

# Identification of *Escherichia coli* strains using MALDI-TOF MS combined with long short-term memory neural networks

Qiqi Mao<sup>1,\*</sup>, Xie Zhang<sup>2,\*</sup>, Zeping Xu<sup>2</sup>, Ya Xiao<sup>3</sup>, Yufei Song<sup>4</sup>, Feng Xu<sup>4</sup>

<sup>1</sup>Department of General Surgery, Li Huili Hospital Affiliated to Ningbo University, Ningbo 315040, China

<sup>2</sup>Department of Medicine and Pharmacy, Li Huili Hospital Affiliated to Ningbo University, Ningbo 315040, China

<sup>3</sup>School of Medicine, Ningbo University, Ningbo 315211, Zhejiang, China

<sup>4</sup>Department of Gastroenterology, Li Huili Hospital Affiliated to Ningbo University, Ningbo 315040, China

\*Equal contribution

**Correspondence to:** Feng Xu, Yufei Song; email: [xufeng200468@126.com](mailto:xufeng200468@126.com), <https://orcid.org/0000-0003-3962-2918>; [517582445@qq.com](mailto:517582445@qq.com), <https://orcid.org/0000-0002-3874-8722>

**Keywords:** matrix-assisted laser desorption ionization time-of-flight mass spectrometry, long short-term memory neural network, *Escherichia coli*, strain-level identification

**Received:** March 18, 2024

**Accepted:** June 3, 2024

**Published:** June 28, 2024

**Copyright:** © 2024 Mao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

The current study aims to develop a new technique for the precise identification of *Escherichia coli* strains, utilizing matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF MS) combined with a long short-term memory (LSTM) neural network. A total of 48 *Escherichia coli* strains were isolated and cultured on tryptic soy agar medium for 24 hours for the generation of MALDI-TOF MS spectra. Eight hundred MALDI-TOF MS spectra were obtained per strain, resulting in a database of 38,400 spectra. Fifty percent of the data was utilized for LSTM neural network training, with fine-tuned parameters for strain-level identification. The other half served as the test set to assess model performance. Traditional PCA dimension reduction of MALDI-TOF MS spectra indicated 47 out of 48 strains to be unclassifiable. In contrast, the LSTM neural network demonstrated remarkable efficacy. After 20 training epochs, the model achieved a loss value of 0.0524, an accuracy of 0.999, a precision of 0.985, and a recall of 0.982. When tested on the unseen data, the model attained an overall accuracy of 92.24%. The integration of MALDI-TOF MS and LSTM neural network markedly enhances the identification of *Escherichia coli* strains. This innovative approach offers an effective and accurate tool for MALDI-TOF MS-based strain-level identification, thus expanding the analytical capabilities of microbial diagnostics.

## INTRODUCTION

Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) has become an invaluable tool in the rapid identification of microbial species. This technology employs laser energy to enable sample desorption and ionization prior to analysis in a time-of-flight mass spectrometer, determining the sample's precise molecular weights [1–3]. MALDI-TOF MS focuses on whole bacterial proteins, yielding a characteristic “protein fingerprint.” These fingerprints are predominantly composed of

highly-expressed, conserved ribosomal proteins, providing a reliable means for microbial species-level identification [4–6].

Despite its efficacy in rapid and accurate microbial identification, limitations of MALDI-TOF MS currently exist at the species level, owing to its reliance on microbial proteins [7, 8]. Strains within the same species exhibiting high protein expression similarity often remain undifferentiated. Advances in deep learning algorithms, especially long short-term memory (LSTM) neural networks, present a solution

to this limitation. LSTMs are noted for their ability to manage long-term information through a network of input, forget, and output gates, enabling them to identify subtle variations in complex data sequences [9–11].

*Escherichia coli* is an exemplary subject for extending MALDI-TOF MS applications to strain-level identification. *Escherichia coli* strains function both as a harmless component of human flora and as a clinical pathogen. Strain-level identification is essential for tracing the origin of nosocomial infections and reducing associated risks. Recognizing this pressing need and the limitations of existing technologies, this study seeks to explore the utility of integrating MALDI-TOF MS with LSTM neural networks for strain-level identification of *Escherichia coli*. This exploration aims to establish a novel method that advances the analytical capabilities of MALDI-TOF MS, particularly in microbial strain-level diagnostics.

This introduction provides a foundation for the research by first examining the current state of MALDI-TOF MS technology, emphasizing its limitations, and then exploring the potential of LSTM neural networks to overcome these limitations [12, 13]. It also emphasizes the clinical importance of strain-level identification, especially for *Escherichia coli*, thereby establishing the relevance and significance of this study.

## MATERIALS AND METHODS

### Material and chemicals

In this study, 48 strains of *Escherichia coli* were extracted and purified from clinical biological samples. All isolates underwent biochemical testing and were subsequently confirmed as *Escherichia coli* via 16S rRNA gene sequencing. Tryptic Soy Agar (TSA) medium, sourced from Merck Millipore, Germany, was utilized for bacterial culture for MALDI-TOF MS analysis,  $\alpha$ -Cyano-4-hydroxycinnamic acid (CHCA) served as the matrix and was obtained from Sigma-Aldrich, USA. The key instruments included a 4800 Plus MALDI-TOF MS mass spectrometer from Absciex, USA, and a DRP-9272 electric thermostatic microbial incubator provided by Shanghai Senxin Experimental Instrument Co., Ltd.

### MALDI-TOF MS analysis

Colonies of each bacterial strain, cultured over a 24-hour period, were prepared for analysis. A portion of the colony biomass was spread across assigned target sites on the MALDI plate. Subsequently, one microlitre of  $\alpha$ -Cyano-4-hydroxycinnamic acid

(CHCA) matrix solution was applied onto each sample spot. Afterward, the plate was left to air-dry, enabling matrix-sample co-crystallization. The prepared MALDI plate was then inserted into the MALDI-TOF MS instrument set to linear scanning mode. Laser intensity was adjusted to 3500 units, and the mass-to-charge ( $m/z$ ) scanning range was established from 0 to 12,000 Da. For each bacterial strain, a total of 40 sample points were analyzed on the plate. From each point, 20 individual spectra were obtained, accumulating in a composite dataset of 800 spectra per strain. The signal-to-noise ratio for the most intense peak in each spectrum needed to exceed 10 for the data to be deemed valid. Additionally, intra-strain spectral variability was evaluated using Hotelling's  $T^2$  statistical test, with an allowable variance of no more than 5%.

### Preparation of the dataset for MALDI-TOF MS spectral analysis

A comprehensive spectral database was created from the 38,400 acquired MALDI-TOF MS spectra, each meticulously categorized according to their originating strains. Each entry in the database represents a unique spectral signature. For each bacterial strain in the database, the dataset is divided into two mutually exclusive subsets. Specifically, 50% of the individual spectra for each strain are chosen using a stochastic sampling algorithm to form the training set. The remaining 50% comprise the test set. Before this division, all spectra undergo a quality control check to ensure compliance with pre-defined data quality standards, including, but not limited to, signal-to-noise ratios and Hotelling's  $T^2$  statistical thresholds.

### LSTM network model architecture and training protocol

The LSTM model is built using the Tensorflow v2.0 framework. It comprises an LSTM layer, a fully connected layer, and a Dropout layer, with parameters set at 128, 64, and 0.3, respectively. Layers are sequentially connected. Details such as activation functions and output sizes are provided in Table 1. The training loss function is categorical cross-entropy, the optimizer is Adam, and 80% of the spectra in the training set are randomly chosen for training, with the remaining 20% used for cross-validation. The maximum training duration is 20 epochs. Model training results are assessed using precision, accuracy, and recall metrics. The calculation formulae are presented in formula [1–3], where TP represents the positive samples correctly predicted by the model, TN the negative samples correctly predicted, FP the negative samples incorrectly predicted as positive,

**Table 1. Structure, activation function, and parameters of the LSTM model.**

Layer	Layer (type)	Activation function	Output size	Number of parameters	Total parameters
1	Lstm (LSTM)	Relu	(None, 128)	1348608	1359984
2	Dense1 (Dense)	Relu	(None, 64)	8256	
3	Dropout (Dropout)		(None, 64)	0	
4	Dense (Dense)	Softmax	(None, 48)	3120	

and FN the positive samples incorrectly predicted as negative.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

### Model evaluation metrics and analysis

The model’s predictive performance is evaluated using a detailed confusion matrix. This matrix classifies multi-class classification outcomes into categories of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), each quantified with absolute numerical values. A crucial metric for model assessment is the Comprehensive Recognition Rate, defined as the ratio of accurately classified samples to the total number of samples in the test dataset. Mathematically, this rate is expressed as:

Comprehensive Recognition Rate = Total Number of Test Samples/Number of Correctly Classified Samples.

### Availability of data and material

The datasets generated and/or analyzed during the current study are available in the SimTK repository ([https://simtk.org/plugins/datashare/?group\\_id=2836](https://simtk.org/plugins/datashare/?group_id=2836)). The 16S rRNA sequencing data are also available in the SimTK repository ([https://simtk.org/plugins/datashare/?group\\_id=2836](https://simtk.org/plugins/datashare/?group_id=2836)).

## RESULTS

### Strain MALDI-TOF MS spectrum database

The constructed database includes a total of 38,400 MALDI-TOF MS spectra, equally distributed among 48 distinct strains of *Escherichia coli*. Each strain contributes 800 individual spectra, providing

a balanced dataset for further analysis. Figure 1 shows the typical MALDI-TOF MS spectra for the 48 *Escherichia coli* strains. Significant peaks, specific to each strain, are primarily noted in the m/z range of 2000 to 10,000. Principal Component Analysis (PCA) is utilized for dimensionality reduction, as shown in Figure 2. Interestingly, the primary component for the spectra of strain LHL40080 (Strain No. 23) appears in the upper left quadrant of the scatter plot, demonstrating distinct data separability from the other 47 strains. However, the principal components for the spectra of the other 47 strains overlap significantly, making them challenging to distinguish. Upon closer examination, it was noted that the MALDI-TOF MS spectra for strain No. 23 differed markedly from those of other strains. This differentiation is presumed to result from subspecies-level variation in this particular strain. Further research is needed to verify this hypothesis.

### Model training and performance metrics

After completing 20 training epochs, as illustrated in Figure 3, our Long Short-Term Memory (LSTM) model demonstrated exemplary performance metrics, validating its effectiveness for the intended application. Specifically, the model recorded a remarkably low loss value of 0.0524, reflecting optimal minimization of prediction errors. This was paired with an exceptional accuracy rate of 0.999, highlighting the model’s near-flawless class label predictions. Additionally, a precision metric of 0.985 and a recall score of 0.982 together indicate the model’s high specificity and sensitivity, affirming its robustness in minimizing false positives and false negatives. These combined metrics underscore the model’s overall predictive prowess and reliability.

### Model evaluation

The confusion matrix presented in Figure 4 provides crucial insights into classification discrepancies among various strains. For example, strain No. 5 was predominantly misclassified as strain No. 18 at a 71% rate (284 out of 400 samples), strain No. 14 was misidentified as strain No. 38 in 25.75% of cases (103 out of 400 samples), and 31.25% of samples from

strain No. 34 were misclassified as strain No. 37 (125 out of 400 samples). Despite these specific instances of misclassification, the model exhibited robust performance for the remaining strains, achieving an identification accuracy exceeding 90%. The overall identification accuracy across all 48 strains reached a commendable 92.24%.

## DISCUSSION

In the realm of microbial diagnostics, traditional MALDI-TOF MS technology has proven to be a rapid and accurate tool by generating specific bacterial fingerprint spectra through the analysis of microbial cellular proteins and peptides [14, 15]. However, this

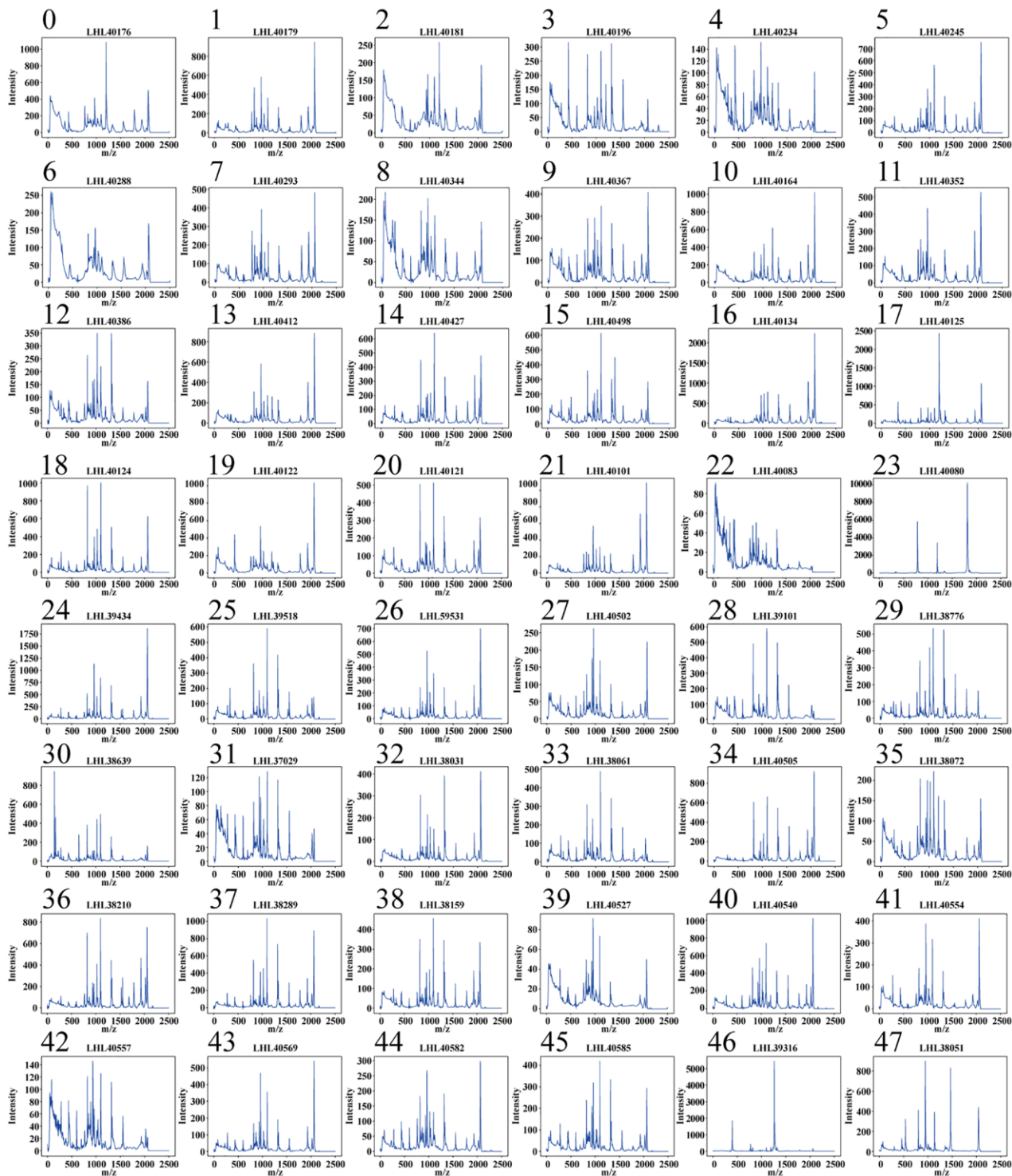
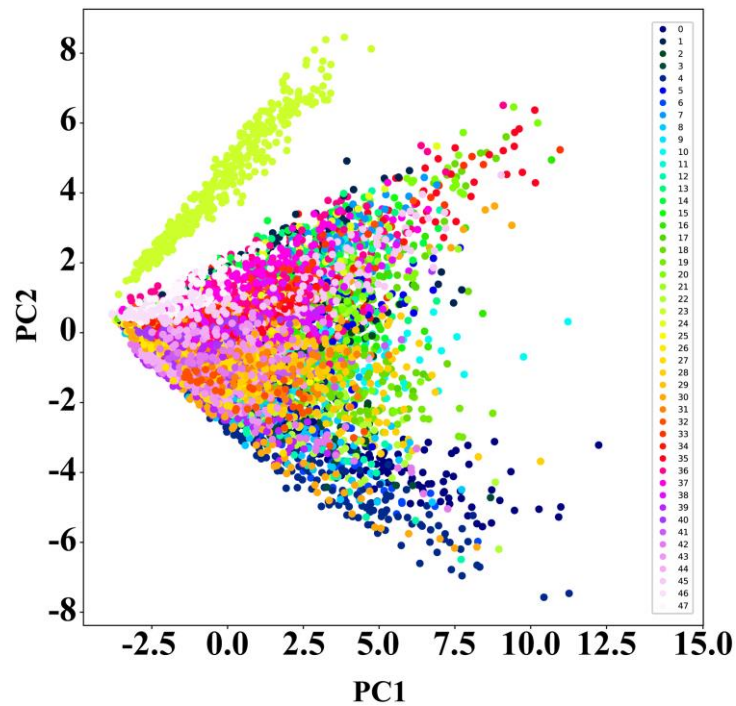
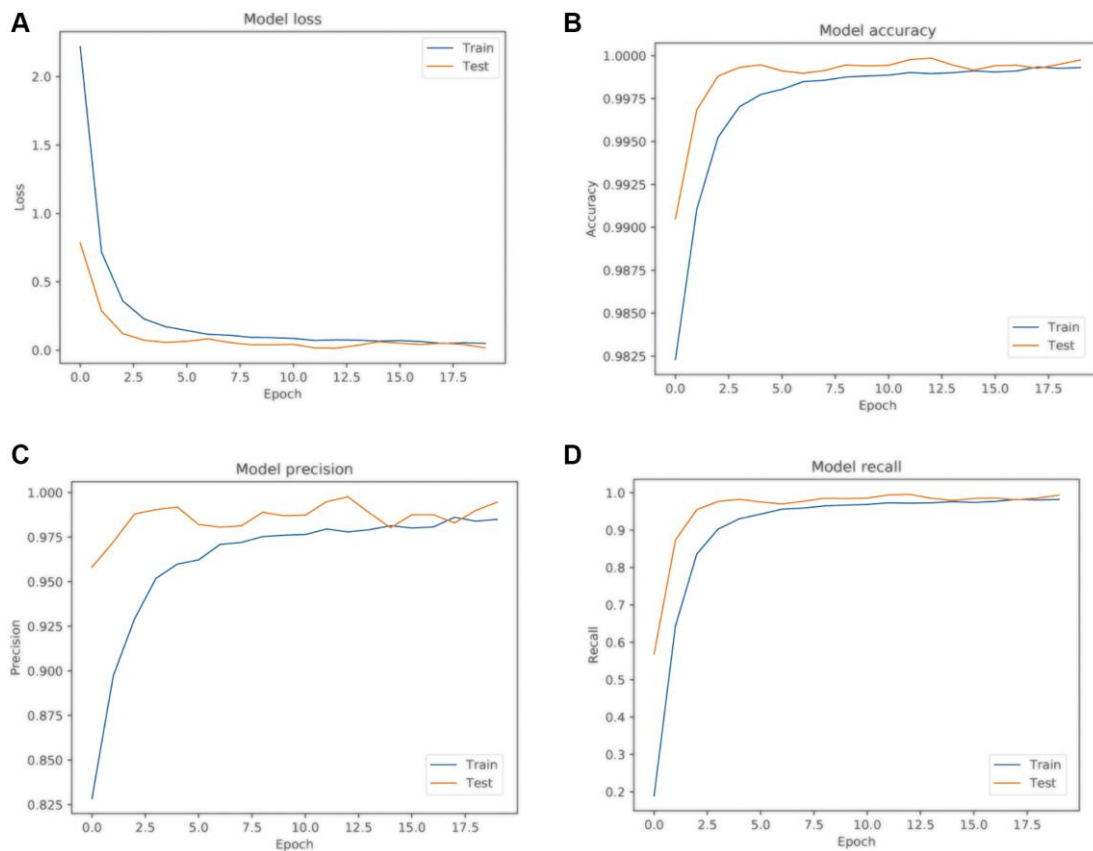


Figure 1. Typical MALDI-TOF MS patterns for 48 *Escherichia coli* strains.



**Figure 2. PCA dimensionality reduction results for *Escherichia coli* strain MALDI-TOF MS spectra.** The numbers 0–47 represent the 48 *Escherichia coli* strains used in the study. The isolated cluster of points in the upper left corner corresponds to strain No. 23 LHL40080.



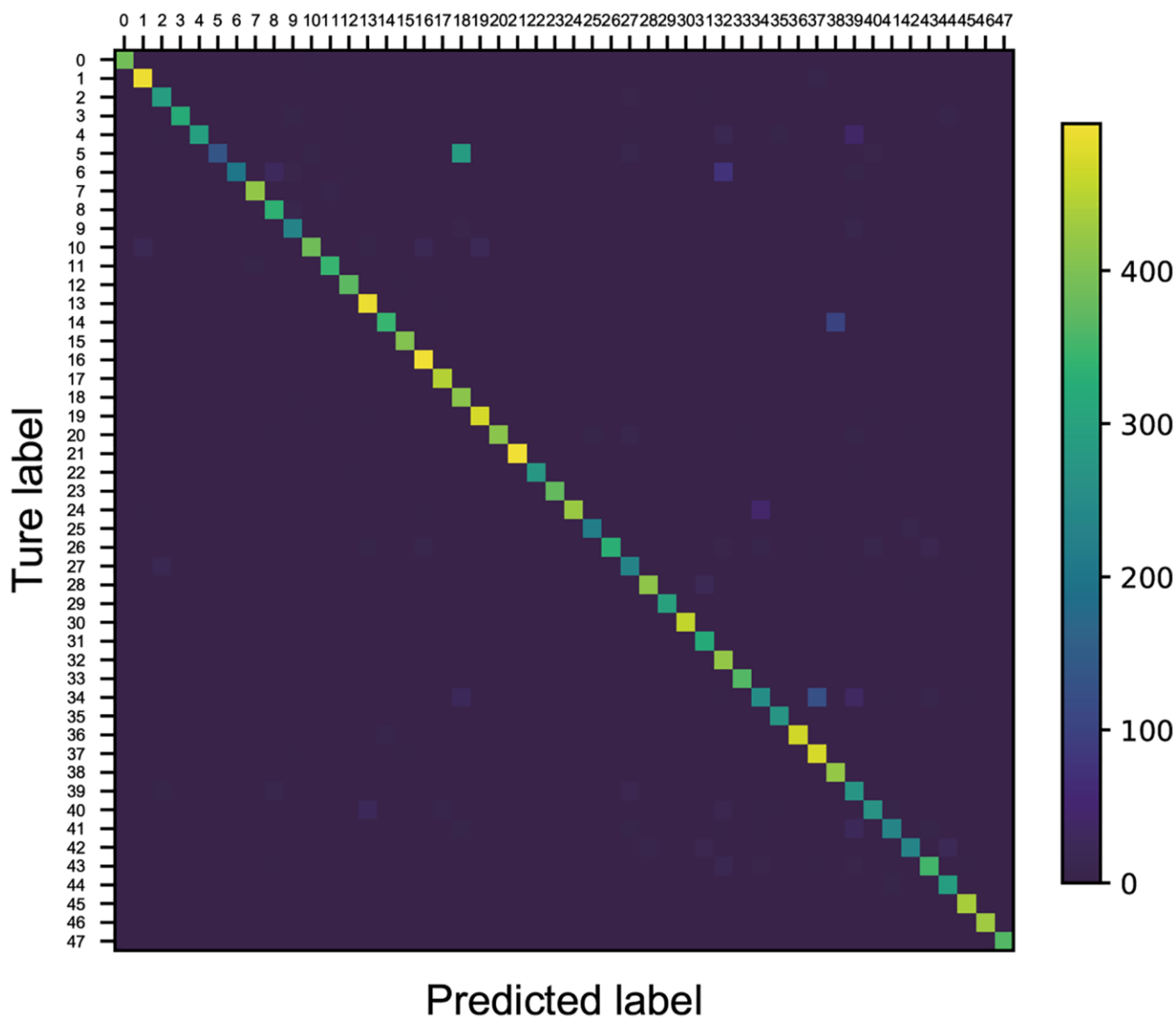
**Figure 3. LSTM model training results.** (A) Model loss curve; (B) Model accuracy curve; (C) Model precision curve; (D) Model recall curve. Blue represents the training sample curve, and yellow represents the test sample curve.

method encounters challenges in differentiating at the subspecies level or among similar microorganisms [16]. Traditional algorithms, while providing statistical validity to matching scores using probabilistic frameworks, are limited in distinguishing closely related strains due to the randomness in the MALDI-TOF MS sampling process [17, 18].

In our newly developed method, by incorporating LSTM neural networks, we are able to overcome these limitations. The unique architecture of LSTMs enhances control over information flow and improves data processing capabilities, particularly in handling long-term dependencies related to time series [19]. The LSTM networks can perform finer analysis of subtle differences within complex biological samples, identifying specific spectral peaks or patterns associated

with virulence factors, toxins, or other biomarkers. This ability is crucial for differentiating microbes that have minor variations in their biomarkers [20].

Additionally, compared to traditional methods, the combined MALDI-TOF MS and LSTM approach is more efficient in handling large datasets, as LSTM networks are designed to manage extensive datasets, offering quicker processing times and more efficient data handling than conventional statistical methods. This is particularly useful for the voluminous data often generated by MALDI-TOF MS. The attributes of LSTM networks make them an ideal choice for predictive modeling in the complex biological systems analyzed by MALDI-TOF MS, leading to the development of superior prognostic and diagnostic tools.



**Figure 4. Confusion matrix for model evaluation.** The numbers 0–47 refer to the 48 *Escherichia coli* strains used in the study.

In this study, 800 maps of each strain were gathered in batches to form a training dataset, preserving the unique characteristics of all maps. Utilizing an LSTM neural network, feature extraction and classification training were conducted on 2505-dimensional atlas data [21, 22]. Retaining the feedback mechanism of the recurrent neural network (RNN), the LSTM enhances data processing by introducing gating units such as forgetting gates, input gates, and output gates. This approach enables better control over information transmission and addresses issues like RNN gradient disappearance and difficulty in capturing long-term sequence dependencies [20, 23–25]. Consequently, inputting the 2505-length map data into the LSTM neural network results in effective feature extraction and classification. Through training on 19,200 images, the LSTM model learned the characteristics of the MALDI-TOF MS spectra at the strain level, achieving the identification of 48 *Escherichia coli* strains with a comprehensive accuracy of 92.24%. These results demonstrate that MALDI-TOF MS provides high-resolution mass spectrometry data, enabling precise and accurate analysis of complex biological samples. When integrated with LSTM neural networks, this precision is further enhanced as the LSTM can efficiently process and interpret the intricate mass spectrometry data. MALDI-TOF MS often generates large datasets. LSTM neural networks are well-suited for managing such large datasets efficiently, offering quicker processing times and more efficient data handling than traditional statistical methods. LSTM networks' ability to learn from and remember long sequences makes them ideal for predictive modeling in complex biological systems analyzed by MALDI-TOF MS, leading to better prognostic and diagnostic tools.

While our study marks a significant advancement in the field of microbial diagnostics, we acknowledge certain limitations that must be considered. One of the most notable limitations is the absence of a direct experimental comparison between our novel method and existing microbial identification techniques. Such a comparison could have provided a more robust foundation for validating our approach. Additionally, the lack of a definitive determination method, like Multi-Locus Sequence Typing (MLST) or genome MLST (gMLST), to confirm whether certain strains, such as strain 23, are subspecies of *E. coli*, is a notable limitation of our study. Despite these constraints, the integration of LSTM neural networks with MALDI-TOF MS technology represents a significant leap forward in microbial diagnostics.

Moreover, our study delves deeper into the potential of this methodology in identifying specific markers for the

accurate discrimination of *E. coli* categories and its applicability in identifying pathogenic strains. This exploration not only highlights the novel contributions of the present study but also opens new avenues for future research in the field of microbial diagnostics. It suggests the possibility of developing more refined tools for microbial identification that could significantly impact clinical diagnostics and public health.

## AUTHOR CONTRIBUTIONS

MQQ, ZX and XF conceptualized the study. SYF, XZP and XY executed the experiments, analyzed the data, and prepared the initial draft of the manuscript. MQQ, SYF and XY ensured the authenticity of all raw data. ZX, XY, and XF critically reviewed and revised the manuscript. MQQ and XY oversaw the research. XF and ZX were responsible for grant proposal writing. All authors have read and approved the manuscript for publication.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

## ETHICAL STATEMENT AND CONSENT

In the present study, all patients provided written informed consent for the use of the specimens for research purposes. Studies were approved by and conducted in accordance with the University of Ningbo Medical Center Lihuili Hospital (KY2021SL237-01).

## FUNDING

This study was supported by the Zhejiang Natural Fund Project (LQ21H090001), the Zhejiang Public Welfare Technology Application Research Program (LGF19H030008) and the Public Welfare Project of Ningbo (2021S183) and Li Huili Hospital, Ningbo Medical Center, “Huili Fund” (2022ZD005).

## REFERENCES

1. Guobin H, Dandan L, Qiuyuan L, Jia Y, Qian L, Qingwei M, Liang Q. Exponential isothermal amplification coupled MALDI-TOF MS for microRNAs detection. *Chin Chem Lett.* 2023; 34:354–8. <https://doi.org/10.1016/j.cclet.2022.04.019>
2. Klyasova GAR, Malchikova AOR, Dzhulakyan ULR. Method For Identification of Bacteria From Positive Hemocultures By Matrix Laser Desorption Ionization Time-of-Flight Mass Spectrometry (Maldi-Tof Ms), In Patients With Bloodstream Infection. 2021.

3. Zengnan W, Ning X, Weiwei L, Jin-Ming L. A membrane separation technique for optimizing sample preparation of MALDI-TOF MS detection. *Chin Chem Lett.* 2019; 30:95–98.  
<https://doi.org/10.1016/j.cclet.2018.01.048>
4. Wang SS, Wang YJ, Zhang J, Xiang J, Sun TQ, Guo YL. Using MALDI-TOF MS coupled with a high-mass detector to directly analyze intact proteins in thyroid tissues. *Science China (Chemistry).* 2018; 61:871–8.  
<https://doi.org/10.1007/s11426-017-9230-1>
5. Sauget M, Valot B, Bertrand X, Hocquet D. Can MALDI-TOF Mass Spectrometry Reasonably Type Bacteria? *Trends Microbiol.* 2017; 25:447–55.  
<https://doi.org/10.1016/j.tim.2016.12.006>  
PMID:[28094091](https://pubmed.ncbi.nlm.nih.gov/28094091/)
6. De Cesare V, Davies P. High-Throughput MALDI-TOF Mass Spectrometry-Based Deubiquitylating Enzyme Assay for Drug Discovery. *Methods Mol Biol.* 2023; 2591:123–34.  
[https://doi.org/10.1007/978-1-0716-2803-4\\_8](https://doi.org/10.1007/978-1-0716-2803-4_8)  
PMID:[36350546](https://pubmed.ncbi.nlm.nih.gov/36350546/)
7. van Belkum A, Chatellier S, Girard V, Pincus D, Deol P, Dunne WM Jr. Progress in proteomics for clinical microbiology: MALDI-TOF MS for microbial species identification and more. *Expert Rev Proteomics.* 2015; 12:595–605.  
<https://doi.org/10.1586/14789450.2015.1091731>  
PMID:[26472137](https://pubmed.ncbi.nlm.nih.gov/26472137/)
8. Oviaño M, Rodríguez-Sánchez B. MALDI-TOF mass spectrometry in the 21st century clinical microbiology laboratory. *Enferm Infecc Microbiol Clin (Engl Ed).* 2021; 39:192–200.  
<https://doi.org/10.1016/j.eimc.2020.02.027>  
PMID:[32345492](https://pubmed.ncbi.nlm.nih.gov/32345492/)
9. Xu S, Li W, Zhu Y, Xu A. A novel hybrid model for six main pollutant concentrations forecasting based on improved LSTM neural networks. *Sci Rep.* 2022; 12:14434.  
<https://doi.org/10.1038/s41598-022-17754-3>  
PMID:[36002466](https://pubmed.ncbi.nlm.nih.gov/36002466/)
10. Füllsack M, Kapeller M, Plakolb S, Jäger G. Training LSTM-neural networks on early warning signals of declining cooperation in simulated repeated public good games. *MethodsX.* 2020; 7:100920.  
<https://doi.org/10.1016/j.mex.2020.100920>  
PMID:[32509538](https://pubmed.ncbi.nlm.nih.gov/32509538/)
11. Cheng N, Kuo A. Using Long Short-Term Memory (LSTM) Neural Networks to Predict Emergency Department Wait Time. *Stud Health Technol Inform.* 2020; 272:199–202.  
<https://doi.org/10.3233/SHTI200528>  
PMID:[32604635](https://pubmed.ncbi.nlm.nih.gov/32604635/)
12. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nat Clin Pract Oncol.* 2008; 5:588–99.  
<https://doi.org/10.1038/nponc1187>  
PMID:[18695711](https://pubmed.ncbi.nlm.nih.gov/18695711/)
13. Tuaevea NO, Falzone L, Porozov YB, Nosyrev AE, Trukhan VM, Kovatsi L, Spandidos DA, Drakoulis N, Kalogeraki A, Mamoulakis C, Tzanakakis G, Libra M, Tsatsakis A. Translational Application of Circulating DNA in Oncology: Review of the Last Decades Achievements. *Cells.* 2019; 8:1251.  
<https://doi.org/10.3390/cells8101251>  
PMID:[31615102](https://pubmed.ncbi.nlm.nih.gov/31615102/)
14. Jang KS, Kim YH. Rapid and robust MALDI-TOF MS techniques for microbial identification: a brief overview of their diverse applications. *J Microbiol.* 2018; 56:209–16.  
<https://doi.org/10.1007/s12275-018-7457-0>  
PMID:[29492868](https://pubmed.ncbi.nlm.nih.gov/29492868/)
15. Dingle TC, Butler-Wu SM. Maldi-tof mass spectrometry for microorganism identification. *Clin Lab Med.* 2013; 33:589–609.  
<https://doi.org/10.1016/j.cll.2013.03.001>  
PMID:[23931840](https://pubmed.ncbi.nlm.nih.gov/23931840/)
16. Weis CV, Jutzeler CR, Borgwardt K. Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. *Clin Microbiol Infect.* 2020; 26:1310–7.  
<https://doi.org/10.1016/j.cmi.2020.03.014>  
PMID:[32217160](https://pubmed.ncbi.nlm.nih.gov/32217160/)
17. Mortier T, Wieme AD, Vandamme P, Waegeman W. Bacterial species identification using MALDI-TOF mass spectrometry and machine learning techniques: A large-scale benchmarking study. *Comput Struct Biotechnol J.* 2021; 19:6157–68.  
<https://doi.org/10.1016/j.csbj.2021.11.004>  
PMID:[34938408](https://pubmed.ncbi.nlm.nih.gov/34938408/)
18. Yang Y, Lin Y, Qiao L. Direct MALDI-TOF MS Identification of Bacterial Mixtures. *Anal Chem.* 2018; 90:10400–8.  
<https://doi.org/10.1021/acs.analchem.8b02258>  
PMID:[30091898](https://pubmed.ncbi.nlm.nih.gov/30091898/)
19. Keya AJ, Shajeeb HH, Rahman MS, Mridha MF. FakeStack: Hierarchical Tri-BERT-CNN-LSTM stacked model for effective fake news detection. *PLoS One.* 2023; 18:e0294701.  
<https://doi.org/10.1371/journal.pone.0294701>  
PMID:[38039283](https://pubmed.ncbi.nlm.nih.gov/38039283/)
20. Ghislieri M, Cerone GL, Knafnitz M, Agostini V. Long short-term memory (LSTM) recurrent neural network for muscle activity detection. *J Neuroeng Rehabil.* 2021; 18:153.



- <https://doi.org/10.1186/s12984-021-00945-w>  
PMID:[34674720](https://pubmed.ncbi.nlm.nih.gov/34674720/)
21. Nikkonen S, Korkalainen H, Leino A, Myllymaa S, Duce B, Leppanen T, Toyras J. Automatic Respiratory Event Scoring in Obstructive Sleep Apnea Using a Long Short-Term Memory Neural Network. *IEEE J Biomed Health Inform.* 2021; 25:2917–27.  
<https://doi.org/10.1109/JBHI.2021.3064694>  
PMID:[33687851](https://pubmed.ncbi.nlm.nih.gov/33687851/)
22. Zhao Y, Liu Y. OCLSTM: Optimized convolutional and long short-term memory neural network model for protein secondary structure prediction. *PLoS One.* 2021; 16:e0245982.  
<https://doi.org/10.1371/journal.pone.0245982>  
PMID:[33534819](https://pubmed.ncbi.nlm.nih.gov/33534819/)
23. Zhao J, Deng F, Cai Y, Chen J. Long short-term memory - Fully connected (LSTM-FC) neural network for PM<sub>2.5</sub> concentration prediction. *Chemosphere.* 2019; 220:486–92.  
<https://doi.org/10.1016/j.chemosphere.2018.12.128>  
PMID:[30594800](https://pubmed.ncbi.nlm.nih.gov/30594800/)
24. Gong CA, Su CS, Chao KW, Chao YC, Su CK, Chiu WH. Exploiting deep neural network and long short-term memory method-ologies in bioacoustic classification of LPC-based features. *PLoS One.* 2021; 16:e0259140.  
<https://doi.org/10.1371/journal.pone.0259140>  
PMID:[34941869](https://pubmed.ncbi.nlm.nih.gov/34941869/)
25. Zhang Z, Wang D, Harrington Pde B, Voorhees KJ, Rees J. Forward selection radial basis function networks applied to bacterial classification based on MALDI-TOF-MS. *Talanta.* 2004; 63:527–32.  
<https://doi.org/10.1016/j.talanta.2003.11.034>  
PMID:[18969464](https://pubmed.ncbi.nlm.nih.gov/18969464/)