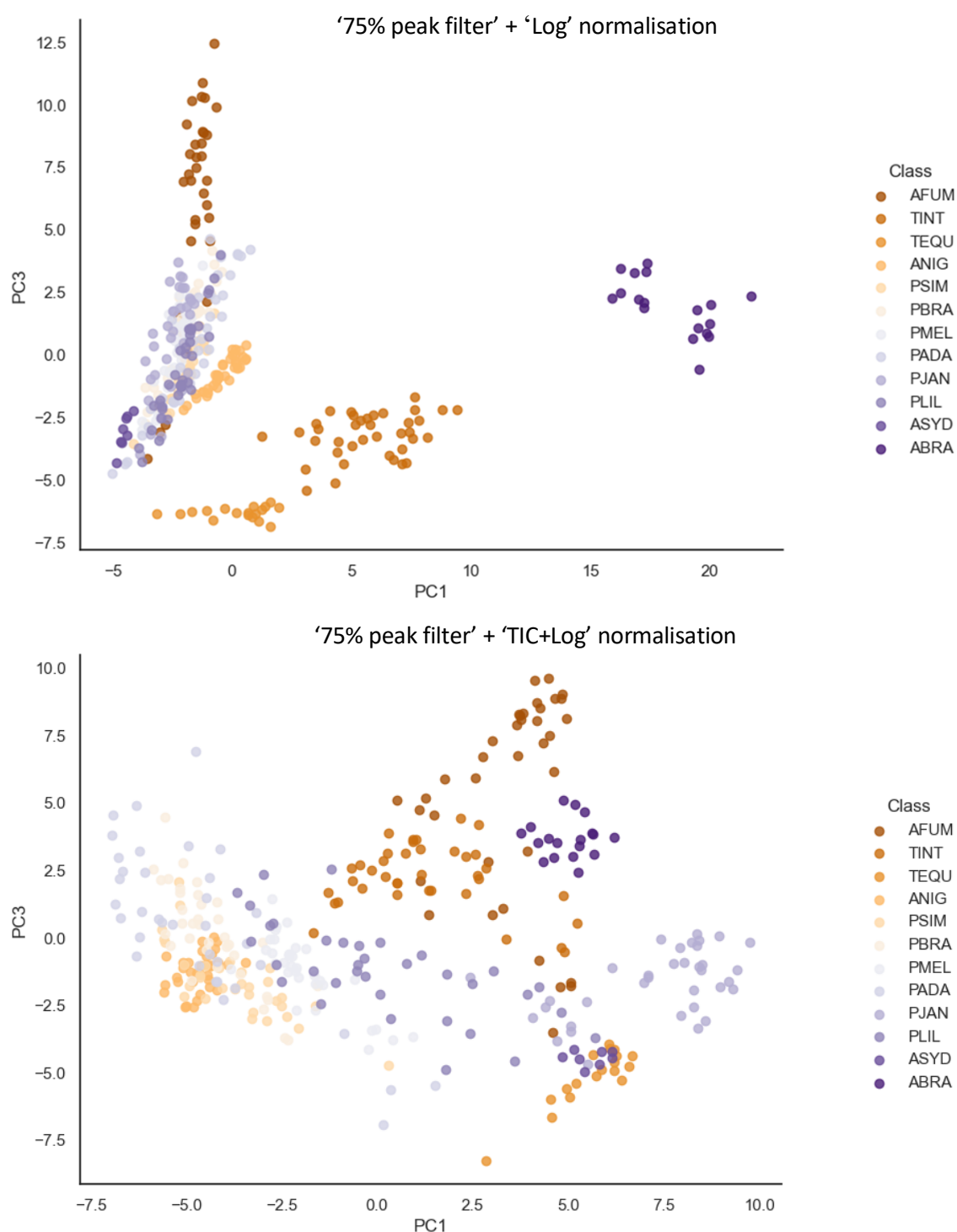


Figure 5.8 PCA scores plots displaying PC1 vs PC3 of two different data normalisation methods of the microbial species used to develop a large-scale classification training model.

Initial reference classification model containing 359 fungal isolates representing 12 different species, achieved a classification accuracy of F1 score of 100% after leave one sample out cross validation (Figure 5.9).

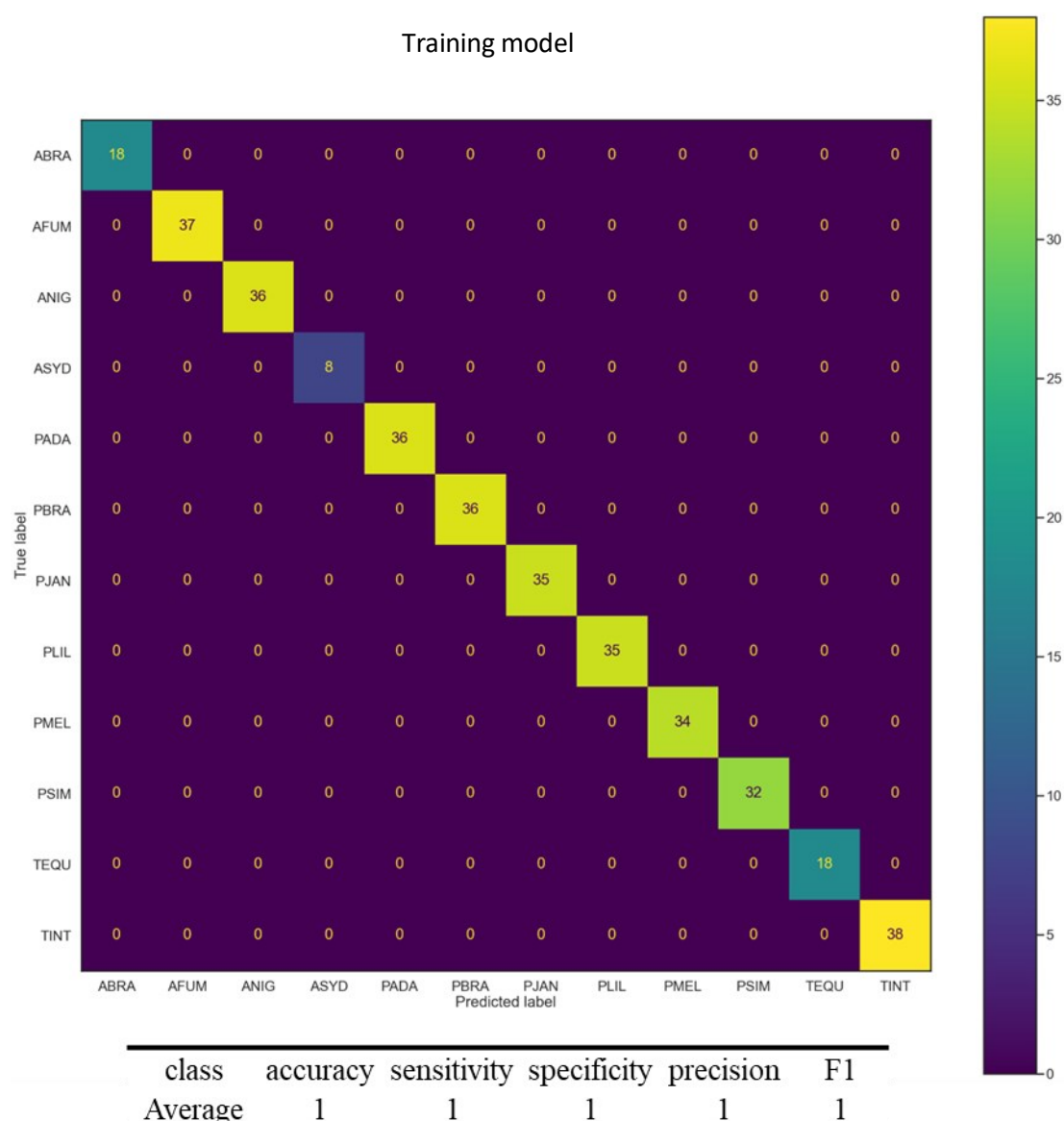


Figure 5.9 Confusion matrix and average training model classification accuracy representing model build using ‘75% peak filter’, ‘Pre-processing 2’ and ‘Log’ pre-processing parameters.

Independent validation using an independent validation set containing six of the twelve fungal species from the reference model and analysed six months later after multiple subculturing achieved classification accuracy of F1 score of 87% (Table 5.10).

Table 5.10 Average performance of the independent validation set, consisting of six out of the twelve fungal species, analysed six months after the initial reference model isolates and multiple subculturing. Overall species-level classification accuracy 87%.

class	accuracy	sensitivity	specificity	precision	F1
AFUM	1.00	1.00	1.00	1.00	1.00
ASYD	0.90	0.80	1.00	1.00	0.89
PBRA	0.80	0.60	1.00	1.00	0.75
PJAN	0.90	0.80	1.00	1.00	0.89
PLIL	0.98	1.00	0.96	0.83	0.91
PMEL	0.94	1.00	0.88	0.63	0.77
Average	0.92	0.87	0.97	0.91	0.87

High classification accuracy results, especially when considering that the model includes numerous same genus isolates (e.g., *Aspergillus* and *Penicillium* species) and isolates analysed after multiple subcultures, demonstrate an exceptional ability to identify and categorise unknown fungus directly from the agar growth plate, which no other method can do.

LA-REIMS platform has demonstrated the capability to be identify species specific features and shifts in their intensities, which facilitate the differentiation of species during the classification process. Box plots in the Figure 5.10 shows an example of two of the features from 12 analysed fungal species. Both features: 391.2217 m/z (LPA 16:0) and 709.4897 m/z (PA 37:4) shows significant variation in peak intensity across the various fungal species.

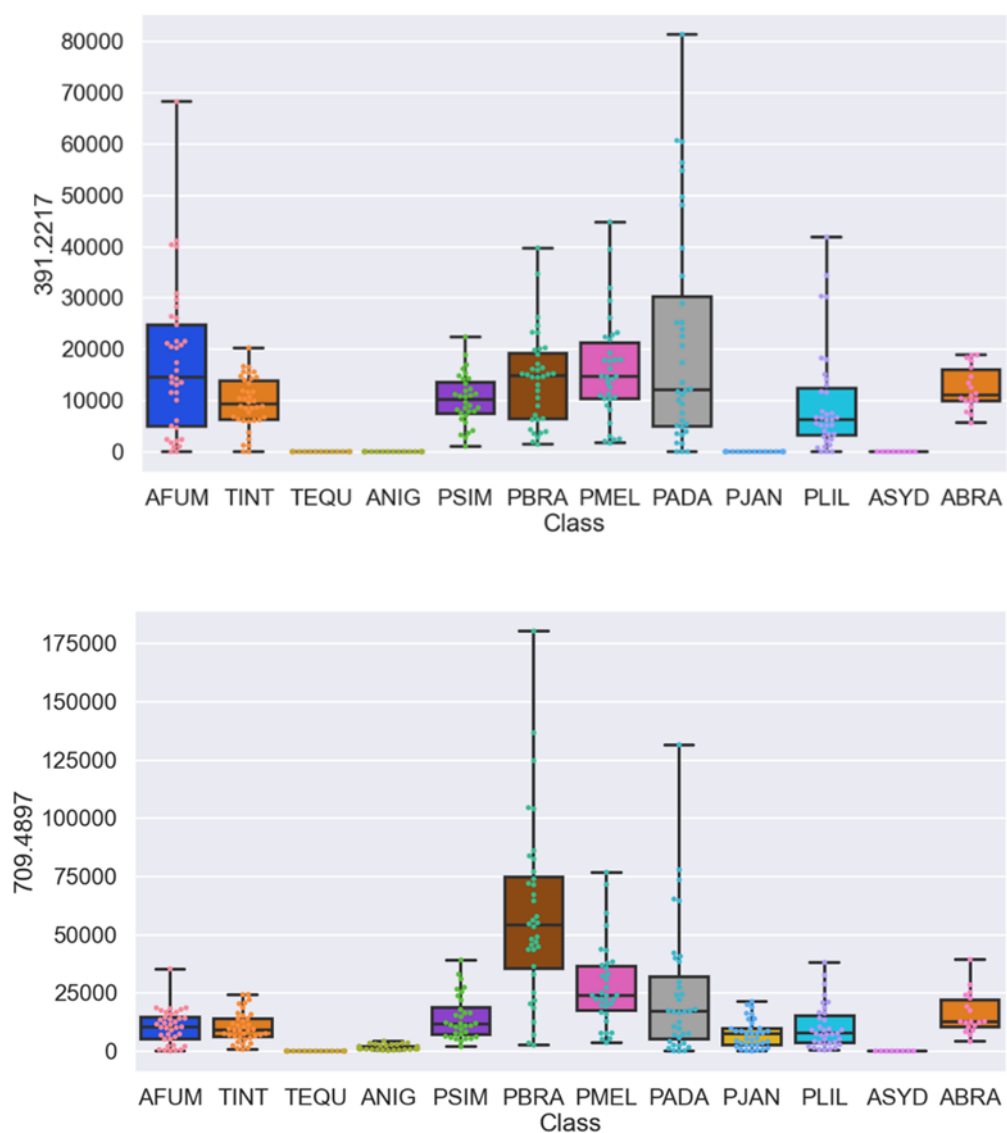


Figure 5.11 Box plots of two metabolic features (391.2217 m/z - LPA 16:0, 709.4897 m/z - PA 37:4) appearing across 12 different fungal species and their peak intensity variations.

5.4 Conclusion

In this chapter, the aim was to evaluate and demonstrate the possibility of LA-REIMS platform to be used as fungal environmental contaminants identification method for industrial pharmaceuticals.

First, to account for the intricate nature of filamentous fungi, such as their growth rate, size, and density, 1mm LA-REIMS low resolution imaging was utilised. This method allowed to assess the distribution of metabolic activity across different fungal species. Our aim was to identify the most suitable sampling approach for data collection, which would facilitate the development of a robust fungal classification model. The results indicate that the metabolic intensity at the edge of the fungal colony is higher than that at the middle of the colony, starting from 24 hours of growth and continuing thereafter. Due to possibility of fungal having different growth patterns within the colony it has been decided that the sampling for the classification model should consist of mixture of burns covering the edge and the middle to make sure that variations of different growth stages/patterns are captured in the model.

A classification model with high accuracy was then developed and validated through cross-validation and independent validation. The process of independent validation was used to tackle the primary challenge in identifying filamentous fungi from different time points, which is usually very difficult due to metabolic changes resulting from culture age.

Even after subculturing over a period of six months, the LA-REIMS-enabled classification still yielded an average accuracy of 92% and an F1 score accuracy of 87% on the independent validation set, indicating significant potential for the technique.

Chapter 6

Other applications for LA-REIMS in
pharmaceutical microbiological
monitoring

Chapter 6 Other applications for LA-REIMS in pharmaceutical microbiological monitoring

6.1 Aims of the chapter

The aims of this chapter were to evaluate other potential applications for the LA-REIMS platform when integrated into the industrial pharmaceutical microbiological monitoring workflow.

Aim 1

The first aim of this chapter was to look into the potential of directly analysing water contaminants growth from the agar plate's filter membrane on the LA-REIMS platform. Also, to use the acquired data to create a water pathogens classification model, which will be used in combination with the LA-REIMS platform for predictive modelling-based water pollutant identification directly from membrane filter agar plates.

Aim 2

The second aim of this chapter was to look into the feasibility of using the LA-REIMS platform to analyse specified microbial species grown in the presence of drugs or excipients. Also, to use acquired data to develop a classification model to confirm the presence and identification of these specified organisms in the presence of drug or excipient.

6.2 Water analysis

6.2.1 Introduction

Pharmaceutical water systems play a critical role in the processing, formulation, and manufacture of various pharmaceutical products, including dosage forms, APIs, API intermediates and analytical reagents. Water serves as a raw material, inactive ingredient, medicinal vehicle, and solvent in these applications, as well as in cleaning processes. Therefore, it is essential to ensure that pharmaceutical water systems are well-maintained, validated, and monitored^{36,262}

Pour plates, spread plates, membrane filtration, and most probable number (MPN) assays are examples of traditional cultural procedures for microbiological testing of water. These approaches are often simple to implement and give sufficient sample processing throughput¹⁷. Membrane filtration is the most often utilised method; hence it was chosen to adapt on the LA-REIMS platform. The membrane filtration method follows the same procedures as the environmental sample monitoring method (*Chapter 4*), with the exception of the first step of sample collection and filtration, where water samples are collected, filtered, and the filter membrane is incubated on an agar plate for the approximately 5 days.

6.2.1.1 Current industrial pharmaceutical water analysis pipeline

The current industrial pharmaceutical water analysis procedure involves filtering collected water samples via filter membrane funnels, then removing the membrane and placing it on an agar plate for incubation for up to 5 days, to allow any microbiological contaminants trapped in the filter to grow. After the initial incubation period, any colonies spotted on the filter membrane in the agar plate are subcultured onto the purity plate (two per plate) and placed to the incubator for the secondary incubation, which can take up to 5 days. Following this, Gram staining is conducted on any detected growth, and microbiological contaminants are then tested using either a VITEK or a MicroSEQ equipment, which requires extra sample preparation prior to analysis (Figure 6.1).

6.2.1.2 Potential LA-REIMS pipeline for water analysis

As shown in the Figure 6.1 below, in comparison to current industrial pharmaceutical pipeline, the suggested new pipeline using LA-REIMS would allow the water contaminants growth obtained on the filter membrane to be analysed directly on the LA-REIMS set up. The purity plate and gram stain phases would be completely removed by this new proposed approach, allowing for a significantly quicker turnaround for identifying water contaminants.

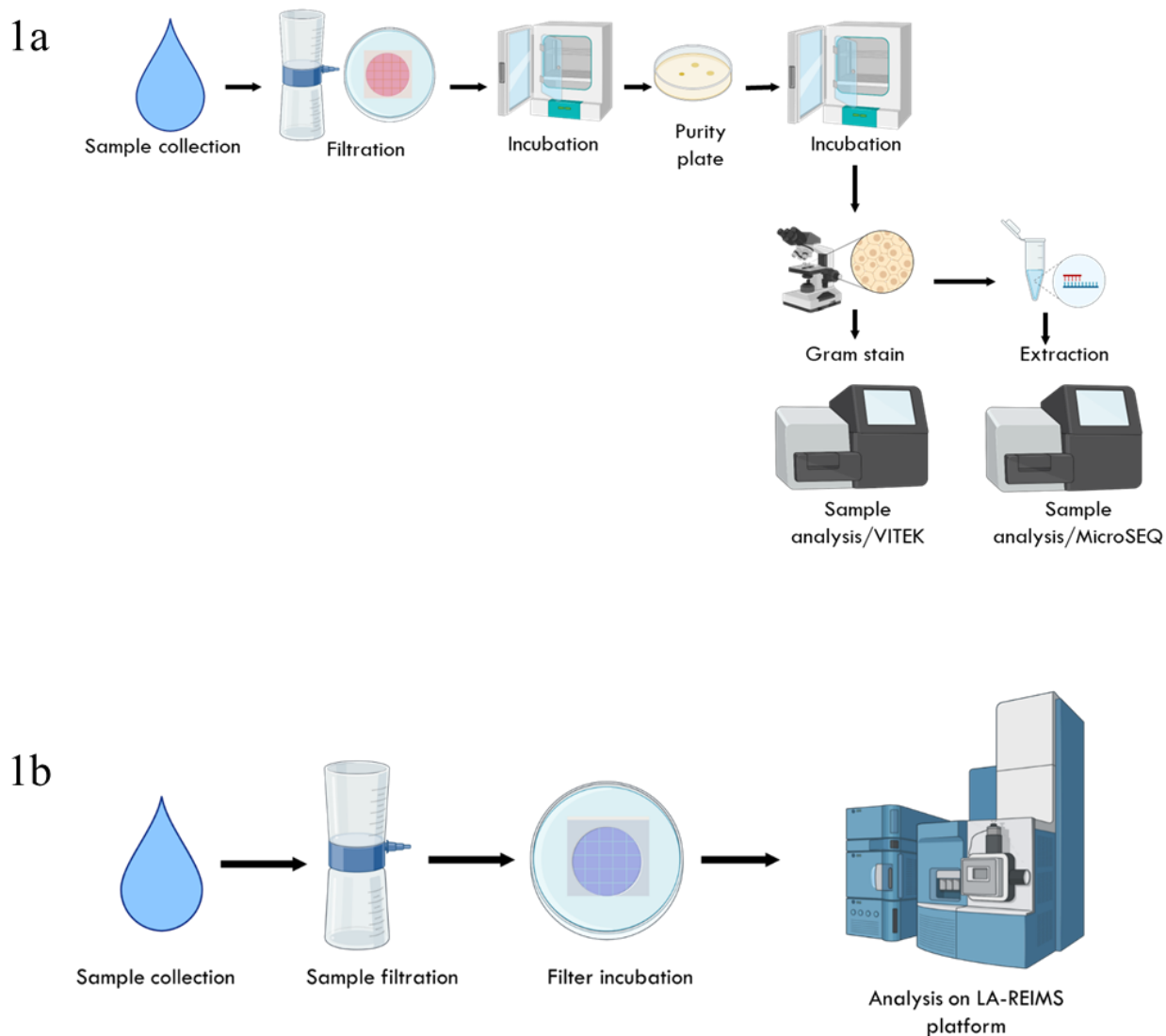


Figure 6.1 Current industrial pharmaceutical water contaminant analysis process (1a) (current timeline 9-18 days depending on the species) in comparison to proposed LA-REIMS method (proposed timeline - 5 days) (1b).

6.2.2 Results and Discussion

6.2.2.1 Data collection from filter membrane using LA-REIMS

A total of 446 water contaminants test isolates were analysed straight from the filter membrane on the agar plate using the LA-REIMS platform, which contained 10 different most commonly found water borne species:

- *Acinetobacter baumannii* (ABAU)
- *Citrobacter freundii* (CFRE)
- *Enterobacter aerogenes* (EAEG)
- *Klebsiella pneumoniae* (KPNE)
- *Listeria monocytogenes* (LMON)
- *Proteus mirabilis* (PMIR)
- *Fusobacterium necrophorum* (FNEC)
- *Staphylococcus aureus* (SAUR)
- *Ralstonia pickettii* (RPIC)
- *Escherichia coli* (ECOL)

For the purpose of the analysis, both biological and technical replicates were combined, with 15 biological isolates for each type of contamination.

The filter membrane had negligible effect on the spectral signal, as can be seen in the Figure 6.2 below, which shows a representative spectrum of the two water pollutants that were investigated.

Even though the membrane filters were washed in 150 mM ammonium acetate solution (please refer for the method to the *Chapter 3*, section 3.4.2) some background was seen in the range of 450 m/z to 550 m/z , which may have been remaining salts from the media or filter cellulose background. However, the intensity of this background was much lower in comparison to the signal that was obtained in the phospholipid region, which is 650 m/z to 800 m/z , as well as the fatty acid region, which is 150 m/z to 300 m/z , the noise was consistent in all sample, moreover it did not interfere with the further analysis and building of the classification model.

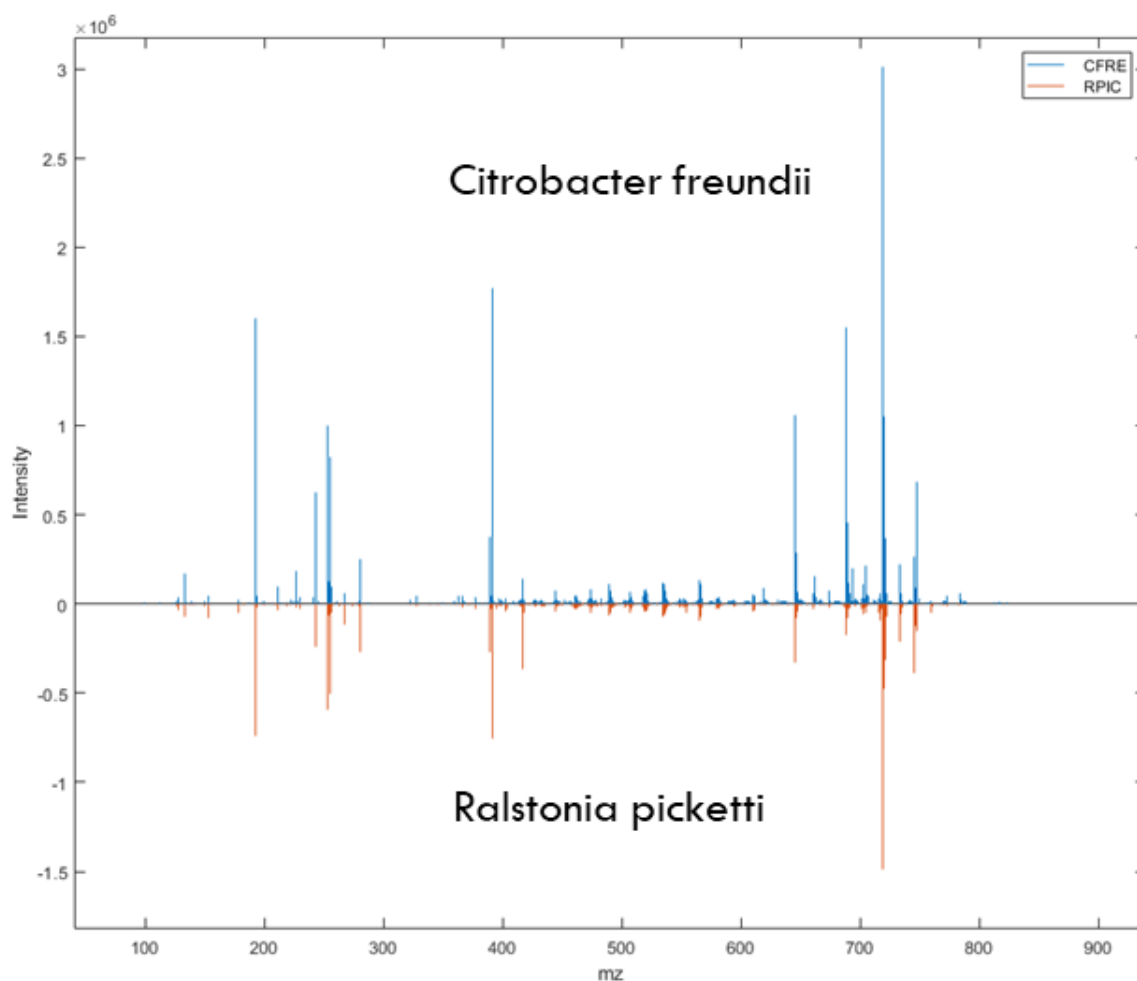


Figure 6.2. Representative mass spectra of the *Citrobacter freundii* and *Ralstonia picketti* water pollutants.

6.2.2.2 Classification model building for water contaminant identification

Here, the classification model was built by employing pre- and post-processing parameters that were determined to be most optimal in *Chapter 4*.

The PCA scores plot of the first two components, shown here as Figure 6.3, reveals clear separation between the various classes to which the species have been assigned, as well as grouping between the scans representing each water contaminant selected for this analysis.

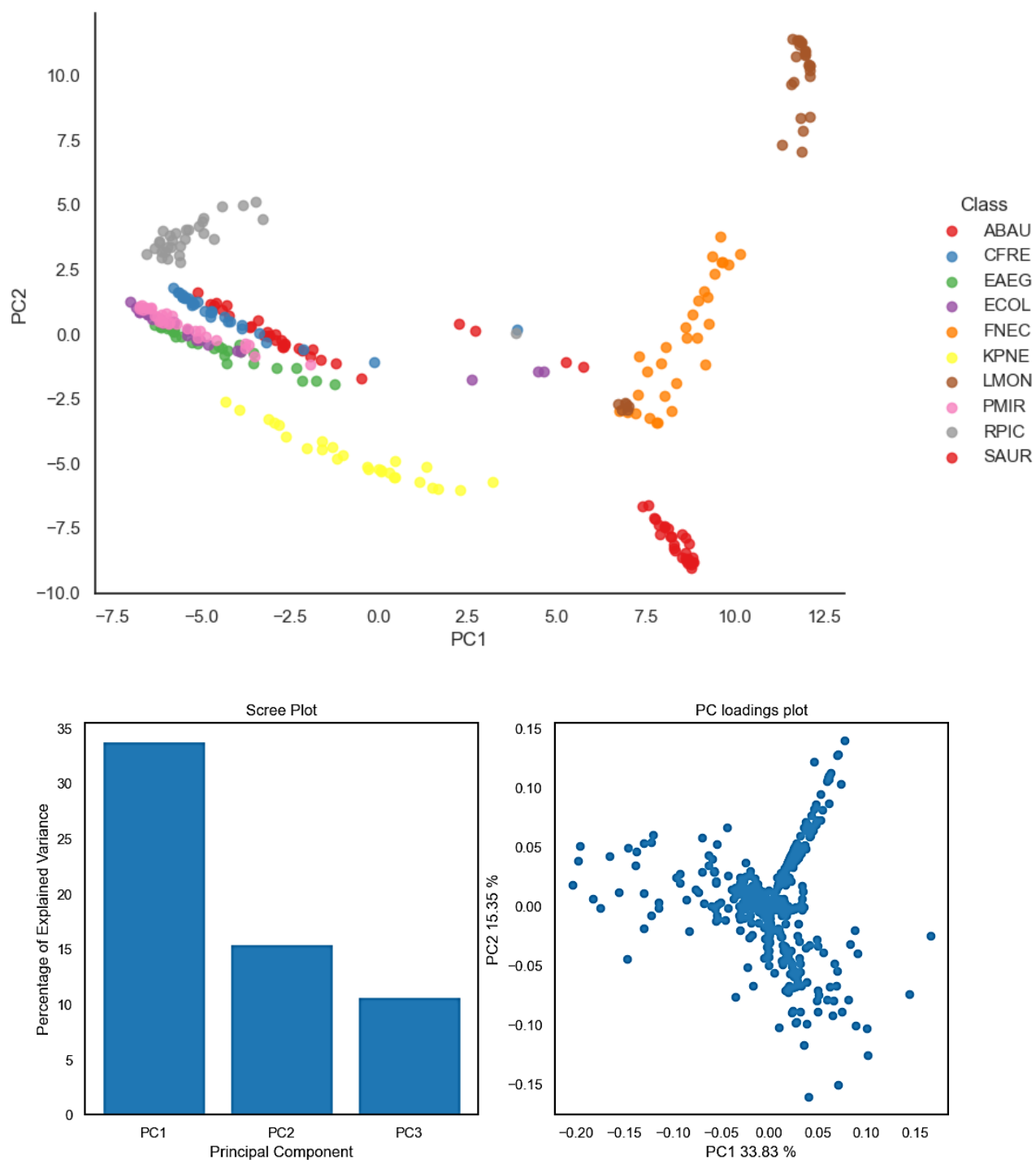


Figure 6.3 PCA scores plot of PC1 vs PC2 of the top 10 industrial pharmaceutical water contaminants analysed with LA-REIMS platform directly from filter membrane on the agar plate.

Here, an SVM-based training model incorporating 10 out of 15 biological replicates as well as 296/446 technical replicates for each of 10 water pathogens was built using cross validation with a leave one sample out (loo) approach. To validate the training model, five remaining biological replicates with three technical replicates each (150 total data points) were used as an independent validation set.

The confusion matrix, which is displayed in Figure 6.4, demonstrates high classification accuracy for the water pollutants that were examined using loo cross-validation. The achieved accuracy of the classification models was 99.34%.

The F1 classification accuracy that was obtained for the independent validation set was 97.97% (Figure 6.5), demonstrating that the training model had excellent performance and the capacity to identify unknown water pollutants with a high level of accuracy.

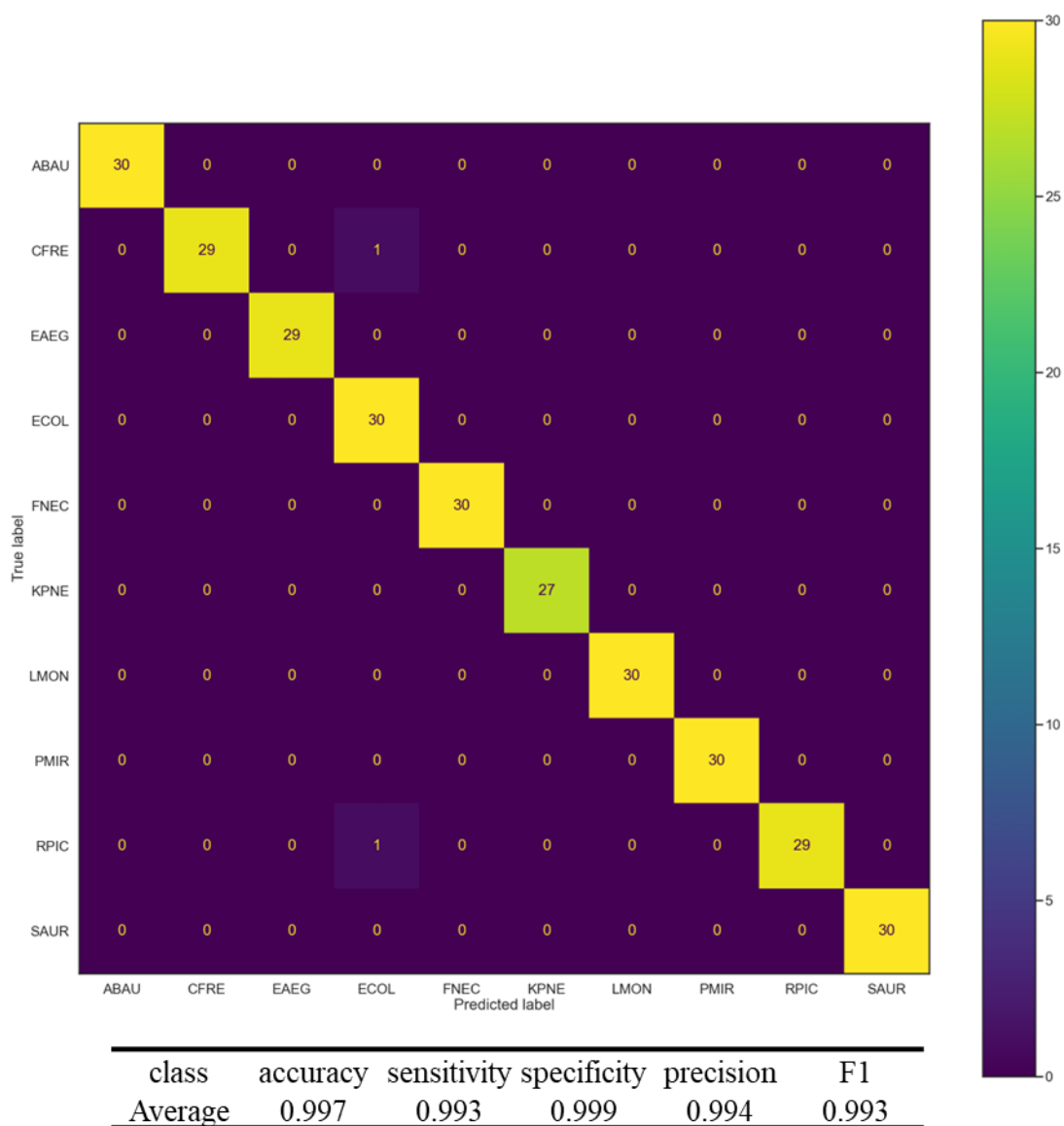


Figure 6.4 Confusion matrix and the average performance of the training model based on SVM classification model build using top 10 industrial pharmaceutical water contaminants with an overall species-level accuracy of F1 99.34 %.

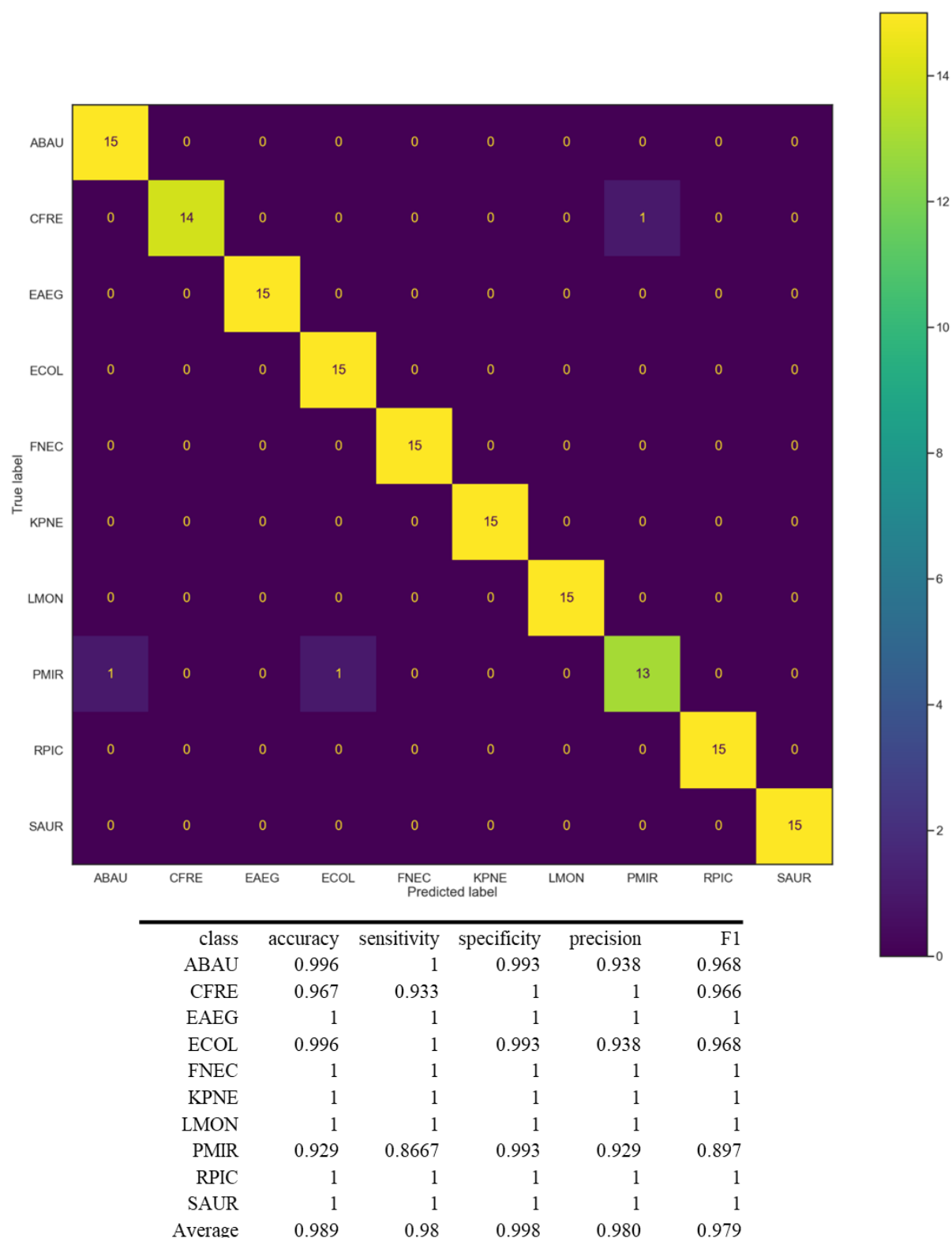


Figure 6.5 Confusion matrix and the performance of the independent validation set based on SVM classification model build using top 10 industrial pharmaceutical water contaminants with an overall species-level accuracy of 97.97 %.

As a means of additional verification, to confirm the robustness of the independent validation, a ROC analysis (please refer to *Chapter 3*), which is a method for determining how well a model based on machine learning can differentiate between a number of distinct species, was carried out in order to evaluate the accuracy of a predictive classification model build above. Here, the macro-averaged plot of ROC for AUC (Area Under the Curve) for all 10 water contaminants are shown in Figure 6.6 below. With an area under the curve (AUC) of 0.99, ROC analysis demonstrates that the classifier model has a high measure of species and classes contained in the model separability. In the ROC plot shows the True Positive Rate (TPR) against the False Positive Rate (FPR) which measures the proportion of true positive predictions correctly identified, while the FPR represents the proportion of false positive predictions. With TRE rate being close to 1 and FPR rate close to 0, the model shows high capability of accurately discriminating between classes.

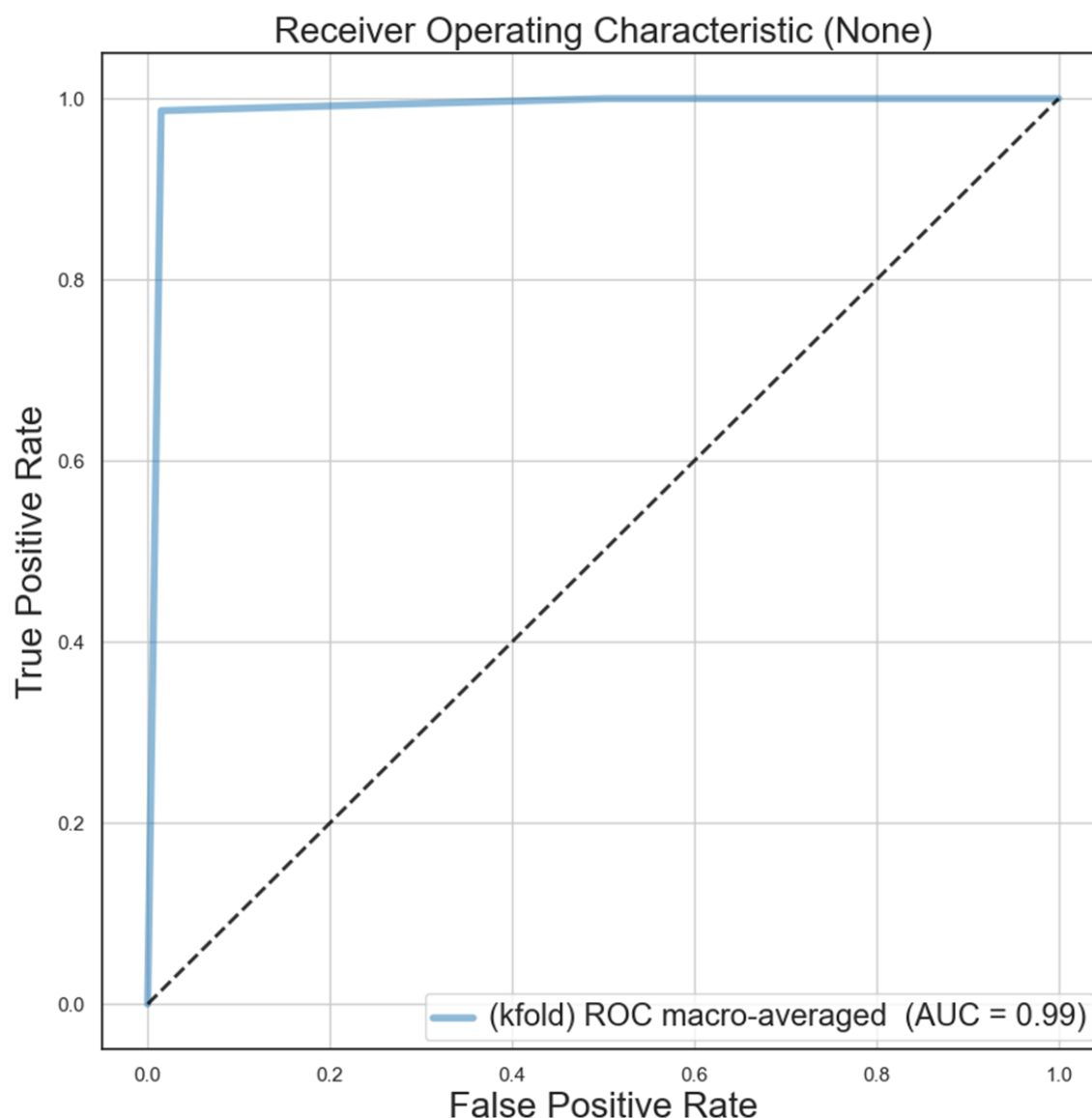


Figure 6.6 Macro-averaged plot of ROC for AUC for top 10 industrial pharmaceutical water pollutants classification (independent validation set) with AUC = 0.99.

6.3 Absence or Presence of specified organisms

6.3.1 Introduction

Microbiological examination of some non-sterile products requires tests for specified organisms. The tests allow determination of the absence of, or limited occurrence of, specified microorganisms that may be detected in the presence of various drug substances or excipients^{17,31}. The process includes the following three steps: first, preparation of the sample and pre- incubation to enrich the sample; second, selection and subculture utilising selective media in order to enrich the specified organism; and third, interpretation of the results to identify

them by following the same steps as an environmental monitoring test (purity plate, secondary incubation, Gram stain). The LA-REIMS platform on the other hand, can be used to directly identify these specified organisms from the final selective media subculturing phase in the presence of drug material.

6.3.1.1 Current industrial pharmaceutical specified microorganisms' detection pipeline

The absence or presence of a specified microorganism is currently determined via a two-way method. The most common process begins with suspending the medicine or excipient of choice into the media. Next, depending on if the method performed for a first time, to demonstrate method suitability, the appropriate microbial spike is added, followed by incubation to allow growth; otherwise, if the routine test is being performed, the suspension is incubated as it is, to allow for any growth that might be present in the test material. Following that, a small amount of liquid culture containing microorganisms and drug, or excipient is put into the appropriate selective media for enrichment, followed by incubation once again. Enrichment culture is then subcultured onto a selective media agar plate, incubated, and subcultured again onto a purity plate for any growth that might occur. Following that, Gram staining is performed on any observed growth, and microbiological contaminants are examined using validated identification system for example either a VITEK or a MicroSEQ, which requires further sample preparation prior to analysis (Figure 6.7).

The second method, which is used much less frequently, involves performing a filtration technique after the selective media enrichment has been incubated (Figure 6.7). As the filtration phase is subsequently followed by the microbiological growth on the membrane filter on the agar, and as mentioned in the previous paragraph for water analysis, LA-REIMS analysis was already researched for the possibility of analysing directly from the filter membrane, the subculturing after enrichment step procedure was investigated in this particular instance.

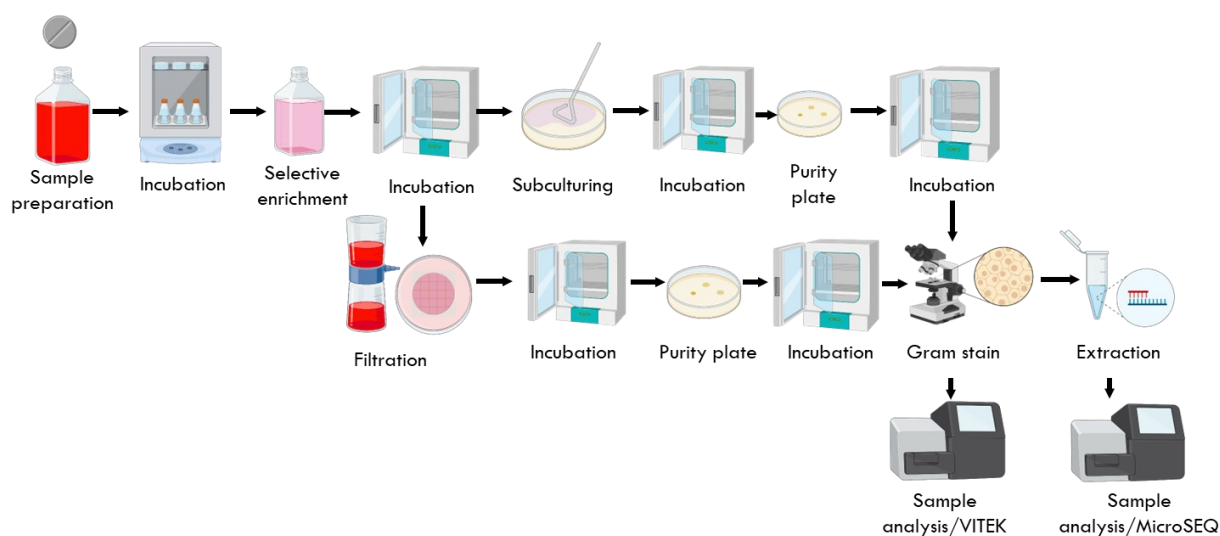


Figure 6.7 Current industrial pharmaceutical absence or presence of specified organisms' workflow. Current timeline 9-18 days depending on the species.

6.2.1.2 Potential LA-REIMS pipeline for detection of specified microorganisms

As shown in the Figure 6.8 below, in comparison to current industrial pharmaceutical pipeline, the suggested new pipeline using LA-REIMS would allow selected microorganisms detection in the presence of drug or excipient to be analysed straight after subculturing on selective media agar plate stage (Figure 6.7). The purity plate and gram stain phases would be completely eliminated by this new proposed approach, allowing for a significantly quicker turnaround for identifying and conforming these specified microorganisms.

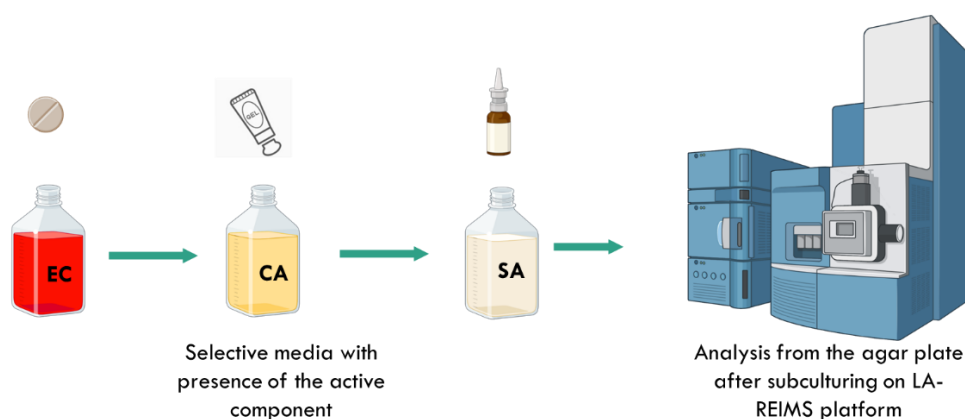


Figure 6.8 Proposed approach for detection of specified microorganisms using LA-REIMS (proposed timeline- 5 days). *E.coli* (EC), *Candida albicans* (CA) and *Staphylococcus aureus* (SA).

6.3.2 Results and Discussion

6.3.2.1 Data collection from the microorganisms in the presence of drug substance

To evaluate the effectiveness of the LA-REIMS platform in collecting raw data from microbial cultures in the presence of drug substances, three distinct types of drug compounds were chosen. As according USP Pharmacopoeia (USP 62 and USP 1111) guidelines²⁶³, *E.coli* was tested in the presence of a tablet (Panadol), *Candida albicans* in the presence of an ointment (Vagisil gel), and *Staphylococcus aureus* in the presence of a nose spray (Otrivine). Cultures were prepared in accordance with USP Pharmacopoeia requirements, including an enrichment stage following selective media enrichment, followed by selective media subculturing and incubation (*Chapter 3*, section 3.4.3). The combination of microbes, drug, enrichment medium, and selective agar used in the experiment is shown in Table 6.1 media selection aligned with the requirements of USP 62²⁹.

Microbial contaminants in the presence of the three chosen drugs were assessed on the LA-REIMS platform using 5 biological replicates and 5 technical replicates for each isolate. Quality control (QC) of the same three microbial cultures without the presence of drugs was also generated using the same sample preparation and analysis method.

Table 6.1 The combination of microbes, drug, enrichment medium, and selective agar used in the experiment.

Microorganism	Drug	Enrichment media	Selective agar
<i>E.coli</i>	Tablet	<i>MacConkey Broth</i>	<i>MacConkey Agar</i>
<i>Candida albicans</i>	Ointment	<i>Sabouraud Dextrose Broth</i>	<i>Sabouraud Dextrose Agar</i>
<i>Staphylococcus aureus</i>	Spray	<i>Soybean–Casein Digest Broth</i>	<i>Mannitol Salt Agar</i>

The PCA score plot in Figure 6.9 indicated distinct clusters of the three separate microorganisms grown in the presence of the drug, together with the QC plates that did not contain any pharmaceuticals in the preparation. This means that data acquired from specific microorganisms in the presence of the drug substance using LA-REIMS revealed no quantifiable difference between species when the drug was not present.

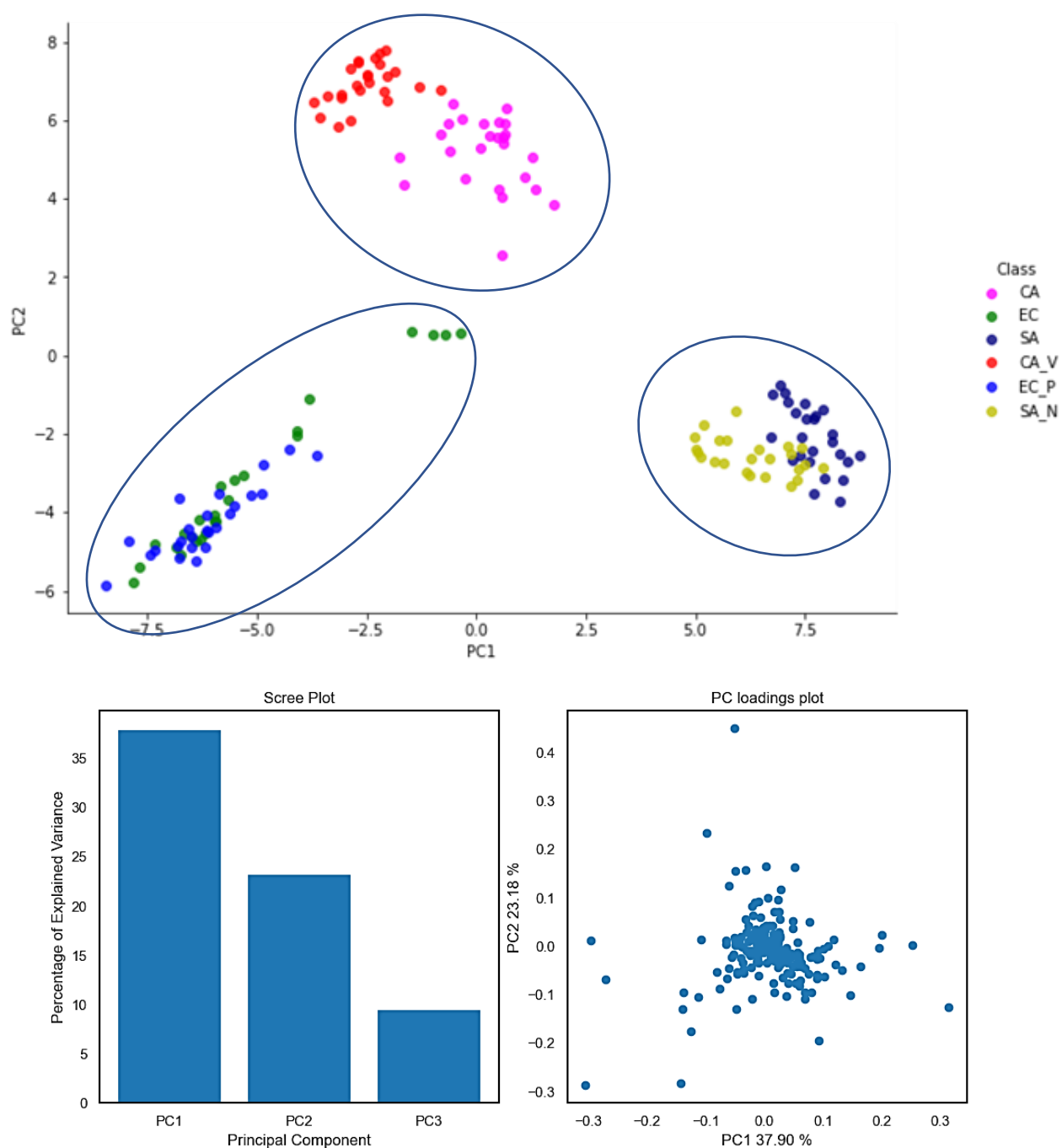


Figure 6.9 PCA score plot of PC1 vs PC2 of data collected from *E.coli*, *Candida albicans* and *Staphylococcus aureus* cultured with and without the drug substance present.

The LA-REIMS spectra of *E.coli* with and without the drug present were examined in order to further confirm no differences between the two samples prepared with and without the drug prior to the LA-REIMS analysis. In general, the ion intensities in both samples were found to be similar. Furthermore, no significant features in spectra from the drug presence in one of the samples were noticed. This observation is consistent with the findings of PCA score plot clusters above.

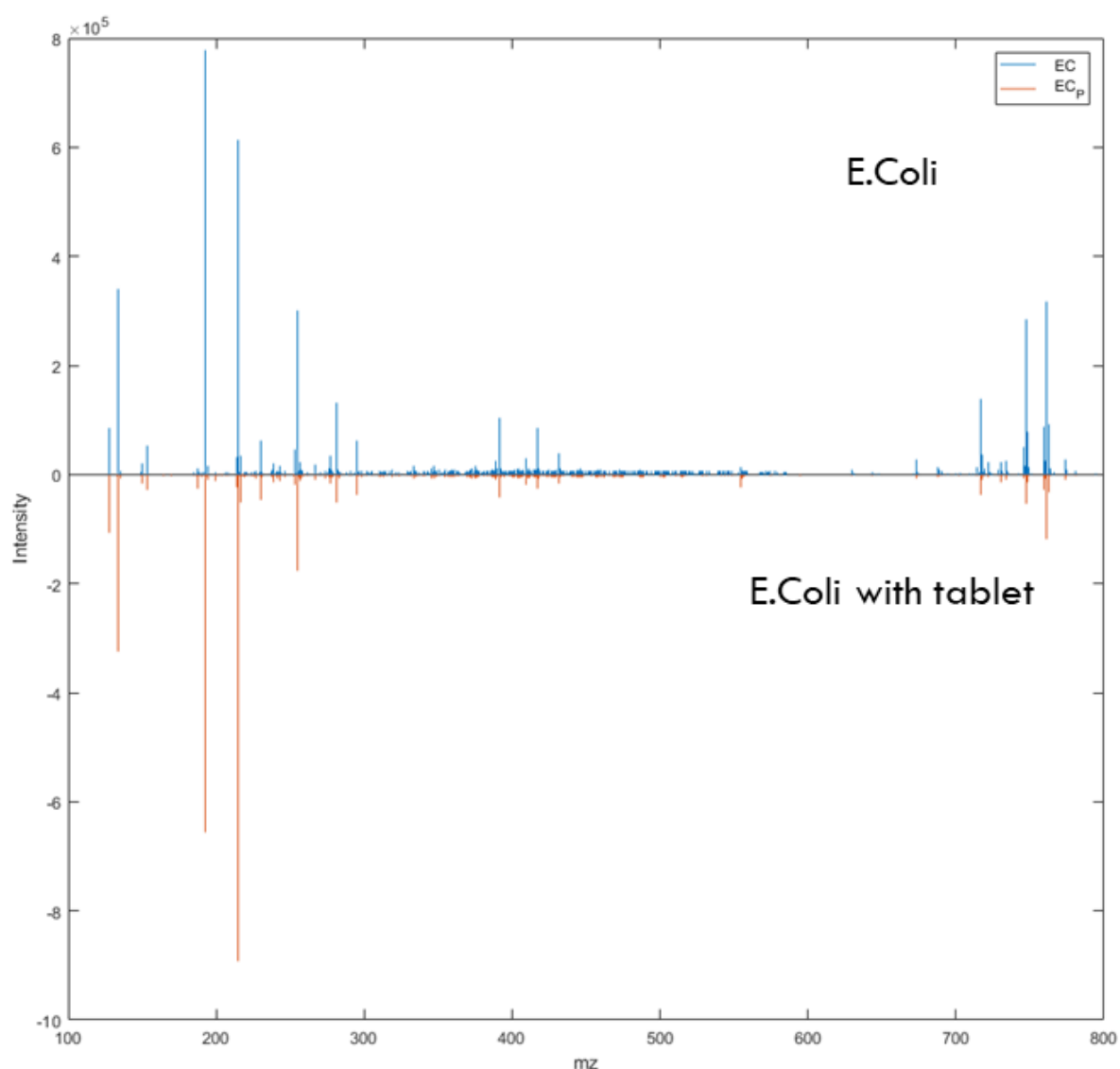


Figure 6.10 Representative mass spectra of *E.coli* with and without the drug presence during sample preparation prior to the LA-REIMS analysis.

6.3.2.2 Classification model building for microorganisms in the presence of drug substance

Here, the classification model was built by employing pre-processing parameters that were determined to be most optimal in Chapter 4.

An SVM-based classification model was developed by employing cross validation with a leave one sample out (loo) approach. Please refer to *Chapter 3* for details on model evaluation selection.

In the first place, in response to the observation made earlier about the absence of a difference in the data collected from microorganisms with and without the drug, a classification model containing all of the isolates (both QC and samples containing the drug) was developed. The purpose of the experiment was to test the model in order to determine whether the classes that had been assigned would still be correctly classified or whether they would be incorrectly assigned to the opposite class that contained samples with the drug present (for example, a sample that contained drug would be misclassified as a sample that did not contain drug).

Here, the confusion matrix in the Figure 6.11, demonstrates classification accuracy after loo cross-validation of the microbial species that were grown in the presence of the drug as well as no drug. The aim was to show that LA-REIMS technique can identify the micro-organisms correctly with or without the presence of the drug. The F1 score was observed at 96.50%, where some of the samples were incorrectly classified, which decreased the accuracy of the classification model. Sometimes, *E. coli* without the drug present was labelled as *E. coli* with drug present, and vice versa. The obtained results supported the aim that was set to achieve.

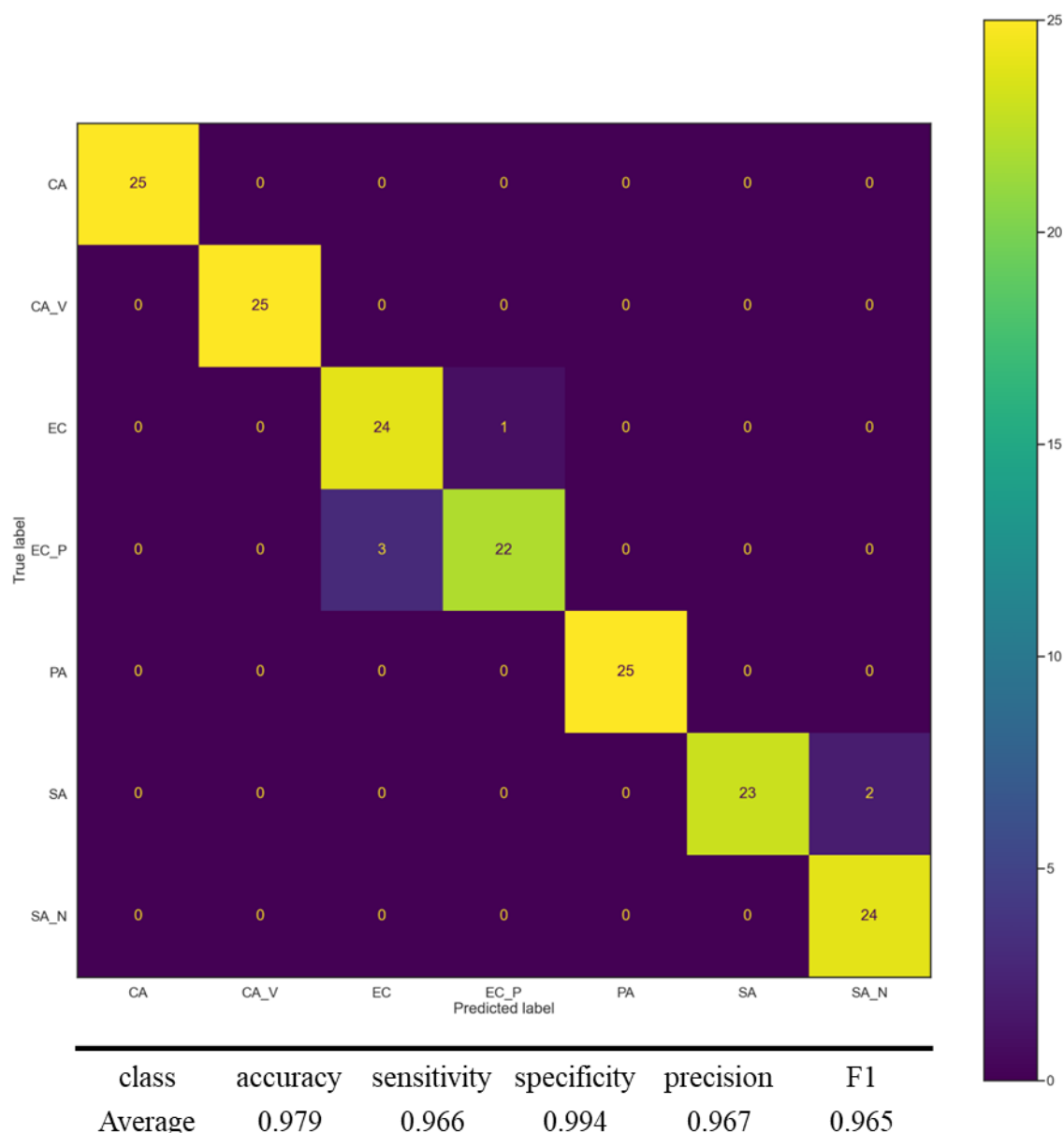


Figure 6.11 Confusion matrix and average performance based on the SVM classification model validated using loo, built using the specified organisms with and without the presence of the drug substance.

The confusion matrix was also generated for the classification models built only using the species with the drug present, in order to confirm the precise accuracy for the classification as well as to confirm the possibility of the predictive modelling-based LA-REIMS platform method for the identification of the specified species in the presence of the drug or excipient.

Here, the confusion matrix in Figure 6.12, shows the F1 score off 100% for the classification model built using SVM and leave one sample out cross validation.

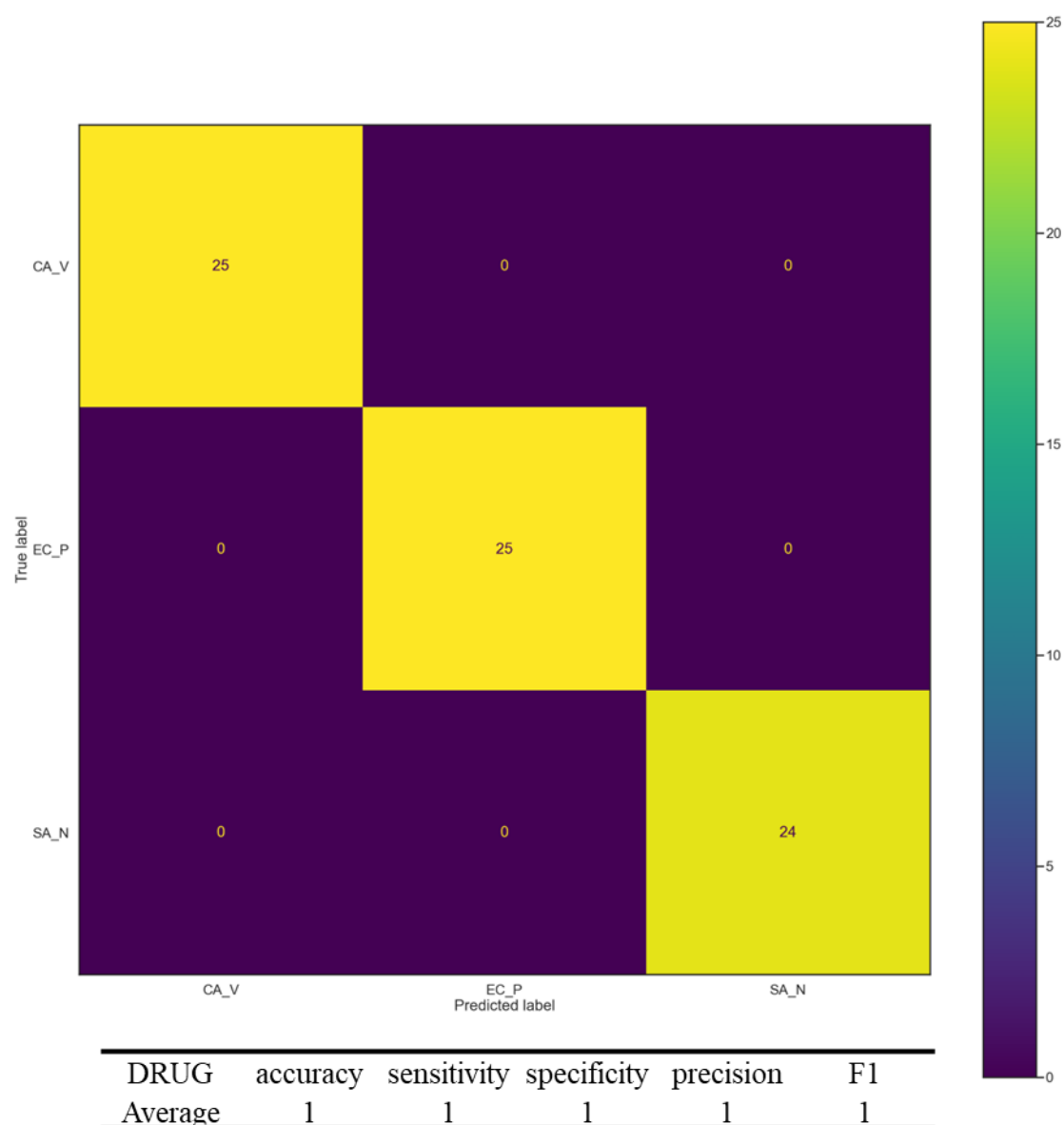


Figure 6.12 Confusion matrix and the average performance based on the SVM classification model built using the specified organisms cultured with the presence of the drug substances.

6.4 Conclusion

In this chapter, the aim was to evaluate and demonstrate the possibility of the LA-REIMS platform to be incorporated into the industrial pharmaceutical pipeline for microbial contaminant identification.

The main differences have been outlined that exist between the industrial pharmaceutical and the potential pipeline that could be developed using LA-REIMS in combination with machine learning, for water pollutant identification methods that are currently in use, as well as the methods for identifying specified organisms with the drug or excipients present. Furthermore, the main steps were then demonstrated that were taken to successfully evaluate and develop accurate predictive model for the classification of unknown water contaminants and the specified organisms.

First, in order to identify the water borne contaminants, method was developed that uses the LA-REIMS platform to acquire data directly from microorganisms growing on the filter membrane. Acquired data was then used to develop a predictive model that was able to identify unknown water borne contaminants with an F1 score accuracy of 97.97% of the independent validation set.

Secondly, work performed for the specified organisms in the presence of drug or excipient confirmed that the presence of the drug substance in the preparative steps of the samples prior to the data acquisition using LA-REIMS gave no distinctive signal or features in comparison to species that were grown in normal conditions without the drug. Obtained results for this part of the chapter showed that LA-REIMS technique can identify the microorganisms correctly with or without the presence of the drug as well as demonstrated that predictive modelling approach can be successfully used to confirm the presence of these microorganisms in the drug testing cultures, as well as identify them with the accuracy of F1 100% after loo cross validation.

Chapter 7

Conclusions, Strengths and Limitations, Future Work

Chapter 7 Conclusion, Strengths and Limitations, Future Work

7.1 Conclusion

The main objective of this research project was to develop and validate an automated LA-REIMS platform, which would be used in conjunction with machine learning multivariate model building, for the direct analysis/identification of microbial samples. This project focused on exploring three key microbial applications in the industrial pharmaceutical sector. These applications include environmental monitoring, detection of water pollutants, and identification of microbial presence in drugs or excipients.

The objective shared by all projects was to establish a stable, repeatable, and reproducible multivariate classification model capable of accurately identifying unknown microbial species with high precision.

The most accurate parameters for the in-house developed data pre-processing pipeline for environmental monitoring of bacteria and yeast were determined through rigorous model evaluation. The selected parameters included 'Pre-processing 2' with '75% peak filter' and with 'TIC+Log' normalisation. The final results showcased a large-scale multivariate classification model, encompassing 162 microbial classes representing 211 species. This model achieved an average accuracy of F1 score of 98.1% through K-fold cross-validation. The generalisability was also validated, as perfect average classification accuracy of F1 score of 100% for an independent validation set was also achieved. Additionally, the utilization of RFE feature selection identified 75 statistically significant biological features. These features were employed to validate the entire data matrix model and explore the potential for biomarker-specific multivariate modelling. The RFE feature training model achieved an average accuracy of F1 score of 92.8% and correctly identified 46 species in the independent validation set with a F1 score of over 90%, highlighting the potential for robust, fast predictive modelling by incorporating a handful of biomarkers. Furthermore, the model also demonstrated resilience to sample storage in the refrigerator for up to 72 hours and exhibited good repeatability across multiple days of data collection and analysis. A reproducibility study, conducted with a three-

year time gap and different instruments, also achieved an average classification accuracy of over 90% for unknown species.

This project also addressed the challenge of accurately identifying filamentous fungi caused by metabolic changes resulting from culture age, which is faced by other applications and techniques (e.g., MALDI, PCR). An independent validation process was conducted to address this issue. By utilising species that had undergone multiple subculturing over six months, the LA-REIMS platform achieved a classification accuracy of 92% and an F1 score accuracy of 87% on the independent validation set, indicating substantial potential for the technique.

For waterborne contaminants, a novel method was developed using the LA-REIMS platform to directly acquire data from microorganisms growing on filter membranes. This data was utilised to construct a predictive model capable of identifying unknown waterborne contaminants with an impressive F1 score accuracy of 97.97%.

Regarding specified organisms in the presence of drugs or excipients, experiments confirmed that the presence of drug substances during the sample preparation steps prior to LA-REIMS data acquisition did not generate distinct signals or features compared to organisms grown without the drug. The results indicated that the LA-REIMS technique can accurately identify microorganisms in the presence of drugs. The predictive modelling approach successfully confirmed the presence of these microorganisms in drug testing cultures, achieving a remarkable classification accuracy of 100%.

In conclusion, the LA-REIMS platform, when combined with the optimised data pre-processing, exhibits significant potential for integration into the industrial pharmaceutical environmental monitoring pipeline. This technology offers faster and more accurate identification of unknown microbial contaminants, resulting in substantial cost savings for companies. The streamlined process eliminates the need for extensive sample preparation, saving both time and resources. Accurate identification plays a crucial role in directing investigations and initiating prompt corrective actions. Overall, the LA-REIMS platform offers a promising solution for efficient and cost-effective microbial identification in pharmaceutical industries.

Overall, LA-REIMS could provide industrial pharmaceutical facilities with a powerful tool for rapid, accurate, and efficient microbial detection and identification. Its capabilities support effective quality control and contamination management, ensuring the production of safe and high-quality pharmaceutical products.

7.2 Strengths and Limitations

LA-REIMS is a novel analytical technique that offers a swift and non-invasive method for microbial analysis. Unlike conventional methods, LA-REIMS does not necessitate prolonged culturing or isolation of microorganisms. When compared to other analytical techniques, LA-REIMS stands out as a promising alternative. The technology allows for the analysis of microbial samples without the need for sample pre-processing, resulting in a reduction of time and effort required. The LA-REIMS method is not dependent on particular DNA sequences and can accurately examine a broad spectrum of microorganisms instantaneously, obviating the necessity for DNA extraction and amplification, thereby resulting in a more rapid and efficient process.

Although LA-REIMS is a promising novel technique, it also has some limitations that should be considered. In its current state, LA-REIMS may not always provide species-level identification for certain microorganisms. While it can differentiate broad groups or classes of microorganisms, it may struggle to discriminate closely related species or strains. The accuracy of LA-REIMS heavily relies on the quality and comprehensiveness of the reference spectral databases used for comparison. Limited or incomplete databases may result in misidentification or inability to identify certain microorganisms.

Despite these limitations, LA-REIMS continues to show promise as a rapid and non-destructive method for microbial identification. Ongoing advancements and improvements in technology, database development, and standardization efforts are expected to address some of these limitations and enhance the utility of LA-REIMS in the future.

LA-REIMS shows promise as a scalable and adaptable technology for industrial pharmaceutical applications. Its compatibility with existing infrastructure, potential for high-volume testing, cost-effectiveness, regulatory acceptance, and user-friendliness make it a compelling option for the industry's evolving needs.

7.3 Future Work

This research projects results have demonstrated the promising potential POC of the LA-REIMS setup where achieved results were user-independent and stable over a range of environmental factors without compromising the high detection accuracy.

Funding from BBSRC IAA and BBSRC ICURe has been secured to support ongoing research and development efforts aimed at enhancing the reliability and effectiveness of the existing LA-REIMS POC. The ultimate goal is to transform the POC into a prototype that can be easily implemented across a range of industrial environments, both within and beyond the laboratory. The targeted applications comprise environmental monitoring, bioburden estimation, non-sterile product microbiological examination, and antimicrobial effectiveness testing. Specifically, further streamlining to the prototype platform is aimed to be tackled with industrial collaborators at AstraZeneca. Future experiments aimed at transforming the current POC into a more portable, fully automated prototype that is independent of laser wavelength are in the planning stage. The objective is to test the prototype with a variety of pharmaceutical samples to determine its relevance.

Specifically, the focus will be on expanding and improving the reference spectral databases for various microorganisms utilised in LA-REIMS. The process involves the integration of a wider range of microbial species and strains, along with the enhancement of database quality and coverage, as well as time series data from various stages of subculturing. The accuracy and reliability of microbial identification through LA-REIMS can be improved by utilising comprehensive and current databases. Additionally, further research is necessary to investigate the wider range of drugs and excipients for the purpose of testing the absence or presence of specific organisms. This may involve the development of a spectral library of microorganisms that have been grown in the presence of various substances.

By addressing these aspects, future work can contribute to the advancement and wider adoption of LA-REIMS as a powerful tool for microbial identification not only for industrial pharmaceutical sector, but also providing valuable insights and applications in various fields of research and industry (e.g., food, healthcare, cosmetics).

Appendix

Table 4.10 List of 162 microorganisms with the sample quantities used for large-scale environmental model development.

	Microbial genera	Class	Replicate Count
1	<i>Acinetobacter baumannii</i>	AcineBauma	15
2	<i>Acinetobacter lwoffii</i>	AcineIwoff	76
3	<i>Acinetobacter radioresistens</i>	AcineRadio	15
4	<i>Acinetobacter schindleri</i>	AcineSchin	13
5	<i>Acinetobacter townneri</i>	AcineTowne	9
6	<i>Acinetobacter ursingii</i>	AcineUrsin	15
7	<i>Aerococcus urinaeequi</i>	AerocUrina	6
8	<i>Aerococcus viridans</i>	AerocVirid	9
9	<i>Agrococcus jenensis</i>	AgrocJenen	9
10	<i>Arthrobacter luteolus</i>	ArthrLuteo	9
11	<i>Bacillus altitudinis</i>	BacilAltit	38
12	<i>Bacillus amyloliquefaciens</i>	BacilAmylo	60
13	<i>Bacillus atrophaeus</i>	BacilAtrop	15
14	<i>Bacillus cereus</i>	BacilCereu	29
15	<i>Bacillus circulans</i>	BacilCircu	42
16	<i>Bacillus flexus</i>	BacilFlexu	23
17	<i>Bacillus halosaccharovorans</i>	BacilHalos	13
18	<i>Bacillus infantis</i>	BacilInfan	32
19	<i>Bacillus licheniformis</i>	BacilLiche	77
20	<i>Bacillus massiliosenegalensis</i>	BacilMassi	9
21	<i>Bacillus megaterium</i>	BacilMegat	36
22	<i>Bacillus mycoides</i>	BacilMycoi	22
23	<i>Bacillus oceanisediminis</i>	BacilOcean	26
24	<i>Bacillus paramycoides</i>	BacilParam	15
25	<i>Bacillus pumilus</i>	BacilPumil	11
26	<i>Bacillus safensis</i>	BacilSafen	10
27	<i>Bacillus species</i>	BacilSP	99
28	<i>Bacillus subtilis</i>	BacilSubti	18
29	<i>Bacillus tequilensis</i>	BacilTequi	15
30	<i>Bacillus thuringiensis</i>	BacilThuri	12
31	<i>Bhargavaea changchunensis</i>	BhargChang	15
32	<i>Brevibacillus aydinogluensis</i>	BrachSP	12

33	<i>Brevibacillus parabrevis</i>	BreviAydin	9
34	<i>Brevibacterium casei</i>	BreviCasei	14
35	<i>Brevibacterium celere</i>	BreviCeler	38
36	<i>Brevibacterium frigoritolerans</i>	BreviFrigo	107
37	<i>Brevibacterium iodinum</i>	BreviIodin	9
38	<i>Brevibacillus parabrevis</i>	BreviParab	9
39	<i>Brevibacterium pityocampae</i>	BreviPityo	9
40	<i>Candida albicans</i>	CandiAlbic	28
41	<i>Candida guilliermondii</i>	CandiGuill	31
42	<i>Candida parapsilosis</i>	CandiParap	29
43	<i>Citricoccus zhacaiensis</i>	CitriZhaca	9
44	<i>Citrobacter freundii</i>	CitroFreun	9
45	<i>Corynebacterium accolens</i>	CorynAccol	9
46	<i>Corynebacterium afermentans</i>	CorynAferm	10
47	<i>Corynebacterium aurimucosum</i>	CorynAurim	14
48	<i>Corynebacterium coyleae</i>	CorynCoyle	16
49	<i>Corynebacterium propinquum</i>	CorynPropi	18
50	<i>Corynebacterium sanguinis</i>	CorynSangu	15
51	<i>Corynebacterium tuberculostearicum</i>	CorynTuber	27
52	<i>Corynebacterium xerosis</i>	CorynXeros	33
53	<i>DermaCoccus nishinomiyaensis</i>	DermaNishi	16
54	<i>Dietzia papillomatosis</i>	DietzPapil	21
55	<i>Enterobacter aerogenes</i>	EnterAerog	9
56	<i>Enterobacter kobei</i>	EnterFaeca	17
57	<i>Enterococcus faecalis</i>	EnterKobei	11
58	<i>Escherichia coli</i>	EscheColi	9
59	<i>Escherichia fergusonii</i>	EscheFergu	19
60	<i>Exiguobacterium species</i>	ExiguSP	6
61	<i>Exophiala dermatitidis</i>	ExophDerma	23
62	<i>Exophiala species</i>	ExophSP	13
63	<i>Fictibacillus barbaricus</i>	FictiBarba	11
64	<i>Fictibacillus phosphorivorans</i>	FictiPhosp	15
65	<i>Fusobacterium necrophorum</i>	FusobNecro	9
66	<i>Georgenia muralis</i>	GeorgMural	9
67	<i>Gordonia bronchialis</i>	GordoBronc	9
68	<i>Gordonia species</i>	GordoSP	15
69	<i>Gordonia sputi</i>	GordoSputi	20
70	<i>Janibacter species</i>	JanibSP	12
71	<i>Klebsiella pneumoniae</i>	KlebsPneum	9
72	<i>Kocuria arsenatis</i>	KocurArsen	12
73	<i>Kocuria palustris</i>	KocurPalus	40
74	<i>Kocuria rhizophila</i>	KocurRhizo	103
75	<i>Kocuria species</i>	KocurSP	12
76	<i>Lactococcus lactis</i>	LactoLacti	7

77	<i>Leifsonia soli</i>	LeifsSoli	55
78	<i>Listeria monocytogenes</i>	ListeMonoc	9
79	<i>Lysinibacillus composti</i>	LysinCompo	9
80	<i>Lysinibacillus fusiformis</i>	LysinFusif	12
81	<i>Lysinibacillus odyseyi</i>	LysinOdyss	9
82	<i>Metabacillus species</i>	MesobSP	10
83	<i>Metabacillus halosaccharovorans</i>	MetabHalos	34
84	<i>Micrococcus flavus</i>	MicroFlavu	12
85	<i>Microbacterium lacticum</i>	MicroLacti	10
86	<i>Micrococcus luteus</i>	MicroLuteu	100
87	<i>Micrococcus lylae</i>	MicroLylae	21
88	<i>Microbacterium paraoxydans</i>	MicroParao	17
89	<i>Microbacterium proteolyticum</i>	MicroProte	67
90	<i>Microbacterium species</i>	MicroSP	12
91	<i>Micrococcus terreus</i>	MicroTerre	26
92	<i>Moraxella osloensis</i>	MoraxOsloe	98
93	<i>Niallia circulans</i>	NiallCircu	29
94	<i>Niallia species</i>	NiallSP	9
95	<i>OceanoBacillus chironomi</i>	OceanChiro	9
96	<i>Oceanobacillus kimchii</i>	OceanKimch	9
97	<i>OrnithiniBacillus bavariensis</i>	OrnitBavar	12
98	<i>Paenibacillus amylolyticus</i>	PaeniAmylo	17
99	<i>Paenibacillus campinasensis</i>	PaeniCampi	10
100	<i>PaeniBacillus glucanolyticus</i>	PaeniGluca	45
101	<i>PaeniBacillus lautus</i>	PaeniLautu	12
102	<i>PaeniBacillus provencensis</i>	PaeniProve	24
103	<i>Paenibacillus segetis</i>	PaeniSeget	9
104	<i>Paenibacillus species</i>	PaeniSP	42
105	<i>Paenibacillus taichungensis</i>	PaeniTaich	57
106	<i>Pantoea septica</i>	PantoSepti	17
107	<i>Pantoea species</i>	PantoSP	17
108	<i>Paracoccus yeei</i>	ParacYeei	31
109	<i>Peribacillus butanolivorans</i>	PeribButan	12
110	<i>Peribacillus psychrosaccharolyticus</i>	PeribPsych	9
111	<i>Planococcus okeanokoites</i>	PlanoOkean	9
112	<i>Prolinoborus fasciculus</i>	ProliFasci	51
113	<i>Proteus mirabilis</i>	ProteMirab	9
114	<i>Pseudomonas aeruginosa</i>	PseudAerug	18
115	<i>Pseudomonas azotoformans</i>	PseudAzoto	12
116	<i>Pseudomonas chlororaphis</i>	PseudChlor	9
117	<i>Pseudomonas species</i>	PseudSP	18
118	<i>Pseudomonas stutzeri</i>	PseudStutz	21
119	<i>Ralstonia pickettii</i>	RalstPicke	65
120	<i>Raoultella planticola</i>	RaoulPlant	11

121	<i>Rhizobium radiobacter</i>	RhizoRadio	9
122	<i>Rhodotorula mucilaginosa</i>	RhodoMucil	64
123	<i>Roseomonas mucosa</i>	RoseoMucos	31
124	<i>Rothia aeria</i>	RothiAeria	9
125	<i>Rothia dentocariosa</i>	RothiDento	7
126	<i>Rothia kristinae</i>	RothiKrist	9
127	<i>Rothia terrae</i>	RothiTerra	10
128	<i>Shigella flexneri</i>	ShigeFlexn	54
129	<i>Sphingomonas olei</i>	ShingOlei	12
130	<i>Sphingomonas carotinifaciens</i>	SphinCarot	10
131	<i>Sphingomonas melonis</i>	SphinMelon	19
132	<i>Sphingobacterium multivorum</i>	SphinMulti	12
133	<i>Sphingomonas panaciterrae</i>	SphinPanac	12
134	<i>Sphingomonas species</i>	SphinSP	12
135	<i>Sphingopyxis terrae</i>	SphinTerra	11
136	<i>Staphylococcus argenteus</i>	StaphArgen	39
137	<i>Staphylococcus arlettae</i>	StaphArlet	13
138	<i>Staphylococcus aureus</i>	StaphAureu	69
139	<i>Staphylococcus auricularis</i>	StaphAuric	15
140	<i>Staphylococcus capitis</i>	StaphCapit	114
141	<i>Staphylococcus caprae</i>	StaphCapra	66
142	<i>Staphylococcus cohnii</i>	StaphCohni	84
143	<i>Staphylococcus condimenti</i>	StaphCondi	9
144	<i>Staphylococcus edaphicus</i>	StaphEdaph	41
145	<i>Staphylococcus epidermidis</i>	StaphEpide	110
146	<i>Staphylococcus haemolyticus</i>	StaphHaemo	109
147	<i>Staphylococcus hominis</i>	StaphHomin	115
148	<i>Staphylococcus lentus</i>	StaphLentu	9
149	<i>Staphylococcus lugdunensis</i>	StaphLugdu	94
150	<i>Staphylococcus pasteurii</i>	StaphPaste	12
151	<i>Staphylococcus petrasii</i>	StaphPetra	30
152	<i>Staphylococcus pettenkoferi</i>	StaphPette	19
153	<i>Staphylococcus saprophyticus</i>	StaphSapro	78
154	<i>Staphylococcus simulans</i>	StaphSimul	9
155	<i>Staphylococcus succinus</i>	StaphSucci	14
156	<i>Staphylococcus warneri</i>	StaphWarne	111
157	<i>Stenotrophomonas panacihumi</i>	StenoPanac	9
158	<i>Streptococcus salivarius</i>	StrepSaliv	9
159	<i>Streptococcus sanguinis</i>	StrepSangu	9
160	<i>Streptococcus species</i>	StreptocSP	9
161	<i>Streptomyces species</i>	StreptomSP	9
162	<i>Xanthomonas species</i>	XanthSP	9
	Grand Total		4264

Table 4.11 List of 162 species performance in the classification reference model using whole data matrix. Model accuracy after K-fold cross validation - F1 score 98.1%.

class	accuracy	sensitivity	specificity	precision	F1
AcineBauma	0.967	0.933	1.000	1.000	0.966
AcineIwoff	0.971	0.942	1.000	0.975	0.959
AcineRadio	1.000	1.000	1.000	1.000	1.000
AcineSchin	1.000	1.000	1.000	1.000	1.000
AcineTowne	1.000	1.000	1.000	1.000	1.000
AcineUrsin	0.967	0.933	1.000	1.000	0.966
AerocUrina	1.000	1.000	1.000	1.000	1.000
AerocVirid	1.000	1.000	1.000	1.000	1.000
AgrocJenen	1.000	1.000	1.000	1.000	1.000
ArthrLuteo	1.000	1.000	1.000	1.000	1.000
BacilAltit	1.000	1.000	1.000	1.000	1.000
BacilAmylo	1.000	1.000	1.000	1.000	1.000
BacilAtrop	1.000	1.000	1.000	1.000	1.000
BacilCereu	0.885	0.769	1.000	1.000	0.870
BacilCircu	0.983	0.967	1.000	0.955	0.961
BacilFlexu	1.000	1.000	1.000	1.000	1.000
BacilHalos	1.000	1.000	1.000	1.000	1.000
BacilInfan	1.000	1.000	1.000	1.000	1.000
BacilLiche	0.975	0.949	1.000	1.000	0.974
BacilMassi	1.000	1.000	1.000	1.000	1.000
BacilMegat	1.000	1.000	0.999	0.923	0.960
BacilMycoi	0.999	1.000	0.998	0.751	0.858
BacilOcean	1.000	1.000	1.000	1.000	1.000
BacilParam	1.000	1.000	1.000	1.000	1.000
BacilPumil	1.000	1.000	1.000	1.000	1.000
BacilSP	1.000	1.000	1.000	1.000	1.000
BacilSafen	1.000	1.000	1.000	1.000	1.000
BacilSubti	1.000	1.000	1.000	1.000	1.000
BacilTequi	1.000	1.000	0.999	0.917	0.957
BacilThuri	0.917	0.833	1.000	1.000	0.909
BhargChang	1.000	1.000	1.000	1.000	1.000
BrachSP	1.000	1.000	1.000	1.000	1.000
BreviAydin	1.000	1.000	1.000	1.000	1.000
BreviCasei	0.893	0.786	1.000	1.000	0.880
BreviCeler	1.000	1.000	1.000	0.952	0.975
BreviFrigo	1.000	1.000	1.000	1.000	1.000
BreviIodin	1.000	1.000	1.000	1.000	1.000
BreviParab	1.000	1.000	1.000	1.000	1.000

BreviPityo	1.000	1.000	1.000	1.000	1.000
CandiAlbic	1.000	1.000	1.000	1.000	1.000
CandiGuill	1.000	1.000	1.000	1.000	1.000
CandiParap	1.000	1.000	1.000	1.000	1.000
CitriZhaca	1.000	1.000	1.000	1.000	1.000
CitroFreun	1.000	1.000	1.000	1.000	1.000
CorynAccol	1.000	1.000	1.000	1.000	1.000
CorynAferm	1.000	1.000	1.000	1.000	1.000
CorynAurim	1.000	1.000	1.000	1.000	1.000
CorynCoyle	1.000	1.000	1.000	1.000	1.000
CorynPropi	1.000	1.000	1.000	1.000	1.000
CorynSangu	1.000	1.000	1.000	1.000	1.000
CorynTuber	1.000	1.000	1.000	1.000	1.000
CorynXeros	1.000	1.000	1.000	1.000	1.000
DermaNishi	1.000	1.000	1.000	1.000	1.000
DietzPapil	1.000	1.000	1.000	1.000	1.000
EnterAerog	1.000	1.000	1.000	1.000	1.000
EnterFaeca	1.000	1.000	1.000	1.000	1.000
EnterKobei	1.000	1.000	1.000	1.000	1.000
EscheColi	1.000	1.000	1.000	1.000	1.000
EscheFergu	1.000	1.000	1.000	1.000	1.000
ExiguSP	1.000	1.000	1.000	1.000	1.000
ExophDerma	1.000	1.000	1.000	1.000	1.000
ExophSP	1.000	1.000	1.000	1.000	1.000
FictiBarba	0.955	0.909	1.000	1.000	0.952
FictiPhosp	1.000	1.000	1.000	1.000	1.000
FusobNecro	1.000	1.000	1.000	1.000	1.000
GeorgMural	1.000	1.000	1.000	1.000	1.000
GordoBronc	1.000	1.000	1.000	1.000	1.000
GordoSP	1.000	1.000	1.000	0.944	0.971
GordoSputi	1.000	1.000	1.000	1.000	1.000
JanibSP	1.000	1.000	1.000	1.000	1.000
KlebsPneum	1.000	1.000	1.000	1.000	1.000
KocurArsen	1.000	1.000	1.000	1.000	1.000
KocurPalus	1.000	1.000	1.000	1.000	1.000
KocurRhizo	1.000	1.000	1.000	1.000	1.000
KocurSP	1.000	1.000	1.000	1.000	1.000
LactoLacti	1.000	1.000	1.000	1.000	1.000
LeifsSoli	0.975	0.950	1.000	1.000	0.974
ListeMonoc	1.000	1.000	1.000	1.000	1.000
LysinCompo	1.000	1.000	1.000	1.000	1.000
LysinFusif	1.000	1.000	1.000	1.000	1.000

LysinOdyss	1.000	1.000	1.000	1.000	1.000
MesobSP	1.000	1.000	1.000	1.000	1.000
MetabHalos	1.000	1.000	1.000	1.000	1.000
MicroFlavu	1.000	1.000	1.000	1.000	1.000
MicroLacti	1.000	1.000	1.000	1.000	1.000
MicroLuteu	1.000	1.000	1.000	1.000	1.000
MicroLylae	1.000	1.000	1.000	1.000	1.000
MicroParao	1.000	1.000	1.000	1.000	1.000
MicroProte	1.000	1.000	1.000	1.000	1.000
MicroSP	1.000	1.000	1.000	1.000	1.000
MicroTerre	1.000	1.000	1.000	1.000	1.000
MoraxOsloe	1.000	1.000	0.999	0.882	0.938
NiallCircu	0.932	0.864	1.000	0.951	0.905
NiallSP	1.000	1.000	1.000	1.000	1.000
OceanChiro	1.000	1.000	1.000	1.000	1.000
OceanKimch	1.000	1.000	1.000	1.000	1.000
OrnitBavar	1.000	1.000	1.000	1.000	1.000
PaeniAmylo	1.000	1.000	1.000	1.000	1.000
PaeniCampi	1.000	1.000	1.000	1.000	1.000
PaeniGluca	1.000	1.000	0.999	0.917	0.957
PaeniLautu	1.000	1.000	1.000	1.000	1.000
PaeniProve	1.000	1.000	0.999	0.866	0.928
PaeniSP	1.000	1.000	1.000	1.000	1.000
PaeniSeget	1.000	1.000	1.000	1.000	1.000
PaeniTaich	0.977	0.956	0.999	0.870	0.911
PantoSP	0.971	0.941	1.000	1.000	0.970
PantoSepti	1.000	1.000	0.999	0.917	0.957
ParacYeei	1.000	1.000	1.000	1.000	1.000
PeribButan	1.000	1.000	1.000	1.000	1.000
PeribPsych	1.000	1.000	1.000	1.000	1.000
PlanoOkean	1.000	1.000	1.000	1.000	1.000
ProliFasci	0.988	0.976	0.999	0.890	0.931
ProteMirab	1.000	1.000	1.000	1.000	1.000
PseudAerug	1.000	1.000	1.000	1.000	1.000
PseudAzoto	1.000	1.000	1.000	0.938	0.968
PseudChlor	1.000	1.000	1.000	1.000	1.000
PseudSP	1.000	1.000	1.000	1.000	1.000
PseudStutz	1.000	1.000	1.000	1.000	1.000
RalstPicke	1.000	1.000	1.000	1.000	1.000
RaoulPlant	0.909	0.818	1.000	1.000	0.900
RhizoRadio	1.000	1.000	1.000	1.000	1.000
RhodoMucil	1.000	1.000	1.000	1.000	1.000

RoseoMucos	1.000	1.000	1.000	1.000	1.000
RothiAeria	1.000	1.000	1.000	1.000	1.000
RothiDento	1.000	1.000	1.000	1.000	1.000
RothiKrist	1.000	1.000	1.000	1.000	1.000
RothiTerra	0.950	0.900	1.000	1.000	0.947
ShigeFlexn	1.000	1.000	0.999	0.917	0.957
ShingOlei	0.875	0.750	1.000	1.000	0.857
SphinCarot	1.000	1.000	1.000	1.000	1.000
SphinMelon	0.969	0.938	1.000	1.000	0.968
SphinMulti	1.000	1.000	1.000	1.000	1.000
SphinPanac	0.999	1.000	0.998	0.800	0.889
SphinSP	1.000	1.000	1.000	1.000	1.000
SphinTerra	1.000	1.000	1.000	1.000	1.000
StaphArgen	0.950	0.900	1.000	1.000	0.947
StaphArlet	1.000	1.000	1.000	1.000	1.000
StaphAureu	0.998	1.000	0.995	0.572	0.728
StaphAuric	0.900	0.800	1.000	0.960	0.873
StaphCapit	1.000	1.000	1.000	1.000	1.000
StaphCapra	0.990	0.981	0.999	0.827	0.898
StaphCohni	0.950	0.900	1.000	1.000	0.947
StaphCondi	1.000	1.000	1.000	1.000	1.000
StaphEdaph	1.000	1.000	1.000	1.000	1.000
StaphEpide	0.994	0.988	1.000	1.000	0.994
StaphHaemo	0.982	0.963	1.000	0.935	0.949
StaphHomin	0.971	0.941	1.000	1.000	0.970
StaphLentu	1.000	1.000	1.000	1.000	1.000
StaphLugdu	0.952	0.906	0.999	0.806	0.853
StaphPaste	0.875	0.750	1.000	1.000	0.857
StaphPetra	1.000	1.000	1.000	1.000	1.000
StaphPette	0.921	0.842	1.000	1.000	0.914
StaphSapro	0.925	0.850	1.000	1.000	0.919
StaphSimul	1.000	1.000	1.000	1.000	1.000
StaphSucci	1.000	1.000	1.000	1.000	1.000
StaphWarne	0.926	0.852	1.000	0.923	0.886
StenoPanac	1.000	1.000	1.000	1.000	1.000
StrepSaliv	1.000	1.000	1.000	1.000	1.000
StrepSangu	1.000	1.000	1.000	1.000	1.000
StreptocSP	1.000	1.000	1.000	1.000	1.000
StreptomSP	1.000	1.000	1.000	1.000	1.000
XanthSP	1.000	1.000	1.000	1.000	1.000
Average	0.990	0.980	1.000	0.984	0.981

Table 4.12 List of 48 species performance in the validation of the training model using whole data matrix. Classification accuracy of the unknown microbial isolates - F1 score 100%.

class	accuracy	sensitivity	specificity	precision	F1
AcineIwoff	1.000	1.000	1.000	1.000	1.000
BacilAltit	1.000	1.000	1.000	1.000	1.000
BacilAmylo	1.000	1.000	1.000	1.000	1.000
BacilCereu	1.000	1.000	1.000	1.000	1.000
BacilCircu	1.000	1.000	1.000	1.000	1.000
BacilInfan	1.000	1.000	1.000	1.000	1.000
BacilLiche	1.000	1.000	1.000	1.000	1.000
BacilMegat	1.000	1.000	1.000	1.000	1.000
BacilOcean	1.000	1.000	1.000	1.000	1.000
BreviCeler	1.000	1.000	1.000	1.000	1.000
BreviFrigo	1.000	1.000	1.000	1.000	1.000
CandiAlbic	1.000	1.000	1.000	1.000	1.000
CandiGuill	1.000	1.000	1.000	1.000	1.000
CandiParap	1.000	1.000	1.000	1.000	1.000
CorynTuber	1.000	1.000	1.000	1.000	1.000
CorynXeros	1.000	1.000	1.000	1.000	1.000
ExophDerma	1.000	1.000	1.000	1.000	1.000
KocurPalus	1.000	1.000	1.000	1.000	1.000
KocurRhizo	1.000	1.000	1.000	1.000	1.000
LeifsSoli	1.000	1.000	1.000	1.000	1.000
MetabHalos	1.000	1.000	1.000	1.000	1.000
MicroLuteu	1.000	1.000	1.000	1.000	1.000
MicroParao	1.000	1.000	1.000	1.000	1.000
MicroProte	1.000	1.000	1.000	1.000	1.000
MoraxOsloe	1.000	1.000	1.000	1.000	1.000
NiallCircu	1.000	1.000	1.000	1.000	1.000
PaeniGluca	1.000	1.000	1.000	1.000	1.000
PaeniTaich	1.000	1.000	1.000	1.000	1.000
ParacYeei	1.000	1.000	1.000	1.000	1.000
ProliFasci	1.000	1.000	1.000	1.000	1.000
RalstPicke	1.000	1.000	1.000	1.000	1.000
RhodoMucil	1.000	1.000	1.000	1.000	1.000
RoseoMucos	1.000	1.000	1.000	1.000	1.000
ShigeFlexn	1.000	1.000	1.000	1.000	1.000
SphinMelon	1.000	1.000	1.000	1.000	1.000
StaphArgen	1.000	1.000	1.000	1.000	1.000
StaphAureu	1.000	1.000	1.000	1.000	1.000
StaphCapit	1.000	1.000	1.000	1.000	1.000

StaphCapra	1.000	1.000	1.000	1.000	1.000
StaphCohni	1.000	1.000	1.000	1.000	1.000
StaphEdaph	1.000	1.000	1.000	1.000	1.000
StaphEvide	1.000	1.000	1.000	1.000	1.000
StaphHaemo	1.000	1.000	1.000	1.000	1.000
StaphHomin	1.000	1.000	1.000	1.000	1.000
StaphLugdu	1.000	1.000	1.000	1.000	1.000
StaphPetra	1.000	1.000	1.000	1.000	1.000
StaphSapro	1.000	1.000	1.000	1.000	1.000
StaphWarne	1.000	1.000	1.000	1.000	1.000
Average	1.000	1.000	1.000	1.000	1.000

Table 4.13 List of 162 species performance in the classification training model using 75 statistically significant RFE features. Model accuracy F1 score 92.84.

class	accuracy	sensitivity	specificity	precision	F1
AcineBauma	0.833	0.667	1.000	0.909	0.769
AcineIwoff	0.952	0.904	1.000	0.942	0.923
AcineRadio	0.966	0.933	0.998	0.737	0.824
AcineSchin	0.961	0.923	1.000	0.975	0.948
AcineTowne	1.000	1.000	1.000	0.981	0.990
AcineUrsin	0.933	0.867	1.000	1.000	0.929
AerocUrina	1.000	1.000	1.000	1.000	1.000
AerocVirid	1.000	1.000	1.000	1.000	1.000
AgrocJenen	1.000	1.000	1.000	1.000	1.000
ArthrLuteo	1.000	1.000	1.000	1.000	1.000
BacilAltit	0.983	0.966	1.000	1.000	0.982
BacilAmylo	0.969	0.938	1.000	0.934	0.936
BacilAtrop	1.000	1.000	1.000	1.000	1.000
BacilCereu	1.000	1.000	0.999	0.889	0.942
BacilCircu	1.000	1.000	1.000	1.000	1.000
BacilFlexu	0.956	0.913	1.000	0.955	0.933
BacilHalos	0.923	0.846	1.000	1.000	0.917
BacilInfan	0.981	0.962	1.000	1.000	0.980
BacilLiche	0.898	0.797	1.000	0.955	0.869
BacilMassi	1.000	1.000	1.000	0.971	0.985
BacilMegat	0.983	0.967	1.000	0.957	0.962
BacilMycoi	0.954	0.909	0.998	0.773	0.836
BacilOcean	0.978	0.957	0.999	0.906	0.931
BacilParam	0.967	0.933	1.000	1.000	0.966
BacilPumil	1.000	1.000	1.000	1.000	1.000
BacilSP	0.963	0.929	0.997	0.672	0.780
BacilSafen	0.950	0.900	1.000	0.963	0.930
BacilSubti	1.000	1.000	1.000	1.000	1.000
BacilTequi	0.966	0.933	1.000	0.937	0.935
BacilThuri	0.916	0.833	1.000	0.926	0.877
BhargChang	0.900	0.800	1.000	0.923	0.857
BrachSP	1.000	1.000	1.000	1.000	1.000
BreviAydin	0.889	0.778	0.999	0.903	0.836
BreviCasei	0.893	0.786	1.000	0.961	0.864
BreviCeler	0.961	0.923	0.999	0.836	0.878
BreviFrigo	0.994	0.987	1.000	1.000	0.993
BreviIodin	1.000	1.000	1.000	1.000	1.000
BreviParab	1.000	1.000	1.000	1.000	1.000

BreviPityo	1.000	1.000	1.000	1.000	1.000
CandiAlbic	1.000	1.000	1.000	0.958	0.979
CandiGuill	1.000	1.000	1.000	1.000	1.000
CandiParap	0.978	0.957	1.000	1.000	0.978
CitriZhaca	1.000	1.000	1.000	1.000	1.000
CitroFreun	0.999	1.000	0.999	0.810	0.895
CorynAccol	1.000	1.000	1.000	0.952	0.976
CorynAferm	1.000	1.000	1.000	1.000	1.000
CorynAurim	1.000	1.000	1.000	1.000	1.000
CorynCoyle	1.000	1.000	1.000	1.000	1.000
CorynPropi	1.000	1.000	1.000	1.000	1.000
CorynSangu	1.000	1.000	1.000	1.000	1.000
CorynTuber	0.975	0.950	1.000	1.000	0.974
CorynXeros	1.000	1.000	1.000	1.000	1.000
DermaNishi	1.000	1.000	1.000	1.000	1.000
DietzPapil	1.000	1.000	1.000	1.000	1.000
EnterAerog	0.944	0.889	1.000	1.000	0.941
EnterFaeca	1.000	1.000	1.000	1.000	1.000
EnterKobei	1.000	1.000	1.000	1.000	1.000
EscheColi	1.000	1.000	1.000	0.933	0.966
EscheFergu	0.947	0.895	0.999	0.866	0.880
ExiguSP	1.000	1.000	1.000	1.000	1.000
ExophDerma	0.975	0.950	1.000	1.000	0.974
ExophSP	1.000	1.000	1.000	1.000	1.000
FictiBarba	0.954	0.909	0.998	0.789	0.845
FictiPhosp	0.933	0.867	1.000	1.000	0.929
FusobNecro	0.999	1.000	0.997	0.692	0.818
GeorgMural	1.000	1.000	1.000	1.000	1.000
GordoBronc	0.944	0.889	1.000	1.000	0.941
GordoSP	0.900	0.800	1.000	1.000	0.889
GordoSputi	0.975	0.950	0.999	0.863	0.904
JanibSP	1.000	1.000	1.000	1.000	1.000
KlebsPneum	1.000	1.000	1.000	1.000	1.000
KocurArsen	1.000	1.000	0.999	0.885	0.939
KocurPalus	0.968	0.935	1.000	1.000	0.967
KocurRhizo	0.965	0.932	0.998	0.790	0.855
KocurSP	1.000	1.000	1.000	1.000	1.000
LactoLacti	0.929	0.857	1.000	1.000	0.923
LeifsSoli	0.924	0.850	0.998	0.754	0.799
ListeMonoc	0.889	0.778	1.000	1.000	0.875
LysinCompo	1.000	1.000	0.999	0.923	0.960
LysinFusif	1.000	1.000	1.000	1.000	1.000

LysinOdyss	1.000	1.000	1.000	1.000	1.000
MesobSP	1.000	1.000	1.000	1.000	1.000
MetabHalos	0.919	0.839	1.000	0.922	0.878
MicroFlavu	0.958	0.917	1.000	1.000	0.957
MicroLacti	0.850	0.700	1.000	1.000	0.824
MicroLuteu	0.958	0.915	1.000	0.951	0.933
MicroLylae	0.976	0.952	0.999	0.834	0.889
MicroParao	0.999	1.000	0.999	0.833	0.909
MicroProte	0.918	0.837	0.999	0.846	0.842
MicroSP	1.000	1.000	1.000	1.000	1.000
MicroTerre	0.923	0.846	1.000	0.952	0.896
MoraxOsloe	0.993	0.985	1.000	0.961	0.973
NiallCircu	0.954	0.909	1.000	0.941	0.925
NiallSP	0.889	0.778	1.000	0.963	0.860
OceanChiro	1.000	1.000	1.000	1.000	1.000
OceanKimch	1.000	1.000	1.000	1.000	1.000
OrnitBavar	1.000	1.000	1.000	0.973	0.986
PaeniAmylo	0.912	0.824	1.000	0.915	0.867
PaeniCampi	0.950	0.900	1.000	1.000	0.947
PaeniGluca	0.985	0.970	1.000	0.978	0.974
PaeniLautu	1.000	1.000	1.000	1.000	1.000
PaeniProve	1.000	1.000	0.999	0.911	0.954
PaeniSP	1.000	1.000	0.999	0.909	0.952
PaeniSeget	1.000	1.000	1.000	1.000	1.000
PaeniTaich	0.967	0.933	1.000	0.960	0.947
PantoSP	0.853	0.706	0.999	0.857	0.774
PantoSepti	0.941	0.882	0.999	0.907	0.894
ParacYeei	0.980	0.960	1.000	0.985	0.972
PeribButan	1.000	1.000	0.999	0.882	0.938
PeribPsych	1.000	1.000	1.000	1.000	1.000
PlanoOkean	1.000	1.000	1.000	1.000	1.000
ProliFasci	0.964	0.929	0.999	0.873	0.900
ProteMirab	0.944	0.889	0.999	0.889	0.889
PseudAerug	1.000	1.000	0.999	0.913	0.955
PseudAzoto	0.958	0.917	0.999	0.910	0.913
PseudChlor	0.999	1.000	0.999	0.857	0.923
PseudSP	0.889	0.778	0.999	0.901	0.835
PseudStutz	0.952	0.905	1.000	1.000	0.950
RalstPicke	1.000	1.000	1.000	1.000	1.000
RaoulPlant	0.909	0.818	1.000	1.000	0.900
RhizoRadio	0.999	1.000	0.999	0.806	0.893
RhodoMucil	1.000	1.000	1.000	0.952	0.976

RoseoMucos	0.859	0.720	0.999	0.760	0.740
RothiAeria	1.000	1.000	1.000	1.000	1.000
RothiDento	0.785	0.571	0.999	0.837	0.679
RothiKrist	1.000	1.000	1.000	1.000	1.000
RothiTerra	0.950	0.900	1.000	1.000	0.947
ShigeFlexn	0.940	0.881	0.999	0.893	0.887
ShingOlei	0.875	0.750	1.000	1.000	0.857
SphinCarot	1.000	1.000	1.000	0.988	0.994
SphinMelon	0.969	0.938	1.000	1.000	0.968
SphinMulti	1.000	1.000	1.000	1.000	1.000
SphinPanac	0.999	1.000	0.998	0.800	0.889
SphinSP	0.997	1.000	0.995	0.546	0.707
SphinTerra	1.000	1.000	1.000	1.000	1.000
StaphArgen	0.933	0.867	0.999	0.895	0.881
StaphArlet	0.808	0.615	1.000	1.000	0.762
StaphAureu	0.988	0.981	0.995	0.531	0.689
StaphAuric	0.800	0.600	1.000	1.000	0.750
StaphCapit	1.000	1.000	0.999	0.871	0.931
StaphCapra	0.972	0.944	0.999	0.912	0.928
StaphCohni	0.908	0.817	0.999	0.900	0.856
StaphCondi	0.888	0.778	0.998	0.757	0.767
StaphEdaph	1.000	1.000	1.000	0.938	0.968
StaphEpide	0.937	0.875	0.999	0.885	0.880
StaphHaemo	0.914	0.829	0.999	0.864	0.846
StaphHomin	0.958	0.918	0.999	0.834	0.874
StaphLentu	1.000	1.000	1.000	1.000	1.000
StaphLugdu	0.936	0.875	0.997	0.674	0.762
StaphPaste	0.708	0.417	0.999	0.652	0.508
StaphPetra	0.889	0.778	1.000	1.000	0.875
StaphPette	0.921	0.842	1.000	1.000	0.914
StaphSapro	0.900	0.800	1.000	0.941	0.865
StaphSimul	1.000	1.000	1.000	1.000	1.000
StaphSucci	0.999	1.000	0.999	0.856	0.922
StaphWarne	0.912	0.827	0.998	0.696	0.756
StenoPanac	1.000	1.000	1.000	1.000	1.000
StrepSaliv	1.000	1.000	1.000	1.000	1.000
StrepSangu	1.000	1.000	1.000	1.000	1.000
StreptocSP	1.000	1.000	1.000	1.000	1.000
StreptomSP	1.000	1.000	1.000	1.000	1.000
XanthSP	1.000	1.000	1.000	1.000	1.000
Average	0.964	0.928	1.000	0.937	0.928

Table 4.14 List of 48 species performance in the validation of the training model developed using 75 statistically significant RFE features. Classification accuracy F1 score 83.7%.

class	accuracy	sensitivity	specificity	precision	F1
AcineIwoff	1.000	1.000	1.000	1.000	1.000
BacilAltit	1.000	1.000	1.000	1.000	1.000
BacilAmylo	0.833	0.667	1.000	1.000	0.800
BacilCereu	1.000	1.000	1.000	1.000	1.000
BacilCircu	1.000	1.000	1.000	1.000	1.000
BacilInfan	1.000	1.000	1.000	1.000	1.000
BacilLiche	1.000	1.000	1.000	1.000	1.000
BacilMegat	0.917	0.833	1.000	1.000	0.909
BacilOcean	1.000	1.000	1.000	1.000	1.000
BreviCeler	0.000	0.000	0.995	0.000	0.000
BreviFrigo	1.000	1.000	1.000	1.000	1.000
CandiAlbic	0.983	0.967	1.000	1.000	0.983
CandiGuill	1.000	1.000	1.000	1.000	1.000
CandiParap	1.000	1.000	1.000	1.000	1.000
CorynTuber	1.000	1.000	1.000	1.000	1.000
CorynXeros	1.000	1.000	1.000	1.000	1.000
ExophDerma	1.000	1.000	1.000	1.000	1.000
KocurPalus	0.000	0.000	0.997	0.000	0.000
KocurRhizo	1.000	1.000	1.000	1.000	1.000
LeifsSoli	0.000	0.000	0.998	0.000	0.000
MetabHalos	0.000	0.000	0.999	0.000	0.000
MicroLuteu	1.000	1.000	0.999	0.968	0.984
MicroParao	1.000	1.000	1.000	1.000	1.000
MicroProte	0.999	1.000	0.999	0.947	0.973
MoraxOsloe	1.000	1.000	1.000	1.000	1.000
NiallCircu	1.000	1.000	1.000	1.000	1.000
PaeniGluca	1.000	1.000	1.000	1.000	1.000
PaeniTaich	0.944	0.889	1.000	1.000	0.941
ParacYeei	1.000	1.000	1.000	1.000	1.000
ProliFasci	1.000	1.000	1.000	1.000	1.000
RalstPicke	0.000	0.000	0.998	0.000	0.000
RhodoMucil	0.916	0.833	0.998	0.909	0.870
RoseoMucos	1.000	1.000	1.000	1.000	1.000
ShigeFlexn	1.000	1.000	1.000	1.000	1.000
SphinMelon	1.000	1.000	1.000	1.000	1.000
StaphArgen	1.000	1.000	1.000	1.000	1.000
StaphAureu	1.000	1.000	1.000	1.000	1.000
StaphCapit	1.000	1.000	1.000	1.000	1.000

StaphCapra	0.917	0.833	1.000	1.000	0.909
StaphCohni	1.000	1.000	1.000	1.000	1.000
StaphEdaph	0.944	0.889	0.999	0.960	0.923
StaphEpide	0.000	0.000	0.998	0.000	0.000
StaphHaemo	1.000	1.000	1.000	1.000	1.000
StaphHomin	0.950	0.900	1.000	1.000	0.947
StaphLugdu	1.000	1.000	1.000	1.000	1.000
StaphPetra	0.972	0.958	0.986	0.590	0.730
StaphSapro	0.583	0.167	1.000	1.000	0.286
StaphWarne	0.998	1.000	0.996	0.826	0.905
Average	0.975	0.951	0.999	0.858	0.837

Bibliography

1. Prescott LM, Klein PH. *Microbiology 5th Edition*. ©2002 / Hardcover with CDROM; 2002.
2. Carr FJ. *Microbiology : A Fundamental Introduction.*; 2016.
3. Lim D V, Florida S. Introduction, Development of Microbiology. *Life Sci*. 2001.
4. Briški F, Vuković Domanovac M. Environmental microbiology. *Phys Sci Rev*. 2019;2(11). doi:10.1515/psr-2016-0118
5. Pepper IL, Gerba CP, Gentry TJ. Introduction to Environmental Microbiology. In: *Environmental Microbiology: Third Edition*. ; 2015. doi:10.1016/B978-0-12-394626-3.00001-6
6. Guo L, Ye L, Zhao Q, Ma Y, Yang J, Luo Y. Comparative study of MALDI-TOF MS and VITEK 2 in bacteria identification. *J Thorac Dis*. 2014;6(5):534-538. doi:10.3978/j.issn.2072-1439.2014.02.18
7. Sandle T. Good practices for pharmaceutical microbiology laboratories. *GMP Rev*. 2012.
8. McCormick K, Sanders JH. Pharmaceutical microbiology. In: *Quality*. ; 2022. doi:10.1016/b978-0-323-90815-3.00013-x
9. Charry N, Gómez SL. Limits determination of microbial contamination present on surfaces from a pharmaceutical microbiology district reference laboratory. *J Pharm Pharmacogn Res*. 2016.
10. Magdy M. Brief Introduction to Pharmaceutical Microbiologist. *J Microbiol Exp*. 2014. doi:10.15406/jmen.2014.01.00029
11. Kapoor D, Sharma P, Sharma MMM, Kumari A, Kumar R. Microbes in pharmaceutical industry. In: *Microbial Diversity, Interventions and Scope*. ; 2020. doi:10.1007/978-981-15-4099-8_16
12. Vijayakumar R, Saleh Al-Aboody M, Sandle T. A review of melanized (black) fungal contamination in pharmaceutical products-incidence, drug recall and control measures. *J Appl Microbiol*. 2016. doi:10.1111/jam.12888

13. Wackett LP. Pharmaceuticals and microbes in wastewater: An annotated selection of World Wide Web sites relevant to the topics in environmental microbiology. *Environ Microbiol.* 2020. doi:10.1111/1462-2920.15283
14. Severns M. Sterility testing in 7 days -Speeding time to market through a rapid sterility test. *Am Pharm Rev.* 2015.
15. Mohamad Seif El-Din Ashour, Moselhy Salah Mansy MEE. Microbiological Enviroment Monitoring in Pharneutical Facility. *Egypt Acad J Biol Sci.* 2011;3(1):63-74.
16. Cundell AM. Microbial identification strategies in the pharmaceutical industry. *PDA J Pharm Sci Technol.* 2006;60(2):111-123.
17. HARRIS M. Pharmaceutical Microbiology Manual 2014. *Acad Med.* 1964;39(11):xix. doi:10.1097/00001888-196411000-00034
18. Pharmaceutical Microbiology. *J Pharm Pharmacol.* 2010. doi:10.1211/002235707781850078
19. Gholizadeh-Hashjin A, lotfipour farzaneh, Hallaj-Nezhadi S. Quality Control of Non-Sterile Drug Product According to United States' Pharmacopeia Instruction. *Iran J Med Microbiol.* 2019. doi:10.30699/ijmm.13.5.321
20. Sandle T. *Pharmaceutical Microbiology: Essentials for Quality Assurance and Quality Control.*; 2015. doi:10.1016/C2014-0-00532-1
21. *Pharmaceutical Microbiological Quality Assurance and Control.*; 2019. doi:10.1002/9781119356196
22. Lemos MA, Ghisleni DDM, de Jesus Andreoli Pinto T. Validation of Alternative Microbiological Method in Non-sterile Pharmaceutical Product Through the Reference Strains and Productive Environment Bioburden. *J Pharm Innov.* 2020;15(4):641-650. doi:10.1007/s12247-019-09408-1
23. Bagnis S, Fitzsimons M, Snape J, Tappin A, Comber S. Sorption of active pharmaceutical ingredients in untreated wastewater effluent and effect of dilution in freshwater: Implications for an “impact zone” environmental risk assessment approach. *Sci Total Environ.* 2018;624:333-341. doi:10.1016/j.scitotenv.2017.12.092

24. Dao H, Lakhani P, Police A, et al. Microbial Stability of Pharmaceutical and Cosmetic Products. *AAPS PharmSciTech*. 2018;19(1):60-78. doi:10.1208/s12249-017-0875-1
25. Selim NA, Saeed AM, Ibrahim MK. Monitoring and controlling bacteria in pharmaceutical industries water system. *J Appl Microbiol*. 2020. doi:10.1111/jam.14686
26. Sakarikou C, Altieri A, Bossa MC, et al. Rapid and cost-effective identification and antimicrobial susceptibility testing in patients with Gram-negative bacteremia directly from blood-culture fluid. *J Microbiol Methods*. 2018;146(January):7-12. doi:10.1016/j.mimet.2018.01.012
27. Gunar O V., Sakhno NG. Methods of Microbial Identification During Drug Quality Analysis and Applicability Assessment. *Pharm Chem J*. 2019. doi:10.1007/s11094-019-01930-3
28. Jimenez L. Microorganisms in the environment and their relevance to pharmaceutical processes. In: *Microbial Contamination Control in the Pharmaceutical Industry*. ; 2004. doi:10.1201/9780203026267.ch1
29. Vu N, Lou JR, Kupiec TC. Quality Control: Microbial Limit Tests for Nonsterile Pharmaceuticals, Part 2. *Int J Pharm Compd*. 2014.
30. Vu N, Lou JR, Kupiec TC. Test for Absence of Specified Microorganism. *Int J Pharm Compd*. 2014.
31. Surette C, Scherer B, Corwin A, et al. Rapid Microbiology Screening in Pharmaceutical Workflows. *SLAS Technol*. 2018;23(4):387-394. doi:10.1177/2472630318779758
32. Ratajczak M, Kubicka MM, Kamińska D, Sawicka P, Długaszewska J. Microbiological quality of non-sterile pharmaceutical products. *Saudi Pharm J*. 2015;23(3):303-307. doi:10.1016/j.jsps.2014.11.015
33. Steel KJ. Microbial identification. *J Gen Microbiol*. 1965. doi:10.1099/00221287-40-1-143
34. Liu S, Gunawan C, Barraud N, Rice SA, Harry EJ, Amal R. Understanding, monitoring, and controlling biofilm growth in drinking water distribution systems. *Environ Sci Technol*. 2016. doi:10.1021/acs.est.6b00835
35. Hyde T, Anstead J, Schade L, Zellner J. Microbial Identification in Pharmaceutical

- Compounding. *Int J Pharm Compd.* 2016.
36. Sutton SVW, Cundell AM, Porter D. Microbial identification in the pharmaceutical industry. *Pharmaceutial Forum.* 2004.
 37. de Brito NMR, Lourenço FR. Rapid identification of microbial contaminants in pharmaceutical products using a PCA/LDA-based FTIR-ATR method. *Brazilian J Pharm Sci.* 2021. doi:10.1590/s2175-97902020000318899
 38. Jang KS, Kim YH. Rapid and robust MALDI-TOF MS techniques for microbial identification: a brief overview of their diverse applications. *J Microbiol.* 2018. doi:10.1007/s12275-018-7457-0
 39. Van Belkum A, Durand G, Peyret M, et al. Rapid clinical bacteriology and its future impact. *Ann Lab Med.* 2013. doi:10.3343/alm.2013.33.1.14
 40. Berg EL. The future of phenotypic drug discovery. *Cell Chem Biol.* 2021. doi:10.1016/j.chembiol.2021.01.010
 41. Lowy F. Bacterial Classification , Structure and Function. *Columbia Univ.* 2004.
 42. Clarke T, Brinkac L, Manoranjan J, et al. Typing and classification of non-tuberculous mycobacteria isolates. *F1000Research.* 2020. doi:10.12688/f1000research.22620.1
 43. Zahir T, Camacho R, Vitale R, et al. High-throughput time-resolved morphology screening in bacteria reveals phenotypic responses to antibiotics. *Commun Biol.* 2019. doi:10.1038/s42003-019-0480-9
 44. Costa LV da, Miranda RV da SL de, Fonseca EL da, et al. Assessment of VITEK® 2, MALDI-TOF MS and full gene 16S rRNA sequencing for aerobic endospore-forming bacteria isolated from a pharmaceutical facility. *J Microbiol Methods.* 2022. doi:10.1016/j.mimet.2022.106419
 45. Bilecen K, Yaman G, Ciftci U, Laleli YR. Performances and reliability of Bruker microflex LT and VITEK MS MALDI-TOF mass spectrometry systems for the identification of clinical microorganisms. *Biomed Res Int.* 2015;2015(Ivd). doi:10.1155/2015/516410
 46. Yoon SH, Ha SM, Kwon S, et al. Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol*

Microbiol. 2017. doi:10.1099/ijsem.0.001755

47. Klemetsen T, Willassen NP, Karlsen CR. Full-length 16S rRNA gene classification of Atlantic salmon bacteria and effects of using different 16S variable regions on community structure analysis. *Microbiologyopen*. 2019. doi:10.1002/mbo3.898
48. Paul B, Kavia Raj K, Murali TS, Satyamoorthy K. Species-specific genomic sequences for classification of bacteria. *Comput Biol Med.* 2020. doi:10.1016/j.compbiomed.2020.103874
49. Johnson JS, Spakowicz DJ, Hong BY, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun.* 2019. doi:10.1038/s41467-019-13036-1
50. Odds F. Biochemical Tests for Identification of Medical Bacteria. *J Clin Pathol.* 1981. doi:10.1136/jcp.34.5.572-a
51. Christ APG, Ramos SR, Cayô R, Gales AC, Hachich EM, Sato MIZ. Characterization of Enterococcus species isolated from marine recreational waters by MALDI-TOF MS and Rapid ID API® 20 Strep system. *Mar Pollut Bull.* 2017;118(1-2):376-381. doi:10.1016/j.marpolbul.2017.03.025
52. Burke CM, Darling AE. A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. *PeerJ.* 2016. doi:10.7717/peerj.2492
53. Lear G, Dickie I, Banks J, et al. Methods for the extraction, storage, amplification and sequencing of dna from environmental samples. *N Z J Ecol.* 2018. doi:10.20417/nzj ecol.42.9
54. Armanhi JSL, De Souza RSC, De Araújo LM, et al. Multiplex amplicon sequencing for microbe identification in community-based culture collections. *Sci Rep.* 2016. doi:10.1038/srep29543
55. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J Clin Microbiol.* 2007. doi:10.1128/JCM.01228-07
56. Petti CA. Detection and identification of microorganisms by gene amplification and sequencing. *Clin Infect Dis.* 2007. doi:10.1086/512818

57. Nübel U, Garcia-Pichel F, Muyzer G. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol.* 1997. doi:10.1128/aem.63.8.3327-3332.1997
58. Bej AK, Mahbubani MH, Dicesare JL, Atlas RM. Polymerase chain reaction-gene probe detection of microorganisms by using filter-concentrated samples. *Appl Environ Microbiol.* 1991. doi:10.1128/aem.57.12.3529-3534.1991
59. Baidoo EEK, Benke PI, Keasling JD. Mass spectrometry-based microbial metabolomics. *Methods Mol Biol.* 2012. doi:10.1007/978-1-61779-827-6_9
60. Baidoo EEK, Teixeira Benites V. Mass spectrometry-based microbial metabolomics: Techniques, analysis, and applications. In: *Methods in Molecular Biology.* ; 2019. doi:10.1007/978-1-4939-8757-3_2
61. Bauermeister A, Mannocho-Russo H, Costa-Lotufo L V., Jarmusch AK, Dorrestein PC. Mass spectrometry-based metabolomics in microbiome investigations. *Nat Rev Microbiol.* 2022. doi:10.1038/s41579-021-00621-9
62. Biosoft. Mass Spectrometry: Introduction, Principle of Mass Spectrometry, Components of Mass Spectrometer, Applications. *Prem Biosoft.* 2015.
63. Boesl U. Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrom Rev.* 2017. doi:10.1002/mas.21520
64. Welker M. Proteomics for routine identification of microorganisms. *Proteomics.* 2011. doi:10.1002/pmic.201100049
65. Fang J, Dorrestein PC. Emerging mass spectrometry techniques for the direct analysis of microbial colonies. *Curr Opin Microbiol.* 2014;19(1):120-129. doi:10.1016/j.mib.2014.06.014
66. Sandrin TR, Demirev PA. Characterization of microbial mixtures by mass spectrometry. *Mass Spectrom Rev.* 2018;37(3):321-349. doi:10.1002/mas.21534
67. Smith SA, Smith RW, Xia Y, Ouyang Z. Introduction to Mass Spectrometry. *Charact Impurities Degrad Using Mass Spectrom.* 2011. doi:10.1002/9780470921371.ch1
68. Matthiesen R, Bunkenborg J. Introduction to mass spectrometry-based proteomics. *Methods Mol Biol.* 2013. doi:10.1007/978-1-62703-392-3_1
69. Harned A. Mass Spectrometry : Introduction. *Interpret A J Bible Theol.* 2009.

70. Evans CH. Introduction to mass spectrometry. *Trends Biochem Sci.* 1986. doi:10.1016/0968-0004(86)90058-7
71. March RE. An introduction to quadrupole ion trap mass spectrometry. *J Mass Spectrom.* 1997. doi:10.1002/(SICI)1096-9888(199704)32:4<351::AID-JMS512>3.0.CO;2-Y
72. Chernushevich I V., Loboda A V., Thomson BA. An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom.* 2001. doi:10.1002/jms.207
73. Siuzdak G. Mass Spectrometry - Introduction - Ion sources & sample introduction - Mass analyzers - Basics of biomolecule MS - Applications Mass Spectrometry Introduction. *Mass Spectrom Biotechnol.* 1996.
74. Niessen WMA, Falck D. Introduction to Mass Spectrometry, a Tutorial. In: *Analyzing Biomolecular Interactions by Mass Spectrometry.* ; 2013. doi:10.1002/9783527673391.ch1
75. Vorce SP. Mass Spectrometry. In: *Principles of Forensic Toxicology: Fifth Edition.* ; 2020. doi:10.1007/978-3-030-42917-1_14
76. Züllig T, Köfeler HC. HIGH RESOLUTION MASS SPECTROMETRY IN LIPIDOMICS. *Mass Spectrom Rev.* 2021. doi:10.1002/mas.21627
77. Heiles S. Advanced tandem mass spectrometry in metabolomics and lipidomics—methods and applications. *Anal Bioanal Chem.* 2021. doi:10.1007/s00216-021-03425-1
78. van Agthoven MA, Lam YPY, O'Connor PB, Rolando C, Delsuc MA. Two-dimensional mass spectrometry: new perspectives for tandem mass spectrometry. *Eur Biophys J.* 2019. doi:10.1007/s00249-019-01348-5
79. Robert E. Ardrey. LIQUID CHROMATOGRAPHY– MASS SPECTROMETRY: AN INTRODUCTION. In: *Canadian Geographic.* ; 2003.
80. Smyth W. Electrospray Ionisation - Mass Spectrometry (ESI-MS) and Liquid Chromatography -Electrospray Ionisation - Mass Spectrometry (LC-ESI-MS) of Selected Pharmaceuticals. *Curr Pharm Anal.* 2006. doi:10.2174/157341206777934644
81. Bibi F, Naseer MI, Yasir M, Al-Ghamdi AAK, Azhar EI. LC-MS based identification of secondary metabolites from marine antagonistic endophytic bacteria. *Genet Mol Res.* 2017. doi:10.4238/gmr16039857

82. Chromacademy. Mass Spectrometry Fundamental LC-MS Introduction. *Crawford Sci.* 2004.
83. Graham C, McMullan G, Graham RLJ. Proteomics in the microbial sciences. *Bioeng Bugs*. 2011. doi:10.4161/bbug.2.1.14413
84. Demirev PA, Fenselau C. Mass spectrometry for rapid characterization of microorganisms. *Annu Rev Anal Chem.* 2008. doi:10.1146/annurev.anchem.1.031207.112838
85. Ribaud G, Ongaro A, Oselladore E, Memo M, Gianoncelli A. Combining Electrospray Mass Spectrometry (ESI-MS) and Computational Techniques in the Assessment of G-Quadruplex Ligands: A Hybrid Approach to Optimize Hit Discovery. *J Med Chem.* 2021. doi:10.1021/acs.jmedchem.1c00962
86. Di Marco VB, Bombi GG. Electrospray mass spectrometry (ESI-MS) in the study of metal-ligand solution equilibria. *Mass Spectrom Rev.* 2006. doi:10.1002/mas.20070
87. Lavigne JP, Espinal P, Dunyach-Remy C, Messad N, Pantel A, Sotto A. Mass spectrometry: A revolution in clinical microbiology? *Clin Chem Lab Med.* 2013. doi:10.1515/cclm-2012-0291
88. Gao P, Xu G. Mass-spectrometry-based microbial metabolomics: Recent developments and applications. *Anal Bioanal Chem.* 2015. doi:10.1007/s00216-014-8127-7
89. Tsuchida S, Umemura H, Nakayama T. Current Status of Matrix-Assisted Laser. *Molecules.* 2020.
90. Panda A, Kurapati S, Samantaray JC, Srinivasan A, Khalil S. MALDI-TOF mass spectrometry proteomic based identification of clinical bacterial isolates. *Indian J Med Res.* 2014.
91. Bizzini A, Durussel C, Bille J, Greub G, Prod'homme G. Performance of matrix-assisted laser desorption ionization-time of flight mass spectrometry for identification of bacterial strains routinely isolated in a Clinical Microbiology Laboratory. *J Clin Microbiol.* 2010. doi:10.1128/JCM.01794-09
92. Rellos MS, Nievas J, Taie SF, et al. Evaluation of mass spectrometry: MALDI-TOF MS for fast and reliable yeast identification. *Rev Argent Microbiol.* 2015. doi:10.1016/j.ram.2015.02.004

93. Alberici RM, Simas RC, Sanvido GB, et al. Ambient mass spectrometry: Bringing MS into the “real world.” *Anal Bioanal Chem*. 2010. doi:10.1007/s00216-010-3808-3
94. Shi L, Habib A, Bi L, Hong H, Begum R, Wen L. Ambient Ionization Mass Spectrometry: Application and Prospective. *Crit Rev Anal Chem*. 2022. doi:10.1080/10408347.2022.2124840
95. Griffiths RL, Kocurek KI, Cooper HJ. Ambient surface mass spectrometry–ion mobility spectrometry of intact proteins. *Curr Opin Chem Biol*. 2018. doi:10.1016/j.cbpa.2017.11.002
96. Clendinen CS, Monge ME, Fernández FM. Ambient mass spectrometry in metabolomics. *Analyst*. 2017. doi:10.1039/c7an00700k
97. Qi K, Wu L, Liu C, Pan Y. Recent advances of ambient mass spectrometry imaging and its applications in lipid and metabolite analysis. *Metabolites*. 2021. doi:10.3390/metabo11110780
98. Ha J, Hong SK, Han GH, Kim M, Yong D, Lee K. Same-day identification and antimicrobial susceptibility testing of bacteria in positive blood culture broths using short-term incubation on solid medium with the MicroFlex LT, Vitek-MS, and Vitek2 systems. *Ann Lab Med*. 2018;38(3):235-241. doi:10.3343/alm.2018.38.3.235
99. Cheng K, Chui H, Domish L, Hernandez D, Wang G. Recent development of mass spectrometry and proteomics applications in identification and typing of bacteria. *Proteomics - Clin Appl*. 2016. doi:10.1002/prca.201500086
100. Claydon MA, Davey SN, Edwards-Jones V, Gordon DB. The Rapid Identification of Intact Microorganisms Using Mass Spectrometry. *Nat Biotechnol*. 1996. doi:10.1038/nbt1196-1584
101. Denyer SP, Hodges NA, Gorman SP. Introduction to Pharmaceutical Microbiology. *Hugo Russell's Pharm Microbiol Seventh Ed*. 2007. doi:10.1002/9780470988329.ch1
102. Sauer S, Freiwald A, Maier T, et al. Classification and identification of bacteria by mass spectrometry and computational analysis. *PLoS One*. 2008. doi:10.1371/journal.pone.0002843
103. Karlsson R, Gonzales-Siles L, Boulund F, et al. Proteotyping: Proteomic characterization, classification and identification of microorganisms - A prospectus. *Syst*

Appl Microbiol. 2015. doi:10.1016/j.syapm.2015.03.006

104. Wilkendorf LS, Bowles E, Buil JB, et al. Update on matrix-assisted laser desorption ionization-time of flight mass spectrometry identification of filamentous fungi. *J Clin Microbiol.* 2020. doi:10.1128/JCM.01263-20
105. Andrade LDO, Awasthi R, Dua K, Andreoli Pinto TDJ. Matrix-assisted laser desorption ionization–time of flight mass spectrometry for identification of bacteria isolated from pharmaceutical clean rooms. *Interv Med Appl Sci.* 2018. doi:10.1556/1646.9.2017.40
106. Fothergill A, Kasinathan V, Hyman J, Walsh J, Drake T, Wayne Wang YF. Rapid identification of bacteria and yeasts from positive-blood-culture bottles by using a lysis-filtration method and matrix- assisted laser desorption ionization-time of flight mass spectrum analysis with the SARAMIS database. *J Clin Microbiol.* 2013;51(3):805-809. doi:10.1128/JCM.02326-12
107. Martin IW. Mass spectrometry in the clinical microbiology laboratory. In: *Mass Spectrometry for the Clinical Laboratory.* ; 2017. doi:10.1016/B978-0-12-800871-3.00011-0
108. Šedo O, Sedláček I, Zdráhal Z. Sample preparation methods for MALDI-MS profiling of bacteria. *Mass Spectrom Rev.* 2011. doi:10.1002/mas.20287
109. Feucherolles M, Frache G. MALDI Mass Spectrometry Imaging: A Potential Game-Changer in a Modern Microbiology. *Cells.* 2022. doi:10.3390/cells11233900
110. Eydmann M, Ball D, Sadler G, Wilks M. Introduction of MALDI-TOF : a revolution in diagnostic. *Biomed Sci.* 2011.
111. El-Aneed A, Cohen A, Banoub J. Mass spectrometry, review of the basics: Electrospray, MALDI, and commonly used mass analyzers. *Appl Spectrosc Rev.* 2009. doi:10.1080/05704920902717872
112. Wahl KL, Wunschel SC, Jarman KH, et al. Analysis of microbial mixtures by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Anal Chem.* 2002;74(24):6191-6199. doi:10.1021/ac0203847
113. Lara BR, De Camargo BB, Paula CR, et al. Comparing the phenotypic, genotypic, and proteomic identification of *Trichosporon* species: A globally emerging yeast of medical importance. *Med Mycol.* 2021. doi:10.1093/mmy/myab050

114. Mahmud I, Garrett TJ. Mass Spectrometry Techniques in Emerging Pathogens Studies: COVID-19 Perspectives. *J Am Soc Mass Spectrom*. 2020. doi:10.1021/jasms.0c00238
115. Valentine N, Wunschel S, Wunschel D, Petersen C, Wahl K. Effect of culture conditions on microorganism identification by matrix-assisted laser desorption ionization mass spectrometry. *Appl Environ Microbiol*. 2005;71(1):58-64. doi:10.1128/AEM.71.1.58-64.2005
116. Chen JHK, Ho PL, Kwan GSW, et al. Direct bacterial identification in positive blood cultures by use of two commercial matrix-assisted laser desorption ionization-time of flight mass spectrometry systems. *J Clin Microbiol*. 2013;51(6):1733-1739. doi:10.1128/JCM.03259-12
117. Jarman KH, Cebula ST, Saenz AJ, et al. An algorithm for automated bacterial identification using matrix- assisted laser desorption/ionization mass spectrometry. *Anal Chem*. 2000;72(6):1217-1223. doi:10.1021/ac990832j
118. Russell SC. Microorganism characterization by single particle mass spectrometry. *Mass Spectrom Rev*. 2009. doi:10.1002/mas.20198
119. Cheng D, Qiao L, Horvatovich P. Toward Spectral Library-Free Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Bacterial Identification. *J Proteome Res*. 2018;17(6):2124-2130. doi:10.1021/acs.jproteome.8b00065
120. Seng P, Drancourt M, Gouriet F, et al. Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. *Clin Infect Dis*. 2009;49(4):543-551. doi:10.1086/600885
121. Katase M, Tsumura K. High-throughput identification of coliforms in processed soybean products using matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Food Sci Technol Res*. 2014;20(5):1093-1097. doi:10.3136/fstr.20.1093
122. Kitagawa K, Shigemura K, Onuma KI, et al. Improved bacterial identification directly from urine samples with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *J Clin Lab Anal*. 2018;32(3):1-7. doi:10.1002/jcla.22301
123. Wilkins CL, Lay JO. *Identification of Microorganisms by Mass Spectrometry*.; 2005. doi:10.1002/0471748641
124. Ashfaq MY, Da'na DA, Al-Ghouti MA. Application of MALDI-TOF MS for

- identification of environmental bacteria: A review. *J Environ Manage.* 2022. doi:10.1016/j.jenvman.2021.114359
125. Demirev PA, Ho YP, Ryzhov V, Fenselau C. Microorganism identification by mass spectrometry and protein database searches. *Anal Chem.* 1999. doi:10.1021/ac990165u
 126. Rahi P, Prakash O, Shouche YS. Matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) based microbial identifications: Challenges and scopes for microbial ecologists. *Front Microbiol.* 2016. doi:10.3389/fmicb.2016.01359
 127. Rychert J. Benefits and Limitations of MALDI-TOF Mass Spectrometry for the Identification of Microorganisms. *J Infect.* 2019. doi:10.29245/2689-9981/2019/4.1142
 128. Schubert S, Weinert K, Wagner C, et al. Novel, improved sample preparation for rapid, direct identification from positive blood cultures using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. *J Mol Diagnostics.* 2011;13(6):701-706. doi:10.1016/j.jmoldx.2011.07.004
 129. Kaleta EJ, Clark AE, Cherkaoui A, et al. Comparative analysis of PCR - Electrospray ionization/mass spectrometry (MS) and MALDI-TOF/MS for the identification of bacteria and yeast from positive blood culture bottles. *Clin Chem.* 2011;57(7):1057-1067. doi:10.1373/clinchem.2011.161968
 130. Santos IC, Hildenbrand ZL, Schug KA. Applications of MALDI-TOF MS in environmental microbiology. *Analyst.* 2016;141(10):2827-2837. doi:10.1039/c6an00131a
 131. Alizadeh M, Yousefi L, Pakdel F, et al. MALDI-TOF Mass Spectroscopy Applications in Clinical Microbiology. *Adv Pharmacol Pharm Sci.* 2021. doi:10.1155/2021/9928238
 132. Christner M, Trusch M, Rohde H, et al. Rapid MALDI-TOF mass spectrometry strain typing during a large outbreak of Shiga-toxigenic *Escherichia coli*. *PLoS One.* 2014;9(7):1-11. doi:10.1371/journal.pone.0101924
 133. Yan Y, He Y, Maier T, et al. Improved identification of yeast species directly from positive blood culture media by combining sepsityper specimen processing and microflex analysis with the matrix-assisted laser desorption ionization biotyper system. *J Clin Microbiol.* 2011;49(7):2528-2532. doi:10.1128/JCM.00339-11

134. Li Y, Shan M, Zhu Z, et al. Application of MALDI-TOF MS to rapid identification of anaerobic bacteria. *BMC Infect Dis*. 2019. doi:10.1186/s12879-019-4584-0
135. Oviaño M, Rodríguez-Sánchez B. MALDI-TOF mass spectrometry in the 21st century clinical microbiology laboratory. *Enferm Infecc Microbiol Clin*. 2021. doi:10.1016/j.eimc.2020.02.027
136. Hou TY, Chiang-Ni C, Teng SH. Current status of MALDI-TOF mass spectrometry in clinical microbiology. *J Food Drug Anal*. 2019;27(2):404-414. doi:10.1016/j.jfda.2019.01.001
137. Dare DJ, Sutton H, Wang X, et al. MALDI-TOF MASS SPECTROMETRY : A RAPID METHOD FOR BACTERIAL IDENTIFICATION FROM WATER SAMPLES The City of New York Department of Environmental. 2005;4(3):2005.
138. Popović NT, Kazazić SP, Strunjak-Perović I, Čož-Rakovac R. Differentiation of environmental aquatic bacterial isolates by MALDI-TOF MS. *Environ Res*. 2017;152(October 2016):7-16. doi:10.1016/j.envres.2016.09.020
139. Akimowicz M, Bucka-Kolendo J. MALDI-TOF MS-Application in food microbiology. *Acta Biochim Pol*. 2020. doi:10.18388/ABP.2020_5380
140. Wang H, Li Y, Fan X, Chiueh TS, Xu YC, Hsueh PR. Evaluation of Bruker Biotyper and Vitek MS for the identification of *Candida tropicalis* on different solid culture media. *J Microbiol Immunol Infect*. 2019. doi:10.1016/j.jmii.2017.11.002
141. Li Y, Wang H, Zhao YP, Xu YC, Hsueh PR. Evaluation of the Bruker Biotyper matrix-assisted laser desorption/ionization time-of-flight mass spectrometry system for identification of *Aspergillus* species directly from growth on solid agar media. *Front Microbiol*. 2017. doi:10.3389/fmicb.2017.01209
142. de Alegría Puig CR, Torres MF, Marfil-Pérez E, et al. Comparison between Vitek MS, Bruker Biotyper, Vitek2, and API20E for differentiation of species of the genus *Raoultella*. *Eur J Clin Microbiol Infect Dis*. 2019. doi:10.1007/s10096-018-03444-4
143. Hamal P, Vavrova A, Mrazek J, Svobodova L. Identification of filamentous fungi including dermatophytes using MALDI-TOF mass spectrometry. *Folia Microbiol (Praha)*. 2022. doi:10.1007/s12223-021-00917-6
144. Robert MG, Romero C, Dard C, et al. Evaluation of ID fungi plates medium for

- identification of molds by MALDI biotyper. *J Clin Microbiol.* 2020;58(5):1-8. doi:10.1128/JCM.01687-19
145. Iqbal KK, Farrance CE, Khot PD, Laboratories CR. Evaluation and Optimization of MALDI-TOF Mass Spectrometry for Identification of Filamentous Fungi during Environmental Monitoring. (Id):7342.
 146. Hsueh PR, Lee TF, Du SH, et al. Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry system for identification of *Nocardia*, *Rhodococcus*, *Kocuria*, *Gordonia*, *Tsukamurella*, and *Listeria* species. *J Clin Microbiol.* 2014. doi:10.1128/JCM.00456-14
 147. Zhu W, Sieradzki K, Albrecht V, et al. Evaluation of the Biotyper MALDI-TOF MS system for identification of *Staphylococcus* species. *J Microbiol Methods.* 2015. doi:10.1016/j.mimet.2015.07.014
 148. Veloo ACM, Jean-Pierre H, Justesen US, et al. Validation of MALDI-TOF MS Biotyper database optimized for anaerobic bacteria: The ENRIA project. *Anaerobe.* 2018;54:224-230. doi:10.1016/j.anaerobe.2018.03.007
 149. Veloo ACM, de Vries ED, Jean-Pierre H, et al. The optimization and validation of the Biotyper MALDI-TOF MS database for the identification of Gram-positive anaerobic cocci. *Clin Microbiol Infect.* 2016. doi:10.1016/j.cmi.2016.06.016
 150. Florio W, Tavanti A, Barnini S, Ghelardi E, Lupetti A. Recent advances and ongoing challenges in the diagnosis of microbial infections by MALDI-TOF mass spectrometry. *Front Microbiol.* 2018. doi:10.3389/fmicb.2018.01097
 151. Han SS, Jeong YS, Choi SK. Current scenario and challenges in the direct identification of microorganisms using maldi tof ms. *Microorganisms.* 2021. doi:10.3390/microorganisms9091917
 152. Su H, Jiang ZH, Chiou SF, et al. Rapid Characterization of Bacterial Lipids with Ambient Ionization Mass Spectrometry for Species Differentiation. *Molecules.* 2022;27(9):1-18. doi:10.3390/molecules27092772
 153. Rankin-Turner S, Heaney LM. Applications of ambient ionization mass spectrometry in 2020: An annual review. *Anal Sci Adv.* 2021. doi:10.1002/ansa.202000135
 154. Huang MZ, Cheng SC, Cho YT, Shiea J. Ambient ionization mass spectrometry: A

- tutorial. *Anal Chim Acta*. 2011. doi:10.1016/j.aca.2011.06.017
155. Lu H, Zhang H, Xu S, Li L. Review of recent advances in lipid analysis of biological samples via ambient ionization mass spectrometry. *Metabolites*. 2021. doi:10.3390/metabo11110781
 156. Lebedev AT. Ambient ionization mass spectrometry. *Russ Chem Rev*. 2015. doi:10.1070/rcr4508
 157. Rankin-Turner S, Reynolds JC, Turner MA, Heaney LM. Applications of ambient ionization mass spectrometry in 2021: An annual review. *Anal Sci Adv*. 2022. doi:10.1002/ansa.202100067
 158. Chen R, Deng J, Fang L, et al. Recent applications of ambient ionization mass spectrometry in environmental analysis. *Trends Environ Anal Chem*. 2017. doi:10.1016/j.teac.2017.07.001
 159. Li L-H, Hsieh H-Y, Hsu C-C. Clinical Application of Ambient Ionization Mass Spectrometry. *Mass Spectrom*. 2017. doi:10.5702/massspectrometry.s0060
 160. ZHANG XL, ZHANG H, WANG XC, HUANG KK, WANG D, CHEN HW. Advances in Ambient Ionization for Mass Spectrometry. *Chinese J Anal Chem*. 2018. doi:10.1016/S1872-2040(18)61122-3
 161. Su H, Jiang ZH, Chiou SF, et al. Rapid Characterization of Bacterial Lipids with Ambient Ionization Mass Spectrometry for Species Differentiation. *Molecules*. 2022. doi:10.3390/molecules27092772
 162. Takáts Z, Wiseman JM, Gologan B, Cooks RG. Mass spectrometry sampling under ambient conditions with desorption electrospray ionization. *Science (80-)*. 2004. doi:10.1126/science.1104404
 163. Takáts Z, Wiseman JM, Cooks RG. Ambient mass spectrometry using desorption electrospray ionization (DESI): Instrumentation, mechanisms and applications in forensics, chemistry, and biology. *J Mass Spectrom*. 2005;40(10):1261-1275. doi:10.1002/jms.922
 164. Pasilis SP. Desorption Electrospray Ionization Mass Spectrometry. In: *Encyclopedia of Analytical Chemistry*. ; 2010. doi:10.1002/9780470027318.a9013

165. Wójtowicz A, Wietecha-Posłuszny R. DESI-MS analysis of human fluids and tissues for forensic applications. *Appl Phys A Mater Sci Process*. 2019. doi:10.1007/s00339-019-2564-2
166. Watrous J, Hendricks N, Meehan M, Dorrestein PC. Capturing bacterial metabolic exchange using thin film desorption electrospray ionization-imaging mass spectrometry. *Anal Chem*. 2010. doi:10.1021/ac9027388
167. Watrous J, Roach P, Alexandrov T, et al. Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A*. 2012. doi:10.1073/pnas.1203689109
168. Meetani MA, Shin YS, Zhang S, Mayer R, Basile F. Desorption electrospray ionization mass spectrometry of intact bacteria. *J Mass Spectrom*. 2007. doi:10.1002/jms.1250
169. Wu C, Dill AL, Eberlin LS, Cooks RG, Ifa DR. Mass spectrometry imaging under ambient conditions. *Mass Spectrom Rev*. 2013. doi:10.1002/mas.21360
170. Lanekoff I, Geydebrekht O, Pinchuk GE, Konopka AE, Laskin J. Spatially resolved analysis of glycolipids and metabolites in living *Synechococcus* sp. PCC 7002 using nanospray desorption electrospray ionization. *Analyst*. 2013. doi:10.1039/c3an36716a
171. Zhang JI, Talaty N, Costa AB, et al. Rapid direct lipid profiling of bacteria using desorption electrospray ionization mass spectrometry. In: *International Journal of Mass Spectrometry*. ; 2011. doi:10.1016/j.ijms.2010.06.014
172. Roach PJ, Laskin J, Laskin A. Nanospray desorption electrospray ionization: An ambient method for liquid-extraction surface sampling in mass spectrometry. *Analyst*. 2010. doi:10.1039/c0an00312c
173. Watrous J, Roach P, Heath B, Alexandrov T, Laskin J, Dorrestein PC. Metabolic profiling directly from the petri dish using nanospray desorption electrospray ionization imaging mass spectrometry. *Anal Chem*. 2013;85(21):10385-10391. doi:10.1021/ac4023154
174. Cardoso-Palacios C, Lanekoff I. Direct Analysis of Pharmaceutical Drugs Using Nano-DESI MS. *J Anal Methods Chem*. 2016. doi:10.1155/2016/3591908
175. H.V. L, P. P, R.G. B, et al. Rapid mucosal metabolome profiling by desorption electrospray ionization MS (DESI-MS) for prediction of preterm pre-labour rupture of membranes (PPROM). *BJOG An Int J Obstet Gynaecol*. 2018.

176. Song Y, Talaty N, Datsenko K, Wanner BL, Cooks RG. In vivo recognition of *Bacillus subtilis* by desorption electrospray ionization mass spectrometry (DESI-MS). *Analyst*. 2009. doi:10.1039/b900069k
177. Tata A, Perez CJ, Ore MO, et al. Evaluation of imprint DESI-MS substrates for the analysis of fungal metabolites. *RSC Adv*. 2015. doi:10.1039/c5ra12805f
178. Cody RB, Laramée JA, Durst HD. Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal Chem*. 2005;77(8):2297-2302. doi:10.1021/ac050162j
179. Gross JH. Direct analysis in real time-a critical review on DART-MS. *Anal Bioanal Chem*. 2014. doi:10.1007/s00216-013-7316-0
180. Nilles JM, Connell TR, Durst HD. Quantitation of chemical warfare agents using the direct analysis in real time (DART) technique. *Anal Chem*. 2009. doi:10.1021/ac900682f
181. Shelley JT, Wiley JS, Chan GCY, Schilling GD, Ray SJ, Hieftje GM. Characterization of direct-current atmospheric-pressure discharges useful for ambient desorption/ionization mass spectrometry. *J Am Soc Mass Spectrom*. 2009;20(5):837-844. doi:10.1016/j.jasms.2008.12.020
182. Fox A. Mass spectrometry for species or strain identification after culture or without culture: Past, present, and future. *J Clin Microbiol*. 2006. doi:10.1128/JCM.00971-06
183. Vaclavik L, Cajka T, Hrbek V, Hajslova J. Ambient mass spectrometry employing direct analysis in real time (DART) ion source for olive oil quality and authenticity assessment. *Anal Chim Acta*. 2009. doi:10.1016/j.aca.2009.04.043
184. Cody RB, McAlpin CR, Cox CR, Jensen KR, Voorhees KJ. Identification of bacteria by fatty acid profiling with direct analysis in real time mass spectrometry. *Rapid Commun Mass Spectrom*. 2015. doi:10.1002/rcm.7309
185. Cajka T, Riddellova K, Tomaniova M, Hajslova J. Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: A powerful tool for beer origin recognition. *Metabolomics*. 2011. doi:10.1007/s11306-010-0266-z
186. Arora M, Zambrzycki SC, Levy JM, et al. Machine Learning Approaches to Identify Discriminative Signatures of Volatile Organic Compounds (VOCs) from Bacteria and

- Fungi Using SPME-DART-MS. *Metabolites*. 2022. doi:10.3390/metabo12030232
187. Gupta S, Samal N. Application of direct analysis in real-time mass spectrometry (DART-MS) in forensic science: a comprehensive review. *Egypt J Forensic Sci*. 2022. doi:10.1186/s41935-022-00276-4
 188. Domin MA, Steinberg BD, Quimby JM, Smith NJ, Greene AK, Scott LT. Routine analysis and characterization of highly insoluble polycyclic aromatic compounds by direct analysis in real time mass spectrometry (DART). *Analyst*. 2010. doi:10.1039/b923300h
 189. Sisco E, Forbes TP. Forensic applications of DART-MS: A review of recent literature. *Forensic Chem*. 2021. doi:10.1016/j.forc.2020.100294
 190. Musah RA, Espinoza EO, Cody RB, et al. A high throughput ambient mass spectrometric approach to species identification and classification from chemical fingerprint signatures. *Sci Rep*. 2015. doi:10.1038/srep11520
 191. Shen L, Zhang J, Yang Q, Manicke NE, Ouyang Z. High throughput paper spray mass spectrometry analysis. *Clin Chim Acta*. 2013. doi:10.1016/j.cca.2012.10.025
 192. Zhi-Ping Z, Xiao-Ning L, Ya-Jun Z. Ambient ionization-paper spray ionization and its application. *Fenxi Huaxue/ Chinese J Anal Chem*. 2014. doi:10.1016/S1872-2040(13)60706-9
 193. Liu J, He X, He Y, Yu M, Jiang L, Chen B. Development and Application of Paper Spray Ionization Mass Spectrometry. *Prog Chem*. 2017. doi:10.7536/PC170352
 194. Han W, Zheng Y, Muyizere T, Zhang Z. Development of paper substrate for paper spray MS in high-sensitivity analysis of biological samples. *Bioanalysis*. 2018. doi:10.4155/bio-2018-0199
 195. Basuri P, Das S, Jenifer SK, Jana SK, Pradeep T. Microdroplet Impact-Induced Spray Ionization Mass Spectrometry (MISI MS) for Online Reaction Monitoring and Bacteria Discrimination. *J Am Soc Mass Spectrom*. 2021. doi:10.1021/jasms.0c00365
 196. Liu C, Li Y, Xiong C, Nie Z. Development of capillary-paper spray for small-molecule analysis in complex samples. *Anal Bioanal Chem*. 2021. doi:10.1007/s00216-020-03072-y

197. Wei P, Bag S, Pulliam CJ, Snyder DT, Pielak RM, Cooks RG. Analysis of bacteria using zero volt paper spray. *Anal Methods*. 2016. doi:10.1039/c6ay00063k
198. Chamberlain CA, Rubio VY, Garrett TJ. Strain-Level Differentiation of Bacteria by Paper Spray Ionization Mass Spectrometry. *Anal Chem*. 2019. doi:10.1021/acs.analchem.9b00330
199. Espy RD, Muliadi AR, Ouyang Z, Cooks RG. Spray mechanism in paper spray ionization. *Int J Mass Spectrom*. 2012. doi:10.1016/j.ijms.2012.06.017
200. Liu J, Wang H, Manicke NE, Lin JM, Cooks RG, Ouyang Z. Development, characterization, and application of paper spray ionization. *Anal Chem*. 2010. doi:10.1021/ac902854g
201. Smith BL, Boisdon C, Romero-Perez D, et al. Ambient ion focusing for paper spray ionisation. *Int J Mass Spectrom*. 2022. doi:10.1016/j.ijms.2021.116737
202. Chiang S, Zhang W, Ouyang Z. Paper spray ionization mass spectrometry: recent advances and clinical applications. *Expert Rev Proteomics*. 2018. doi:10.1080/14789450.2018.1525295
203. Nguyen TMH, Song WY, Kim TY. Characterization of Spray Modes and Factors Affecting the Ionization Efficiency of Paper Spray Ionization. *Front Chem*. 2022. doi:10.3389/fchem.2022.864184
204. Mclean K, Palarea-Albaladejo J, Currie CG, et al. Rapid and robust analytical protocol for E. coli STEC bacteria subspecies differentiation using whole cell MALDI mass spectrometry. *Talanta*. 2018;182(January):164-170. doi:10.1016/j.talanta.2018.01.055
205. Görgens C, Walker K, Boeser C, et al. Paper spray mass spectrometry – A potential complementary technique for the detection of polar compounds in sports drug testing. *Drug Test Anal*. 2020. doi:10.1002/dta.2909
206. Frey BS, Damon DE, Badu-Tawiah AK. Emerging trends in paper spray mass spectrometry: Microsampling, storage, direct analysis, and applications. *Mass Spectrom Rev*. 2020. doi:10.1002/mas.21601
207. McBride EM, Mach PM, Dhummakupt ES, et al. Paper spray ionization: Applications and perspectives. *TrAC - Trends Anal Chem*. 2019. doi:10.1016/j.trac.2019.06.028

208. Balog J, Sasi-Szabó L, Kinross J, et al. Intraoperative tissue identification using rapid evaporative ionization mass spectrometry. *Sci Transl Med.* 2013;5(194). doi:10.1126/scitranslmed.3005623
209. Paxton T. Rapid evaporative ionization mass spectrometry. In: *Ambient Ionization Mass Spectrometry in Life Sciences: Principles and Applications.* ; 2019. doi:10.1016/B978-0-12-817220-9.00008-4
210. St John E, White E, Balog J, et al. The intelligent knife for detection of invasive breast cancer at radial margins: an intraoperative feasibility trial. *Ann Surg Oncol Conf 18th Annu Meet Am Soc breast Surg ASBS 2017 United states.* 2017.
211. Schäfer KC, Dénes J, Albrecht K, et al. In vivo, in situ tissue analysis using rapid evaporative ionization mass spectrometry. *Angew Chemie - Int Ed.* 2009. doi:10.1002/anie.200902546
212. Phelps DL, Balog J, Gildea LF, et al. The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). *Br J Cancer.* 2018. doi:10.1038/s41416-018-0048-3
213. Van Hese L, Vaysse PM, Siegel TP, Heeren R, Rex S, Cuypers E. Real-time drug detection using a diathermic knife combined to rapid evaporative ionisation mass spectrometry. *Talanta.* 2021. doi:10.1016/j.talanta.2020.121391
214. Takats Z, Strittmatter N, McKenzie JS. Ambient Mass Spectrometry in Cancer Research. In: *Advances in Cancer Research.* ; 2017. doi:10.1016/bs.acr.2016.11.011
215. St John ER, Balog J, McKenzie JS, et al. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: Towards an intelligent knife for breast cancer surgery. *Breast Cancer Res.* 2017. doi:10.1186/s13058-017-0845-2
216. Yau A, Fear MW, Gray N, et al. Enhancing the accuracy of surgical wound excision following burns trauma via application of Rapid Evaporative Ionisation Mass Spectrometry (REIMS). *Burns.* 2022. doi:10.1016/j.burns.2022.08.021
217. Kelis Cardoso VG, Sabin GP, Hantao LW. Rapid evaporative ionization mass spectrometry (REIMS) combined with chemometrics for real-time beer analysis. *Anal Methods.* 2022. doi:10.1039/d2ay00063f

218. Murphy F, Gathercole J, Lee E, et al. Discrimination of milk fermented with different starter cultures by MALDI-TOF MS and REIMS fingerprinting. *Int Dairy J.* 2021. doi:10.1016/j.idairyj.2021.105143
219. Balog J, Szaniszlo T, Schaefer KC, et al. Identification of biological tissues by rapid evaporative ionization mass spectrometry. *Anal Chem.* 2010. doi:10.1021/ac101283x
220. Strittmatter N. Development of novel mass spectrometric methods for the characterisation and identification of microorganisms.
221. Strittmatter N, Jones EA, Veselkov KA, Rebec M, Bundy JG, Takats Z. Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chem Commun.* 2013;49(55):6188-6190. doi:10.1039/c3cc42015a
222. Bodai Z, Cameron S, Bolt F, et al. Effect of Electrode Geometry on the Classification Performance of Rapid Evaporative Ionization Mass Spectrometric (REIMS) Bacterial Identification. *J Am Soc Mass Spectrom.* 2018;29(1):26-33. doi:10.1007/s13361-017-1818-5
223. Cameron SJS, Bolt F, Perdonés-Montero A, et al. Rapid Evaporative Ionisation Mass Spectrometry (REIMS) Provides Accurate Direct from Culture Species Identification within the Genus *Candida*. *Sci Rep.* 2016;6(October):1-10. doi:10.1038/srep36788
224. Strittmatter N, Rebec M, Jones EA, et al. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. *Anal Chem.* 2014;86(13):6555-6562. doi:10.1021/ac501075f
225. Cameron SJS, Perdonés-Montero A, Van Meulebroek L, et al. Sample Preparation Free Mass Spectrometry Using Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry: Applications to Microbiology, Metabolic Biofluid Phenotyping, and Food Authenticity. *J Am Soc Mass Spectrom.* 2021. doi:10.1021/jasms.0c00452
226. Paraskevaïdi M, Cameron SJS, Whelan E, et al. Laser-assisted rapid evaporative ionisation mass spectrometry (LA-REIMS) as a metabolomics platform in cervical cancer screening. *EBioMedicine.* 2020;60. doi:10.1016/j.ebiom.2020.103017
227. Van Meulebroek L, Cameron S, Plekhova V, et al. Rapid LA-REIMS and comprehensive UHPLC-HRMS for metabolic phenotyping of feces. *Talanta.* 2020;217(December 2019):121043. doi:10.1016/j.talanta.2020.121043

228. Bolt F, Cameron SJS, Karancsi T, et al. Automated High-Throughput Identification and Characterization of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionization Mass Spectrometry. *Anal Chem.* 2016;88(19):9419-9426. doi:10.1021/acs.analchem.6b01016
229. Paraskevaïdi M, Cameron SJS, Whelan E, et al. Laser-assisted rapid evaporative ionisation mass spectrometry (LA-REIMS) as a metabolomics platform in cervical cancer screening. *EBioMedicine.* 2020. doi:10.1016/j.ebiom.2020.103017
230. Gowers GOF, Cameron SJS, Perdones-Montero A, et al. Off-Colony Screening of Biosynthetic Libraries by Rapid Laser-Enabled Mass Spectrometry. *ACS Synth Biol.* 2019;8(11):2566-2575. doi:10.1021/acssynbio.9b00243
231. Cameron SJS, Bodai Z, Temelkuran B, et al. Utilisation of Ambient Laser Desorption Ionisation Mass Spectrometry (ALDI-MS) Improves Lipid-Based Microbial Species Level Identification. *Sci Rep.* 2019;9(1):1-8. doi:10.1038/s41598-019-39815-w
232. Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. *Anal Chem.* 2001;73(4):746-750. doi:10.1021/ac0008791
233. Strittmatter N, Jones EA, Veselkov KA, Rebec M, Bundy JG, Takats Z. Analysis of intact bacteria using rapid evaporative ionisation mass spectrometry. *Chem Commun.* 2013;49(55):6188-6190. doi:10.1039/c3cc42015a
234. Cameron SJS, Takáts Z. Mass spectrometry approaches to metabolic profiling of microbial communities within the human gastrointestinal tract. *Methods.* 2018;149:13-24. doi:10.1016/j.ymeth.2018.04.027
235. Barlow RS, Fitzgerald AG, Hughes JM, et al. Rapid evaporative ionization mass spectrometry: A review on its application to the red meat industry with an Australian context. *Metabolites.* 2021. doi:10.3390/metabo11030171
236. Gredell DA, Schroeder AR, Belk KE, et al. Comparison of Machine Learning Algorithms for Predictive Modeling of Beef Attributes Using Rapid Evaporative Ionization Mass Spectrometry (REIMS) Data. *Sci Rep.* 2019;9(1):1-9. doi:10.1038/s41598-019-40927-6
237. Li BQ, Zhang ZY, Kong JL, Zhang L, Guo CH, Li CP. Research Progress of Ambient Ionization Mass Spectrometry. *J Chinese Mass Spectrom Soc.* 2020.

doi:10.7538/zpxb.2019.0013

238. Davis R, Mauer L. Fourier transform infrared (FT-IR) spectroscopy: a rapid tool for detection and analysis of foodborne pathogenic bacteria. *Curr Res Technol Educ Top Appl Microbiol Microb Biotechnol A Méndez-Vilas*. 2010.
239. Naumann D, Fijala V, Labischinski H. The differentiation and identification of pathogenic bacteria using FT-IR and multivariate statistical analysis. *Mikrochim Acta*. 1988. doi:10.1007/BF01205910
240. Pegram S. Use of Fourier Transform Infrared Spectroscopy for the Classification and Identification of Bacteria of Importance to the Food Industry. *Food Microbiol Commons*. 2007.
241. Preisner O, Lopes JA, Menezes JC. Uncertainty assessment in FT-IR spectroscopy based bacteria classification models. *Chemom Intell Lab Syst*. 2008. doi:10.1016/j.chemolab.2008.06.005
242. Oberreuter H, Seiler H, Scherer S. Identification of coryneform bacteria and related taxa by Fourier-transform infrared (FT-IR) spectroscopy. *Int J Syst Evol Microbiol*. 2002. doi:10.1099/00207713-52-1-91
243. Beekes M, Lasch P, Naumann D. Analytical applications of Fourier transform-infrared (FT-IR) spectroscopy in microbiology and prion research. *Vet Microbiol*. 2007. doi:10.1016/j.vetmic.2007.04.010
244. Lefier D, Beccard B, Bradley M. Classification of bacteria using FT-IR. *Spectrosc (Santa Monica)*. 2007.
245. Lorenz B, Wichmann C, Stöckel S, Rösch P, Popp J. Cultivation-Free Raman Spectroscopic Investigations of Bacteria. *Trends Microbiol*. 2017. doi:10.1016/j.tim.2017.01.002
246. Novelli-Rousseau A, Espagnon I, Filiputti D, et al. Culture-free Antibiotic-susceptibility Determination from Single-bacterium Raman Spectra. *Sci Rep*. 2018. doi:10.1038/s41598-018-22392-9
247. Sengupta A, Mujacic M, Davis EJ. Detection of bacteria by surface-enhanced Raman spectroscopy. *Anal Bioanal Chem*. 2006. doi:10.1007/s00216-006-0711-z

248. Pahlow S, Meisel S, Cialla-May D, Weber K, Rösch P, Popp J. Isolation and identification of bacteria by means of Raman spectroscopy. *Adv Drug Deliv Rev.* 2015. doi:10.1016/j.addr.2015.04.006
249. Akanny E, Bonhommé A, Bessueille F, Bourgeois S, Bordes C. Surface enhanced Raman spectroscopy for bacteria analysis: a review. *Appl Spectrosc Rev.* 2021. doi:10.1080/05704928.2020.1796698
250. Tai YH, Lo SC, Montagne K, et al. Enhancing Raman signals from bacteria using dielectrophoretic force between conductive lensed fiber and black silicon. *Biosens Bioelectron.* 2021. doi:10.1016/j.bios.2021.113463
251. Li R, Dhankhar D, Chen J, Krishnamoorthi A, Cesario TC, Rentzepis PM. Identification of live and dead bacteria: A raman spectroscopic study. *IEEE Access.* 2019. doi:10.1109/ACCESS.2019.2899006
252. Gibb S, Strimmer K. Maldiquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics.* 2012. doi:10.1093/bioinformatics/bts447
253. El Morr C, Jammal M, Ali-Hassan H, El-Hallak W. Support Vector Machine. In: *International Series in Operations Research and Management Science.* ; 2022. doi:10.1007/978-3-031-16990-8_13
254. Wong TT, Yeh PY. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Trans Knowl Data Eng.* 2020. doi:10.1109/TKDE.2019.2912815
255. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput.* 2017. doi:10.1007/s11222-016-9696-4
256. Vehtari A, Gelman A, Gabry J. loo package: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models. *Stat Comput.* 2017.
257. Oliveri P, Malegori C, Casale M. Chemometrics and statistics | multivariate classification techniques. In: *Encyclopedia of Analytical Science.* ; 2019. doi:10.1016/B978-0-12-409547-2.14239-8
258. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett.* 2006. doi:10.1016/j.patrec.2005.10.010
259. Lewis MR, Chekmeneva E, Camuzeaux S, et al. An Open Platform for Large Scale LC-

- MS-Based Metabolomics. *ChemRxiv*. 2022.
260. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen KY. Then and now: Use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. *Clin Microbiol Infect*. 2008. doi:10.1111/j.1469-0691.2008.02070.x
261. Elbehiry A, Marzouk E, Hamada M, et al. Application of MALDI-TOF MS fingerprinting as a quick tool for identification and clustering of foodborne pathogens isolated from food products. *New Microbiol*. 2017.
262. The United State Pharmacopeial. <61> Microbiological examination of Nonsterile Products : Microbial Enumeration Tests. *United State Pharmacopeial/National Formul*. 2014;40(2):1-5.
263. Council of Europe. European Pharmacopoeia 8th Edition. *Eur Pharmacopoeia*. 2014.

Permissions to Reprint

Figure 1.2 Permission to reprint

04/10/2022, 03:45 <https://marketplace.copyright.com/rs-ui-web/mp/license/d1d80b4c-20be-4d8e-86a4-6f3aa7c024cf/7fda95b2-aaef-470a-96e...>



This is a License Agreement between Toma Ramonaite/Imperial College London ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rights holder identified in the order details below. The license consists of the order details, the Marketplace Order General Terms and Conditions below, and any Rights holder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Order General Terms and Conditions below.

Order Date	27-Sep-2022	Type of Use	Republish in a
Order License ID	1273127-1		thesis/dissertation
ISSN	1364-5528	Publisher Portion	Royal Society of Chemistry Image/photo/illustration

LICENSED CONTENT

Publication Title	The analyst online	Publication Type	e-Journal
Article Title	Applications of MALDI-TOF MS in environmental microbiology.	Start Page	2827
		End Page	2837
Author/Editor	Society of Public Analysts (Great Britain), Chemical Society (Great Britain), Society for Analytical Chemistry, Society of Public Analysts (Great Britain), Royal Society of Chemistry (Great Britain)	Issue	10
		Volume	141
		URL	http://www.rsc.org/isis/journals/current/analyst/analco.n.htm
Date	01/01/1876		
Language	English		
Country	United Kingdom of Great Britain and Northern Ireland		
Rights holder	Royal Society of Chemistry		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of images / photos / illustrations	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Life of current edition	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 250,000	Currency	GBP
Rights Requested	Main product and any product related to main product		

NEW WORK DETAILS

<https://marketplace.copyright.com/rs-ui-web/mp/license/d1d80b4c-20be-4d8e-86a4-6f3aa7c024cf/7fda95b2-aaef-470a-96e6-edb0d564a453>

1/7

Title	Development of microbial identification platform for industrial pharmaceutical microbiology using high-throughput Laser-Assisted Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS)	Institution name	Imperial College London
		Expected presentation date	2022-12-09
Instructor name	Toma Ramonaite		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Toma Ramonaite/Imperial College London
-------------------------------	-----	--	--

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1	Title of the article/chapter the portion is from	Applications of MALDI-TOF MS in environmental microbiology.
Editor of portion(s)	Santos, InÃ¡s C.; Hildenbrand, Zacariah L.; Schug, Kevin A.	Author of portion(s)	Santos, InÃ¡s C.; Hildenbrand, Zacariah L.; Schug, Kevin A.
Volume of serial or monograph	141	Issue, if republishing an article from a serial	10
Page or page range of portion	2827-2837	Publication date of portion	2016-05-21

Marketplace Order General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation Terms" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

Figure 1.5 Permission to reprint.



 Home

 Help ▾

 Email Support

 Sign in

 Create Account



Paper spray ionization: Applications and perspectives

Author:
Ethan M. McBride, Phillip M. Mach, Elizabeth S. Dhummakupt, Sarah Dowling, Daniel O. Carmany, Paul S. Demond, Gabrielle Rizzo, Nicholas E. Manicke, Trevor Glaros

Publication: TrAC Trends in Analytical Chemistry

Publisher: Elsevier

Date: September 2019

Published by Elsevier B.V.

Creative Commons Attribution-NonCommercial-No Derivatives License (CC BY NC ND)

This article is published under the terms of the [Creative Commons Attribution-NonCommercial-No Derivatives License \(CC BY NC ND\)](#).

For non-commercial purposes you may copy and distribute the article, use portions or extracts from the article in other works, and text or data mine the article, provided you do not alter or modify the article without permission from Elsevier. You may also create adaptations of the article for your own personal use only, but not distribute these to others. You must give appropriate credit to the original work, together with a link to the formal publication through the relevant DOI, and a link to the Creative Commons user license above. If changes are permitted, you must indicate if any changes are made but not in any way that suggests the licensor endorses you or your use of the work.

Permission is not required for this non-commercial use. For commercial use please continue to request permission via RightsLink.

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)
| [For California Residents](#) | [Terms and Conditions](#)Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Figure 1.6 Permission to reprint.



[Home](#)[Help](#)[Email Support](#)

Toma Ramonaite



Sample Preparation Free Mass Spectrometry Using Laser-Assisted Rapid Evaporative Ionization Mass Spectrometry: Applications to Microbiology, Metabolic Biofluid Phenotyping, and Food Authenticity

Author: Simon J. S. Cameron, Alvaro Perdones-Montero, Lieven Van Meulebroek, et al
Publication: Journal of the American Society for Mass Spectrometry
Publisher: American Chemical Society
Date: Jun 1, 2021

Copyright © 2021, American Society for Mass Spectrometry. Published by the American Chemical Society. All rights reserved.

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms and Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your RightsLink request. No additional uses are granted (such as derivative works or other editions). For any uses, please submit a new request.

If credit is given to another source for the material you requested from RightsLink, permission must be obtained from that source.

[BACK](#)[CLOSE WINDOW](#)

© 2022 Copyright - All Rights Reserved | [Copyright Clearance Center, Inc.](#) | [Privacy statement](#) | [Data Security and Privacy](#)
| [For California Residents](#) | [Terms and Conditions](#)Comments? We would like to hear from you. E-mail us at customer@copyright.com

List of Conferences presentations

‘Expanding the Applications of Rapid Evaporative Ionisation Mass Spectrometry (REIMS) to the Pharmaceutical Product Development Workflow’

Oral presentation

Metabolomics 2019, The Hague

‘Expanding the Applications of Rapid Evaporative Ionisation Mass Spectrometry (REIMS) to the Pharmaceutical Product Development Workflow’

Oral presentation

BMSS SIG Meeting 2020, University of Surrey

‘High-throughput Laser-Assisted Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS) platform to the Pharmaceutical Product Development Workflow with No Sample Preparation’

Poster

ASMS Conference on Mass Spectrometry and Allied Topics 2020

‘Rapid microbial identification for pharmaceutical microbiology using Laser-Assisted Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS)’

Poster

BMSS SIG Meeting 2022, University of Surrey

‘Microbial identification platform for pharmaceutical microbiology using high-throughput Laser-Assisted Rapid Evaporative Ionisation Mass Spectrometry (LA-REIMS)’

Oral Presentation

ASMS Conference on Mass Spectrometry and Allied Topics 2022, Minneapolis

Grants and Awards

- ❖ Metabolomics Society Student Travel Award (2019)
- ❖ BBSRC Impact Acceleration Account (IAA) grant (2021)
- ❖ Imperial College Dean's Internship Award (2022)
- ❖ ICURe Innovation to Commercialisation by UKRI BBSRC (2022)