frontiers | Research Topics

Artificial intelligence in forensic microbiology

Edited by
Chen Li, Yu-Dong Yao, Jiangwei Yan and Marcin Grzegorzek

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-2244-8
DOI 10.3389/978-2-8325-2244-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Artificial intelligence in forensic microbiology

Topic editors

Chen Li — Northeastern University, China

Yu-Dong Yao — Stevens Institute of Technology, United States

Jiangwei Yan — Shanxi Medical University, China

Marcin Grzegorzek — University of Lübeck, Germany

Citation

Li, C., Yao, Y.-D., Yan, J., Grzegorzek, M., eds. (2023). *Artificial intelligence in forensic microbiology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2244-8

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 11 April 2023
DOI 10.3389/fmicb.2023.1194390

OPEN ACCESS

EDITED AND REVIEWED BY
Matthias Hess,
University of California, Davis, United States

*CORRESPONDENCE
Chen Li
✉ lichen@bmie.neu.edu.cn

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 27 March 2023
ACCEPTED 27 March 2023
PUBLISHED 11 April 2023

CITATION
Li C (2023) Editorial: Artificial intelligence in forensic microbiology.
Front. Microbiol. 14:1194390.
doi: 10.3389/fmicb.2023.1194390

Editorial: Artificial intelligence in forensic microbiology

Chen Li*

Research Group for Microscopic Image and Medical Image Analysis, College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China

KEYWORDS

artificial intelligence, forensic microbiology, forensic science, data processing, multi-omics technique

Editorial on the Research Topic
Artificial intelligence in forensic microbiology

Identifying drowning and estimating the postmortem interval (PMI) have long been challenging problems in forensic medicine. Traditional examination methods involve observing physical signs, such as white foam in the nose or mouth, lung contraction or overinflation, pulmonary edema, and the presence of water in the stomach, to determine whether death was caused by drowning. Experimental examination generally involves measuring the content of diatoms in postmortem organ tissues. Determination of the time since death still relies mainly on various corpse signs and insect development identification methods. The specificity of various signs is not strong, and diatom testing may produce false positive and false negative results. In addition, these methods are limited by the observer's experience and environmental factors, and cannot meet the requirements of accurate forensic medicine. Accurately identifying drowning as the cause of death and determining the postmortem submersion interval (PMSI) is crucial in forensic science. However, there is a pressing need to develop exact methods and indicators to accomplish these objectives with greater accuracy and specificity. In recent years, microbial research has attracted much interest among forensic professionals. Integrating next-generation sequencing (NGS) with artificial intelligence algorithms has proven to be an effective method for analyzing changes in postmortem microbial communities (Wang Z. et al.). Therefore, our aim is to explore the potential of microbiology in forensic science by focusing on the application of artificial intelligence in forensic microbiology.

In the study conducted by the research team led by Zhao (Wang L. et al.; Zhang et al., 2022), mice were divided into two groups: drowning and post-mortem submersion. Tissue samples were collected at different intervals after death, including cecal contents, liver, brain, and water, which were then amplified and sequenced using the 16s rDNA method. The research results indicated that samples taken from the brain and liver between 5 to 14 days after death are optimal for analysis. Additionally, significant differences in the microbial communities were observed in the brain and liver samples. As the PMSI increased, the dissimilarity in microbial communities between the liver and brain samples of the drowning group and the post-mortem submersion group decreased. Therefore, this method cannot be deemed reliable for determining drowning. Accurate PMSI estimation models were developed for each organ based on their microbiota. The liver had a mean absolute error (MAE) of 1.282 ± 0.189 days, the brain had a MAE of 0.989 ± 0.237d, and the cecum had a MAE of 0.818 ± 0.165d.

Similarly, in Dmitrijs' study, it was demonstrated that microbial communities can be utilized to determine the PMSI in juvenile swine (Dmitrijs et al.). In Yu's study, it was shown that artificial intelligence is better at automatically identifying diatoms in drowning cases (Yu et al.). Pan's experiment proved that there are differences in bacterial communities among different water levels in the Yellow River Basin (Pan et al.).

With the continuous advancement of detection techniques and analytical methods, we can now investigate microbial communities within samples at previously unattainable depths. Microbial communities have the potential to serve as a powerful tool for estimating the time and identifying causes of death in animal models. However, microbial communities can vary significantly depending on different environments and conditions. To promote the widespread use of microbiomes in forensic science, more research professionals must collaborate and establish a comprehensive and systematic microbial database that can be integrated with data from other omics fields.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

We thank the support from National Natural Science Foundation of China (No. 82220108007) in this work.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Zhang, F., Wang, P., Zeng, K., Yuan, H., Wang, Z., Li, X., et al. (2022). Postmortem submersion interval estimation of cadavers recovered from freshwater based on gut microbial community succession. *Front. Microbiol.* 13, 988297. doi: 10.3389/fmicb.2022.988297

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 July 2022
doi: 10.3389/fmicb.2022.951707

Bacterial Succession in Microbial Biofilm as a Potential Indicator for Postmortem Submersion Interval Estimation

Finkelbergs Dmitrijs[1†], Juanjuan Guo[1,2,3†], Yecao Huang[4], Yafei Liu[4], Xinyue Fang[4], Kankan Jiang[4], Lagabaiyila Zha[1], Jifeng Cai[1] and Xiaoliang Fu[1,4*]*

[1] Department of Forensic Science, School of Basic Medical Sciences, Central South University, Changsha, China, [2] Department of Vascular Surgery, Shenzhen Second People's Hospital, The First Affiliated Hospital of Shenzhen University Health Science Center, Shenzhen, China, [3] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [4] Department of Forensic Medicine, School of Basic Medical Sciences and Forensic Medicine, Hangzhou Medical College, Hangzhou, China

Bacteria acts as the main decomposer during the process of biodegradation by microbial communities in the ecosystem. Numerous studies have revealed the bacterial succession patterns during carcass decomposition in the terrestrial setting. The machine learning algorithm-generated models based on such temporal succession patterns have been developed for the postmortem interval (PMI) estimation. However, the bacterial succession that occurs on decomposing carcasses in the aquatic environment is poorly understood. In the forensic practice, the postmortem submersion interval (PMSI), which approximately equals to the PMI in most of the common drowning cases, has long been problematic to determine. In the present study, bacterial successions in the epinecrotic biofilm samples collected from the decomposing swine cadavers submerged in water were analyzed by sequencing the variable region 4 (V4) of 16S rDNA. The succession patterns between the repeated experimental settings were repeatable. Using the machine learning algorithm for establishing random forest (RF) models, the microbial community succession patterns in the epinecrotic biofilm samples taken during the 56-day winter trial and 21-day summer trial were determined to be used as the PMSI predictors with the mean absolute error (MAE) of 17.87 ± 2.48 ADD (≈1.3 day) and 20.59 ± 4.89 ADD (≈0.7 day), respectively. Significant differences were observed between the seasons and between the substrates. The data presented in this research suggested that the influences of the environmental factors and the aquatic bacterioplankton on succession patterns of the biofilm bacteria were of great significance. The related mechanisms of such influence need to be further studied and clarified in depth to consider epinecrotic biofilm as a reliable predictor in the forensic investigations.

Keywords: postmortem interval, postmortem submersion interval, microbial biofilm, bacterial succession, machine learning algorithm

OPEN ACCESS

Edited by:
Chen Li,
Northeastern University, China

Reviewed by:
Fanglin Guan,
Xi'an Jiaotong University Health Science Center, China
Feng Chen,
Augusta University, United States
Jianding Cheng,
Sun Yat-sen University, China

****Correspondence:***
Xiaoliang Fu
fuxiaoliang8893@163.com
Jifeng Cai
cjf_jifeng@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Aquatic Microbiology, a section of the journal Frontiers in Microbiology

Received: *24 May 2022*
Accepted: *13 June 2022*
Published: *22 July 2022*

Citation:
Dmitrijs F, Guo J, Huang Y, Liu Y, Fang X, Jiang K, Zha L, Cai J and Fu X (2022) Bacterial Succession in Microbial Biofilm as a Potential Indicator for Postmortem Submersion Interval Estimation. Front. Microbiol. 13:951707. doi: 10.3389/fmicb.2022.951707

INTRODUCTION

During the forensic investigations, corpses can be found in a variety of natural or artificial aquatic environments, such as ponds, rivers, lakes, seas, and water storage containers. These cases may be suicides, homicides, or accidents, with the most common death cause being asphyxiation by drowning in the water. Due to the lower temperature

and the oxygen-deficit, the postmortem decomposition in the aquatic environments is usually slower than that on the land, which makes the forensic identification of corpses in the water particular. Postmortem submersion interval (PMSI), which is defined as the period between the entry into the water and recovery of the dead body, is approximately equal to the postmortem interval (PMI) in most of the common cases of drowning (Humphreys et al., 2013). It can be used for inferring the location of body falling into the water, delimiting the scope of the search for the suspect, and providing direction for the investigation. For corpses found on land, the PMI estimation relies on the temporal changes of the postmortem decomposition. However, determining PMSI has long been problematic because the postmortem decomposition in water is affected by a variety of biotic and abiotic factors, such as microbial metabolism, algal growth, and adipocere formation (Haefner et al., 2004; Pakosh and Rogers, 2009; Widya et al., 2012; Ueland et al., 2014). For example, corpses submerged in freshwater (rivers) may show more obvious decomposition changes than those submerged in saltwater (seas), as lower temperatures and higher salinity slow down bacterial activity (Byard, 2018). It is evident that microbial metabolism is the leading factor affacting the rate of postmortem decomposition happening in the terrestrial environment (Lauber et al., 2014; Metcalf et al., 2016). Based on this, several studies have monitored the bacterial community successions on the corpses found on land, and suggested that the postmortem changes in microbial communities were dramatic, measurable, and repeatable, allowing PMI to be estimated accurately even within a long time frame (Metcalf et al., 2013, 2016; Burcham et al., 2019). However, such estimation is more complicatied for carcasses found in water, especially for the cases of drowning. When drowning occurs, aquatic microorganisms will enter the respiratory and digestive tracts along with the liquid, changing the structure of endogenous microbial communities, consequently affecting their succession patterns. Moreover, if any unpredictable rupture occurs on the decomposing carcass, fluid from the environment could also enter the body and interrupt previous succession processes. All these complex scenarios are hard to evaluate in the forensic practices. As a result, to this date, limitied research has been done on the succession of the microbial communities on carcasses found in water.

When free-swimming aquatic microbes encounter solid surfaces, they can gradually switch from a planktonic lifestyle to forming biofilms that are wrapped by the extracellular polymeric substance (EPS) (Battin et al., 2007). EPS contributes to plankton attachment and provides independent niches specific to the particular aquatic environments. The human corpse appearing in the water provides attachment surface and organic nutrients for the aquatic microbes to form specific epinecrotic biofilm. A growing body of excellent studies has been done on biofilms in multiple disciplines, such as food hygiology (Liu et al., 2015), ecology (Ram et al., 2005), and medicine (Huang et al., 2011; Johnson et al., 2015), but there is limited data available associated with the human corpse decomposition in the water. The bacterial succession patterns of the epinecrotic biofilms have been studied by several forensic research groups using surrogate models of the vertebrate corpses. Dickson et al. were the first to evaluate the potential use of the bacterial succession for the PMSI estimation (Dickson et al., 2011). They have investigated the microbes involved in decomposition of porcine cadavers in the sea and confirmed that marine bacteria rapidly colonized the skin of the submerged cadavers in a succession manner. Using high-throughput sequencing (HTS), Benbow et al. have provided the first metagenomic data which has described aquatic bacteria succession patterns in the epinecrotic biofilms on the porcine remains submerged in a freshwater habitat (Benbow et al., 2015). The biofilm communities on the submerged remains changed significantly through the PMSI. Although the microbial community differences between the summer and winter trials have been observed, the succession patterns of dominant phyla and genera were similar. In recent years, machine learning (Li et al., 2019), and even deep learning algorithms (Rahaman et al., 2020), have been applied in the medical field on a large scale, as a result providing new tools for forensic research to explore microbial succession patterns. Using replicate swine carcasses, Kaszubinski et al. (2020) performed an experiment to describe the variation of epinecrotic biofilm microbiome in a non-flowing aquatic habitat. Through the sufficient sequencing data obtained, Kaszubinski et al. modeled key taxa for estimating PMSIs using the machine learning algorithm (random forest (RF) regression), correctly estimating the PMSI $\pm$ 3 days with $R^2 = 97.50\%$. These studies have provided compelling evidence that the bacterial succession in the epinecrotic biofilm has a prominent potential to be used for the PMSI estimation in forensic investigations of submerged corpses.

To further the understanding of the biofilm succession in our geographical region and its potential to be used as an indicator in the PMSI estimation during forensic investigations, present study used juvenile swine as proxies for human corpses to establish decomposition models for the use in a non-flowing aquatic environment. Sterile tiles were used as the inorganic solids placed in both close and distant proximity from the decomposing carcasses for the epilithic biofilm attachment. We employed HTS to fully describe the bacterial successions in the epinecrotic biofilms on the decomposing carcasses and epilithic biofilms on the submerged tiles. The objectives of the research were to (1) characterize the succession patterns of the bacterial communities in the epinecrotic and epilithic biofilms; (2) assess the repeatability of the succession patterns in the replicable settings; (3) compare the differences in succession patterns between the summer and winter trials; and (4) seek to provide important data for developing a machine learning algorithm to estimate PMSI.

MATERIALS AND METHODS

Study Design

This study was approved by the Medical Ethics Committee of Xiangya Hospital, Central South University (approval number: 201503465) and followed all applicable institutional and national guidelines for the care and use of animals. A total of two drowning experiments were, respectively, conducted from 31 October to 26 December 2017 (winter trial) and 21 July to

11 August 2018 (summer trial) in three adjacent freshwater ponds (28°34′74.6″ N, 112°81′95.5″ E) within the Xiangjiang River watershed in Changsha, China (**Figure 1**). Each of the ponds was about 1,000–1,500 m^2 and 5 m deep, subjected to direct sunlight (no canopy), and surrounded by some shrubs. The water in the ponds was previously pumped from a vast lake nearby.

Swine carcasses have been frequently used as proxies for human corpses in the forensic research (Schoenly et al., 2007). In our experiments, six female swines (Sus scrofa demesticus, $n = 3$ per trial), each weighing 7.05–11.5 kg, were purchased from a local farm and killed by drowning after anesthesia. Carcasses were individually placed on a fine mesh nylon pad (60 mesh/inch) inside the plastic cages ($0.75 \times 0.55 \times 0.25$ m^3) to facilitate weighing as the carcass disarticulated and to prevent removal of the carcasses by scavengers (e.g., fishes, shrimps, and crabs) (**Figure 1**). Moreover, bricks were attached to the bottom of the cages to prevent the carcasses emerging from the water during the bloated stage, consequently inhibiting colonization by terrestrial insects. Through these measures, the microbial biofilms could form naturally on the carcass surfaces without the random interference from aquatic and terrestrial scavengers. Each cage was placed in a single pond and about 1 m below the water surface. In order to explore the influence of carcass decomposition on the formation of epilithic biofilm, synchronous sampling of 2 tiles which were placed in the pond 0.5 m away from oppsite sides of each carcass was conducted. About 100 m from the location of sunken carcasses, another two sterile tiles were placed 1 m below the water surface in the nearby lake during the summer trial as a negative control. In the process of decomposition, the postmortem changes were recorded and photographed daily. To visually describe the decomposition processes of the carcasses, the duration of each decay stage was assessed using the framework of Zimmerman and Wallace (2008) and Wallace et al. (2008). The visual body score was evaluated using the total aquatic decomposition scoring (TADS) system as described by van Daalen et al. (2017).

Environmental Parameters and Sample Collection

Air temperature and relative humidity were recorded hourly by a data logger (MEACON Automation Technology Co., Ltd., Hangzhou, China). Water quality parameters, including water temperature (°C), dissolved oxygen (DO, mg/L), pH, conductivity (mS/cm), and salinity (ppt), were measured 4 times every day at 1 m beside each carcass and at the location for the negative control using the AZ 86031 Water Quality Checker (AZ Instrument Co., Taiwan, China), with its probe been placed 1 m below the water surface. Water temperature data recorded every 6 h during each of the experiment days was used to calculate the average daily water temperature, and accumulated degree days (ADDs) were calculated by summing the average daily water temperature above the lower development threshold (LDT) (Mateus and Vieira, 2014). Because ADD was used to explore how microorganism communities (not insects) changed during the study, an LDT of 0°C was employed for this calculation, according to Pechal et al. (2014).

A total of three types of microbial specimens were collected: epinecrotic, epilithic, and aquatic samples. Swine skin has been widely used as a proxy for human skin (Sekkat et al., 2002). Epinecrotic samples were collected from the skins of swine carcasses at multiple time points: 10 min before being sacrificed, immediately after being sacrificed, daily in the first week, and then weekly in the following days during the summer trial. Due to the carcass decomposition progressing slowly at low temperatures, the sampling time points after the swine sacrifice for the winter trial were different: every 3 days in the first week, and then every 2 week in the following days. At each sampling time point, a 10×10 cm^2 skin area on one side of the torso of each carcass was gently swabbed for 60 s using sterile cotton applicators with care taken to not duplicate any previously sampled area. Then, the tip of each applicator was cut off with a pair of sterile scissors and placed in a 1.5-ml microcentrifuge tube. Epilithic and aquatic samples were collected from the tiles and ponds weekly during the summer trial and biweekly during the winter trial. The epilithic communities on the surfaces of each tile were sampled in the same way as the epinecrotic samples. For sampling the aquatic communities, one liter of pond water was collected at 1 m distance from each carcass using the sterile syringes, consequently filtered by suction filtration using a Buchner funnel and the nylon filter membranes (pore diameter 0.2 μm). The negative control of the epilithic and aquatic samples were taken further away from the carcasses from the nearby lake during the summer trial. All samples were immediately frozen at −80°C until further processing. Carcasses and tiles were immediately submerged back into the water to their original location upon sample collection completion, while also making sure that the disposable sterile gloves used in the procedure were replaced after each sampling. The detailed information of each sample is shown in **Supplementary Table S1**.

DNA Extraction and PCR Amplification

The cotton tip of each swab or the fragment cut from each filter membrane was put in a bead tube for genome DNA extraction. The succeeding steps were performed as dictated by the manufacturer's specifications of MoBio PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). Swabs and filter membranes that had not been used for sampling were used as the blank controls. DNA concentration and purity were determined *via* Nanodrop ultraviolet spectrophotometric detection and 0.8% agarose gel electrophoresis (**Supplementary Table S1**). All the qualified DNA samples were diluted to a 20 ng/μl working stock using sterile ultrapure water. When insufficient, the stock solution was used directly.

The V4 region of the 16S rDNA gene has been amplified using dual-indexed primers (515F/806R: 5′-GTGYCAGCMGCCGCGGTAA-3′; 5′-GGACTACNVGGGTWTCTAAT-3′) as described previously (Claesson et al., 2010). Each of the forward primers contained a 6 bp barcode unique to each sample. All polymerase chain reactions (PCRs) were conducted in 25 μl reaction volumes containing 5 μl of 5× reaction buffer, 5 μl 5× GC buffer, 2 μl

FIGURE 1 | The geographic position and experimental scenes. Experiments were conducted in the 3 adjacent freshwater ponds and a nearby vast lake in Changsha, China. The white letters labeled on the satellite photo represent the experimental sites for the winter trail (F, G, & H), summer (K, L, & M) trial and the negative control (N). Carcasses were individually placed inside the plastic cages and then sunk into the water. The bottom pictures are, respectively, taken from the time immediately after death, submerged fresh, early floating, floating decay, advanced floating decay, and sunken remains stages of the carcasses in the winter (above) and summer (below) trials. The time points were labeled on each picture.

dNTP (2.5 mM), 1 μl forward and reverse primers (10 μm), 2 μl DNA template, 8.75 μl ddH_2O, and 0.25 μl Q5 DNA polymerase. Thermal cycling conditions were following: initial denaturation at 98°C for 2 min, followed by 30 cycles of denaturation at 98°C for 15 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, with a final extension at 72°C for 5 min. The resulting amplicons were mixed with the same volume of 1× loading buffer (containing SYB green) and detected using electrophoresis on 2% agarose gels. Except for the blank controls which failed amplification, all amplicons were in the size range of 200–300 bp and excised from the agarose gel for further experiments.

Amplicon Sequencing and Data Analysis

All amplicons in the size range of 200–300 bp were purified using the AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Santa Clara, CA, USA) and pooled into equal concentrations. Sequencing libraries were generated using the TruSeq® Nano DNA LT Library Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's recommendations, and index codes were added. The library quality was assessed on the Agilent Bioanalyzer 2100 system using Agilent High Sensitivity DNA Kit (Agilent, Santa Clara, CA, USA) and the Promega QuantiFluor Fluorometer using Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific, Carlsbad, CA, USA). Finally, the library was sequenced on an Illumina MiSeq platform, which generated 300 bp paired-end reads.

After sequencing, the paired-end reads were assigned to the samples based on their unique barcode, truncated by cutting off the barcode and primer sequence, and merged using FLASH (version 1.2.7) (Magoč and Salzberg, 2011). The merged reads containing ambiguous bases (*N*) or low-quality bases were filtered out using QIIME filter (version 1.8.0) (Bokulich et al., 2013), and chimeras were removed using USEARCH (version 5.2.236) (Edgar et al., 2011). After the removal of singleton sequences, operational taxonomic units (OTUs) were classified with the threshold of 97% similarity using the UCLUST in QIIME (Edgar, 2010). A representative sequence was picked by selecting the longest sequence that had the largest hit number to other sequences in each OTU. Representative sequences of 16S OTUs were, respectively, aligned and annotated using the Greengenes database (Release 13.8, http://greengenes.secondgenome.com/) (DeSantis et al., 2006). Total raw sequencing data was published in the Sequence Read Archive (SRA) under the accession number PRJNA841063.

Statistical Analysis

To avoid biases of biodiversity data generated by the number of sequences, the data were rarefied to 90% of the minimum library size. Statistical analysis was conducted using a web-based tool, MicrobiomeAnalyst (Dhariwal et al., 2017). For the hierarchical cluster analysis (HCA), each OTU began as a separate cluster, then the clustering algorithm proceeded to combine them until all OTUs belonged to a single cluster. Distances between OTUs were measured with Minkowski and clustering algorithms using the average linkage (the distance between two clusters is the average of the distances between all the points in those clusters). The results were visualized as a heat map to show the temporal changes in the taxonomic clusters. The alpha diversity indices of Chao1 and Shannon were calculated to evaluate the species richness and evenness of a sample. The statistical significance of differences in alpha diversity indices between experimental groups/decay stages was estimated using the analysis of variance (ANOVA). Principal coordinate analysis (PCoA) using weighted UniFrac dissimilarities was performed to visualize the changes in the community beta diversity according to the decomposition progress/experimental grouping. The statistical significance of the clustering pattern in the ordination plot was evaluated by permutational multivariate analysis of variance (PERMANOVA). The taxonomic composition of each sample was visualized in a stacked bar plot by the chronological order. Linear discriminant analysis effect size (LEfSe) was used to identify taxa with significantly differential abundance across experimental groups. Random forest (RF) analysis, which is a powerful machine learning algorithm for the identification of the predictive biomarkers and establishment of the prediction the regression of model, has been employed for the regressing of the OTU relative abundances against the ADDs using the "randomForest" R package. The relative abundance of an OTU is its percentage in the total amount of sequences in a sample. The OTUs were ranked in the order of their feature importance and selected to generate predictive biomarker sets. The RF regression models were futher established for predicting the PMSI based on generated biomarker sets were further established. The mean absolute error (MAE) and goodness of fit (R^2) were used to evaluate the performance of the models (Metcalf et al., 2013).

RESULTS

Progression of Carcasses Decomposition

During the summer trial, the average air temperature and relative humidity were 34.40 ± 1.98°C and 34.19 ± 3.78%, respectively. The carcasses decomposed much faster than those during the winter trail and progressed to the stage of sunken remains in one week time (**Figure 1**). Sample collection was thus conducted on the bone remains once a week for the following 2 weeks. By visually describing the decomposition phenomenon progression of each carcass, decay stages were determined as follows: submerged fresh stage began at 0.0 ± 0.0 day, early floating stage at 1.0 ± 0.0 days, floating decay stage at 2.8 ± 0.8 days, advanced floating decay stage at 4.0 ± 0.0 days, and sunken remains stage at 6.0 ± 0.0 days. During the 8-week winter trial, the average air temperature and relative humidity were 15.00 ± 4.84°C and 81.50 ± 14.82%, respectively. The carcasses decomposed slowly and entered sunken remains stage in a gradual manner. The submerged fresh stage began at 0.0 ± 0.0 day, early floating stage at 4.0 ± 1.0 days, floating decay stage at 10.0 ± 0.0 days, advanced floating decay stage at 17.3 ± 2.3 days, and sunken remains stage at 47.3 ± 7.5 days.

Water quality parameters were measured at the locations of sunken carcasses and the negative control tiles. The changes of calculated average water temperature, DO, pH, conductivity, and salinity are shown in **Supplementary Figure S1**. The water temperature was mainly affected by the climate and remained relatively constant in the summer but dropped gradually in the

winter. ADDs increased linearly during both seasons. DO was mainly affected by the weather and dropped markedly on the rainy and cloudy days, most likely due to the photosynthesis decline of the underwater plants rather than the carcass decomposition influence. The fluctuations of water DO, pH, conductivity, and salinity in the summer were greater than those during the winter. However, there was no significant difference of water quality between the locations whether the carcasses had presented. These measurements of water quality indicated that the carcasses which would randomly appear in the pond would not cause serious changes to an aquatic environment with a large storage capacity. This also indicates that the water in these conditions can provide a relatively stable physicochemical environment for the drowned carcasses.

The body weight and TADS score changes during the decomposition processes are shown in **Supplementary Figure S2**. The loss of body weight and the increase of TADS scores in each trial were continued. However, the carcasses that decomposed in the summer progressed to the sunken remains stage more rapidly and had made their weight and TADS scores reach the plateau much earlier. In fact, the hyper slow decaying carcasses during the winter trial remained amounts of soft tissue even after 56-day decomposition. It is difficult to accurately describe such a slow process in detail using the protocols described by van Daalen et al. (2017), which limits the TADS system use for the accurate PMSI estimation for the winter cases. On the other hand, due to the impact of the water flow, the linear relationships between the body weight and the ADD in both trials were not significant. These results suggested that it is difficult to determine temporal changes precisely through the decay phenomena and weight loss.

Overview of Sequencing

A total of 131 samples had been collected, including 66 in the winter trial and 65 in the summer trial. The details of each sample are shown in **Supplementary Table S1**. After the DNA extraction and PCR amplification, a total of 5,816,237 high-quality reads were received from the output data of 16S amplicon sequencing, and 140,618 OTUs were obtained after these reads had been classified with 97% similarity. All the read lengths were distributed in 200–300 bp, which fits the size of the 16S V4 region (**Supplementary Figure S3**). Based on the sequencing data rarefying to 90% of the minimum library size, a total of 2,366,112 clean reads and 6,897 simplified OTUs were used for the following statistical analysis. The numbers of clean reads and simplified OTUs of each sample are shown in **Supplementary Table S1**. The average good's coverage value for all samples was 99.80 $\pm$ 0.14%, which suggested that the number of reads was enough to analyze biodiversity in all samples. The rarefaction curves indicated that the species richness in each sample had approached the plateau phase, and it was unlikely that more species would been detected with additional sequencing efforts (**Supplementary Figure S4**). After representative sequences of OTUs had been annotated in the Greengenes database, the taxon number at each taxonomic level of each sample type is shown in **Supplementary Table S2**. The taxa of epinecrotic, epilithic, and aquatic communities in the winter trial were significantly less than those in summer, especially the taxa at each taxonomic level in epinecrotic communities which were less than half of that in summer. These results suggested that the season could be the main factor affacting the aquatic environments, thus leading to the changes in the biofilm community structures on the carcasses in the water.

Community Differences Between Epinecrotic and Epilithic Biofilms

The calculated alpha diversity indices of Chao1 and Shannon for each sample are shown in **Supplementary Table S1**. The statistical differences in alpha diversity indices between the epinecrotic, epilithic, and aquatic communities were estimated using the ANOVA and reviewed as box plots (**Supplementary Figure S5**). During the winter trial, there were significant differences between the alpha diversity indices of the three communities. Each index of epinecrotic community was lower than that in the epilithic community, especially for the late decay periods. However, a significant difference between the communities was only be found in Chao1 indices for the summer trial, which indicated that there was a large difference in species richness, but a small difference in species evenness among the three communities in summer.

By measuring the weighted UniFrac distances, the beta diversity variation of the communities was visualized in PCoA plots and tested by PERMANOVA. As shown in **Figure 2A**, three types of samples were clustered into the significant clusters separately. The epilithic and aquatic community beta diversity was relatively stable during both seasons. Nevertheless, the epinecrotic communities showed a significant succession pattern through the time, which despite being initially close to the epilithic samples, subsequently separated from the other dots on the PCoA plots. Additionally, the communities present on the skeletal remains collected during the sunken remains stage (14 and 21 days) during the summer trial were also similar to the epilithic communities, indicating that the bones' adsorbability to the planktonic microorganisms was similar to that of the inorganic solid surface. **Supplementary Figure S6** shows that there was a significant difference between the epilithic and aquatic community structures regardless of the presence of the decomposing carcasses. The epilithic samples from both the summer trial and negative control clustered together, while, aquatic samples formed into another cluster. Similar to the measurements of the water quality, these results suggested that the carcasses decomposing in the water had little impact on the aquatic environments and nearby epilithic biofilms.

According to the LEfSe analysis of the winter trial data, the phyla, such as Cyanobacteria, Actinobacteria, Planctomycetes, Fusobacteria, Chloroflexi, and Verrucomicrobia in the epinecrotic communities have significantly less presence than in the epilithic communities, whereas Firmicutes and Bacteroidetes were more present in the epinecrotic communities. As for the summer trial data analysis, Actinobacteria, Chloroflexi, Planctomycetes, and Verrucomicrobia in the epinecrotic communities were significantly less present, whereas Firmicutes were more present in the epinecrotic communities than in the

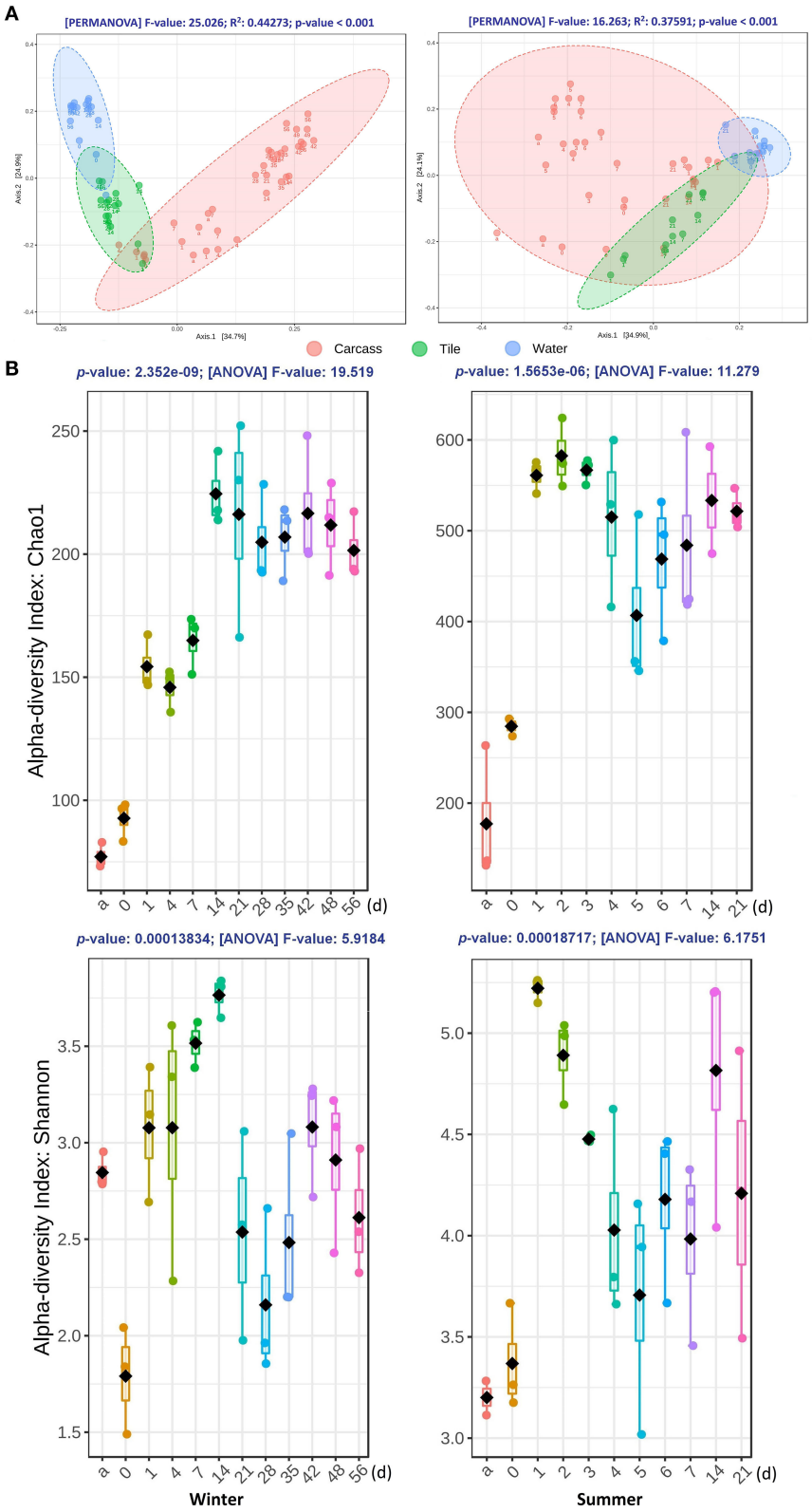


FIGURE 2 | The diversity variations of the bacterial communities during the winter (left panel) and summer (right panel) trials. **(A)** Two-dimensional PCoA plots of weighted UniFrac distance matrices for samples obtained from carcasses, tiles, and water at different time points during both seasons. Samples obtained from

(Continued)

FIGURE 2 | different substrates are presented in different colors. Sampling time points are marked under each dot. The statistical significance of the clustering pattern in each plot was evaluated with PERMANOVA (top of each plot). **(B)** The alpha diversity variation of epinecrotic communities during the both seasons. Samples from each time point were measured with Chao1 and Shannon indices, as shown in each box plot. The *X*-axis of each plot are the sampling days and "a" represents the antemortem. The ANOVA results are showing at the top of each plot.

epilithic communities. The main differentiated phyla between the epinecrotic and epilithic communities were similar during the two seasons (**Supplementary Table S3**). The differentiated genera of the two communities during two seasons are shown in **Supplementary Table S4**.

Epinecrotic Community Temporal Variations During Carcasses Decomposition

The alpha diversity of epinecrotic communities changed significantly during both seasons (**Figure 2B**). For the winter trial, the Chao1 indices had increased after the carcasses entered the water (1 day) and during the floating decay stage (14–56 days), however, the Shannon indices had decreased immediately after death (0 day) and then increased after the carcasses entered the water (1–21 days). The alpha diversity for the summer trial followed a similar pattern. The Chao1 indices had increased after the carcasses entered the water (1 day) and only slightly reduced in the advanced floating decay stage (5–7 days). The Shannon indices had increased first after the carcasses entered the water (1 day), then decreased to a plateau (2–7 days), and finally slightly elevated during the sunken remains stage (14–21 days).

On the PCoA plot for the epinecrotic communities of both seasons (**Supplementary Figure S7**), all the dots of the summer samples clustered with the winter samples which were obtained before the 14 days of the decomposition. The dots of 21–56 days for the winter trial formed another cluster and separated from the previous samples. **Supplementary Figure S8** illustrates that the seasonal factors had altered the beta diversity of aquatic communities. These results indicated that the beta diversity variation in epinecrotic communities during the winter was much more significant than that in summer, which could be attributed to the specific aquatic community structure and longer decomposition processes during the winter trial.

The temporal successions of the epinecrotic communities are shown in the heat maps by hierarchically clustering OTUs with Minkowski distance matrices (**Figure 3A**). Remarkable and repeatable succession patterns have been observed in the epinecrotic biofilms. Significant clusters could been found in the heat map for the winter trial. The features changed significantly across the time points of 10 min before the sacrifice, immediately after the sacrifice, and during the subsequent decay process in the water, especially during the 2–3 weeks when the carcasses were in the advanced floating decay stage. For the summer trial, the epinecrotic community structure changed significantly on the first day in the water and then showed a sequential succession pattern during the subsequent decay process (2–7 days). Additionally, the main classes of the 14 and 21 days which are specific to the communities on the bone remains were similar to the cluster which represented the early decay stage. In the epilithic and aquatic communities, the clustered 16S OTUs with high abundance changed significantly during the winter trial. However, both communities in summer were short of variable OTUs, and performed more like the macro-succession pattern (**Supplementary Figure S9**).

Taxa annotation and abundance variation for chronological samples are shown as histograms in **Figure 3B**. For the winter trial, the phyla of Proteobacteria (79.04 ± 11.09%) and Firmicutes (18.91 ± 10.44%) have been the dominant taxa before the body have entered the water (antemortem and immediately after being sacrificed). During the subsequent decay process in the water, Proteobacteria was still dominant (72.83 ± 17.04%, 1–56 days), whereas Firmicutes had decreased through the submerged fresh stage to the floating decay stage (4.44 ± 5.59%, 0–7 days) and then followed to slowly increase through the advanced floating decay stage to sunken remains stage (15.55 ± 10.35%, 14–56 days). In addition, Bacteroidetes which represented a minor part of the antemortem communities (1.26 ± 0.81%) had increased gradually through the submerged fresh stage to the floating decay stage (16.20 ± 12.61%, 1–14 days). At the advanced submerged decay stage, Bacteroidetes had first decreased (1.54 ± 1.53%, 21–28 days) and then increased up until the sunken remains stage (19.18 ± 9.39%, 35–56 days). Other phyla, such as Cyanobacteria, Acidobacteria, and Fusobacteria were scarce and occurred opportunistically during the winter trial. For the summer trial, the phyla of Proteobacteria (57.95 ± 15.39%) and Firmicutes (28.93 ± 17.03%) were the dominant taxa before the body entered the water. During the subsequent decay process in the water, Proteobacteria was still dominant (53.73 ± 15.62%, 1–21 days), whereas Firmicutes had decreased in the early floating stage (4.54 ± 1.09%, 1–2 days), and then became substantially more abundant during the floating decay stage and advanced floating decay stage (45.56 ± 16.80%, 3–7 days). On the bone remains (14 and 22 days), Firmicutes only accounted for 4.92 ± 1.33% of the communities. Other phyla, such as Acidobacteria, Bacteroidetes, and Chloroflexi, which were abundant right after the carcasses have entered the water diminished during the decay process, but bacame abundant on the bone remains. The variations of the top 40 genera of the epinecrotic communities are shown in **Supplementary Figure S10**.

PMSI Estimation Models Based on the Epinecrotic Biofilm Succession

We regressed the relative abundances of total OTUs against ADDs using the RF machine learning algorithm to establish

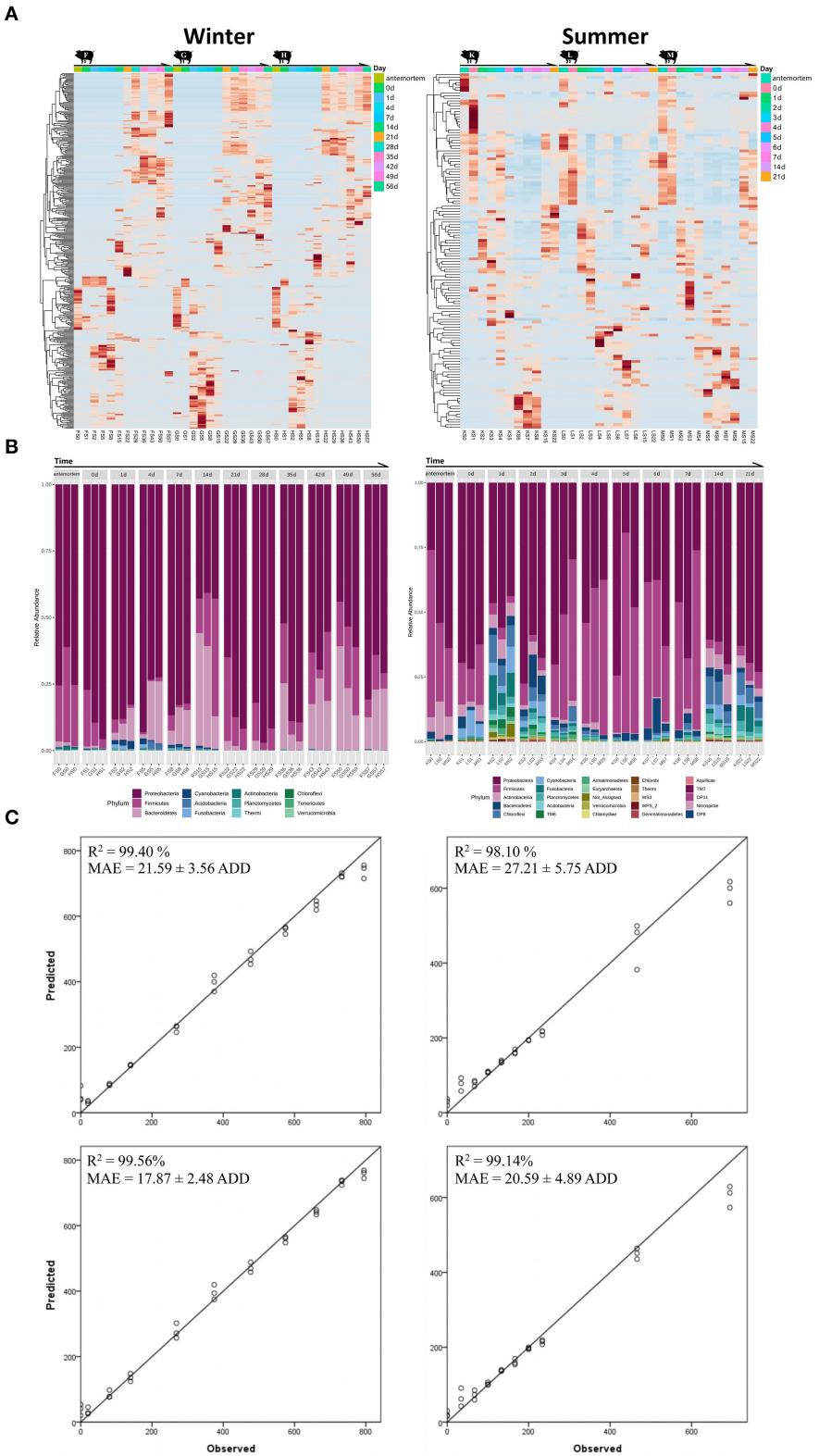


FIGURE 3 | The heat maps, taxa histograms, and RF models for the winter (left panel) and summer (right panel) trials. **(A)** Heat maps for the epinecrotic communities of both seasons. The samples on the *X*-axis are grouped by repeating experimental groups and ordered by the sampling time. The OTUs on the *Y*-axis are hierarchically

(Continued)

FIGURE 3 | clustered with the Minkowski distance matrices. **(B)** The phylum composition and variation of the epinecrotic communities during both seasons. The samples are arranged chronologically on the *X*-axis. The sample time points are marked on the top of each stacked bar and "a" represents the antemortem. The stacked bars of disparate colors show the relative abundance of each phylum in the epinecrotic communities. **(C)** The predicted ADDs calculated by the established RF models vs. observed ADDs in each trial were plotted with a one-to-one line added for reference. The RF models are based on the total OTUs (the top two plots) and biomarker sets (the bottom two plots) illustrate the correlation between observed ADD and predicted ADD.

the models to predict PMSI based on the epinecrotic biofilm succession. The models based on the data of the winter and summer trials explained 98.90 and 95.80%, respectively, of the ADDs since the placement of carcasses in the water. **Figure 3C** depicts the predicted ADDs vs. the observed ADDs, with the one-to-one line added for reference. The R^2 values between observed and predicted ADDs were 99.40 and 98.10%, and the MAEs were 21.59 ± 3.56 and 27.21 ± 5.75 ADD, respectively. Since the average daily water temperature in the non-flowing aquatic environment was mostly stable during the winter and summer trials (14.25 ± 3.70 and 32.51 ± 1.22°C, respectively), ADDs increased linearly during both seasons, which allowed the models to predict PMSI with the error about 1.5 and 0.8 days during 56 and 21 days of decomposition, respectively.

A large proportion of OTUs in the epinecrotic biofilm did not affect PMSI prediction. We list the taxa annotation of the top 167 and 186 important OTUs which made significant contributions (≥0.1%) to the accuracy of the RF models, respectively (**Supplementary Table S5**). It should be noted that these influential taxa use for the PMSI estimation were not present at the higher abundances in the epinecrotic biofilm. They only accounted for 36.12 ± 15.60 and 18.25 ± 12.50% of the communities in each season. For establishing more efficient models, the OTUs with low contribution were removed and the important OTUs were used as the predictive biomarkers in the training sets of the RF machine learning algorithm. In comparison to using total OTUs, the R^2 values of the RF models based on the biomarker sets increased to 99.56 and 99.14%, and the MAEs decreased to 17.87 ± 2.48 and 20.59 ± 4.89 ADD, respectively (**Figure 3C**). It means that the models allow us to predict PMSI with the error of about 1.3 and 0.7 days during 56 and 21 days of decomposition, respectively. Since only 10 biomarkers were shared in the training data of the winter and summer trials, the cross-validation of the different season models was weak in R^2 values (0.20 and 0.69%) and MAEs (202.73 ± 29.57 and 292.25 ± 33.95 ADD). These findings suggested that the RF models based on different seasons cannot be interchangeably used for PMSI prediction.

DISCUSSION

While there have been several studies of the PMI estimation using the microbial succession related to carcasses in the terrestrial ecosystems (Guo et al., 2016; Fu et al., 2019), research of such scope is scarce for the aquatic environments. In the present study, the bacterial succession patterns of epinecrotic and epilithic biofilms in a non-flowing aquatic habitat have been analyzed by sequencing the V4 region of 16S rDNA. Similar to the findings of the previous terrestrial carcass decomposition studies, prominent bacteria successions occurred in a predictable and reproducible manner on the surfaces of the carcasses submerged in water. As shown in the heat maps (**Figure 3A**), the clusters of high abundance taxa changed across each of the decay stages. In addition, these successive changes were highly repetitive between carcasses within coincident experimental settings. In the human cadaver experiments done by Metcalf et al. (2016), the prominent successions of bacterial communities across terrestrial subjects within a season have been observed. The heat maps which were generated with a similar approach as ours have shown the reproducible succession patterns of bacterial communities colonizing the skin of human cadavers. The conclusions of these studies indicated that the epinecrotic community structure variations during the postmortem decomposition processes of terrestrial and aquatic cadavers followed specific succession patterns, which could be quantitatively analyzed by HTS and potentially applied for estimating PMI/PMSI. However, the disparate community structures and succession patterns in different seasons have been observed (**Figures 3A,B**; **Supplementary Figure S10**), meaning that existing findings were not sufficient for the use in diverse environments. Whether this bacteria-dependent method can be used in the forensic practice depends on the determination of the influence of various environmental factors on the succession patterns. Through the present study, we can confirm that the seasonal conditions, such as temperature, light, and water quality, can be the critical factors to community successions and need to be further clarified in the subsequent researches, which should be conducted in the artificially controlled conditions.

The epinecrotic communities of aquatic and terrestrial carcasses shared similar dominant bacterial phyla (Proteobacteria, Firmicutes, Bacteroidetes, and Acidobacteria). Using the replicate swine carcasses as models for the human decomposition research, Pechal et al. (2014) studied the epinecrotic bacteria throughout decomposition in the forest habitat. Each decay stage had a unique profile of four dominant phyla that were changing in disparate trends. Proteobacteria was the most dominant phylum which decreased over time. Firmicutes became the dominant taxon as decomposition progressed. Bacteroidetes occurred only in the fresh and bloated stages. Actinobacteria, which represented the majority of the communities found on the skin of the carcasses, had disappeared during the dry stage. In the present study of the drowned swine carcasses (**Figure 3B**), Proteobacteria and Firmicutes were the dominant phyla before the body had entered the water. During the decay stages in the water, Proteobacteria remained dominant throughout, while. Firmicutes had first decreased, and then increased in the floating decay stage. Bacteroidetes representing a minor part of the antemortem communities had increased first, followed with the short decrease, and then kept

increasing up until the sunken remains stage. Acidobacteria was scarce and occurred opportunistically throughout the decomposition. It can be inferred from the above results that the taxonomic compositions of cadaver epinecrotic communities were similar, whereas the variation trends of each phylum were completely different between the habitats. This may be one of the main reasons why cadaver decomposition is significantly faster in the terrestrial setting (5 days) than that in the water (14 days in summer and 56 days in the winter). Therefore, it is necessary to further study the epinecrotic communities on the aquatic cadavers to widen their extent for the forensic microbiology application.

To date, there have been several studies on the PMSI estimation using succession patterns of the microbial communities on the aquatic carcasses, involving prokaryotic and eukaryotic microflorae. Hyun et al. investigated microeukaryotic biodiversity and community structures on the drowned pig (Hyun et al., 2019). The sequencing analysis showed the water molds and algae were related to the carcass decomposition. Relative abundances of the *Filobasidium*, *Achlya*, *Saprolegnia*, *Hydrodictyon*, *Lobosphaera*, and *Scenedesmus* varied across the decay stages. However, the change in microeukaryotic biodiversity with the decomposition progression was not significantly related to the PMSI. The insufficient biological replication limited the establishment of a mathematical model for the PMSI estimation. By using numerous fresh porcine cadaver bones as the biological replications, Randall et al. determined that eukaryotic community succession had occurred on porcine skeletal remains in a freshwater lake, which allowed the development of the mathematical approach for the PMSI estimation using the RF regression (Randall et al., 2021). Resulting models for the sample data from the ribs and scapulae predicted PMSI with errors of ± 104 (937 ADD) and ± 63 days (564 ADD), respectively. Such high error rates for PMSI estimation suggested that the eukaryotic succession patterns may not be applicable for the rigorous forensic investigation. Notably, the existing body of research on microbial succession on aquatic carcasses was focused mostly on bacterial microflora, especially on that in the epinecrotic biofilms formed on the liquid–solid surfaces. Recently, Cartozzo et al. (2021) have conducted a parallel experiment with Randall et al. (2021), which also used porcine bones submerged in the similar aquatic habitat to explore the temporal changes in the bacterial community structure. They found that the community alpha diversity increased with ADD. Similarly, beta diversity changed significantly with ADD and had been reasonably explained using the environmental parameters and inferred functional pathways. RF models developed using the 24 ribs and 34 scapula family level taxa allowed the prediction of PMSI with root mean square error (RMSE) of 57 (522.97 ADD) and 37 days (333.8 ADD), respectively. By comparing the above parallel experiments, it is apparent that the PMSI could be estimated more accurately from the bacterial succession patterns than that from eukaryotes. Instead of using the porcine bones, present study explored the bacterial succession on the surfaces of intact swine carcasses. Based on the RF models, epinecrotic communities in winter and summer were determined to estimate PMSI with MAEs of 1.3 (17.87 ± 2.48 ADD) and 0.7 days (20.59 ± 4.89 ADD). Apparently, these results can so far only represent the application of epinecrotic biofilm in the present specific habitats. We also find that the established models based on different seasons cannot be interchangeably used due to the rarely shared taxon. Whether it can be applied to other environments remains to be further explored by studying the mechanisms of influence of the environmental factors (e.g., water quality parameters, and aquatic microbiota) affecting the bacterial succession.

The succession patterns of bacterial communities found in the present study have a certain degree of discrepancy from similar studies conducted in other aquatic habitats. Previously, we conducted the experiment on the rat carcasses in the water containers to control the environmental variables such as water temperature and DO (He et al., 2019). Consistent with the present study, the phyla of Proteobacteria, Bacteroidetes, Firmicutes, and Actinobacteria were abundant in the rat epinecrotic communities. However, the abundance of the Proteobacteria decreased while Firmicutes increased during decomposition in the past, which partially contradicted the results of the present experiments conducted in the field ponds (**Figure 3B**). Similar to the present experimental settings, Benbow et al. (2015) have described the bacterial community succession in the epinecrotic biofilms of swine carcasses in a flowing stream during summer and winter. They also confirmed that Proteobacteria decreased as Firmicutes increased over the decomposition process during both seasons, which remained the different from our experiments conducted in the non-flowing ponds (**Figure 3B**). In fact, the variation trends of Proteobacteria and Firmicutes in our observation seemed much more complicated. They did not show continuous growth or a downward trend, but rather fluctuated across decay stages. For example, in the summer trial, Firmicutes increased significantly from the floating decay stage (3 days) to the sunken remains stage (7 days), while Proteobacteria had just a slight decrease. However, Firmicutes was low in abundance on the final skeleton remains (14 & 21 days), with an increase of Proteobacteria (**Figure 3B**). In a study that most closely matched our experimental settings and conditions, Kaszubinski et al. (2020) detected the bacterial community structures on the surfaces of three replicate swine carcasses submerged in a non-flowing pond over six time points (from submerged fresh to advanced floating decay stages) in summer. They found that Firmicutes increased and then decreased with Proteobacteria changing reversely over the decomposition. The most accurate model for PMSI estimation was a quadratic regression of phyla Firmicutes, Proteobacteria, and Bacteroidetes. Compared with the previous studies, more sampling time points (11 in our summer trial) gave us a more detailed interpretation of the epinecrotic community succession in the aquatic habitats. However, the discrepant and discontinuous variations in the taxa abundance that occurred in different habitats also posed a challenge for establishing mathematical model for the PMSI estimation.

In addition, it can be concluded from our results that although the decomposition had progressed in the same habitats, there

were prominent differences between the succession patterns that occurred during the summer and winter. The water temperature, DO, conductivity, and salinity in the summer were significantly higher than those in winter (**Supplementary Figure S1**), which could produce variations in the aquatic bacterioplankton. In the present study, the species richness of the aquatic communities in winter was significantly lower than that in summer (**Supplementary Figure S4**). A great discrepancy in the community diversity between the two seasons is illustrated in the **Supplementary Figure S8**. There were no significant differences of the water quality and community diversity between the locations whether the carcasses had presented (**Supplementary Figure S6**). The above information provides the implication that the effect of the seasonal factors on the bacterioplankton community was greater than that of the presence of the decomposing carcasses. In this case, the succession patterns of the epinecrotic community are different between seasons but repeatable in the parallel experimental settings (**Figures 3A,B**), correlating with the carcasses in summer decomposing much faster than those in winter, while having the equal decomposition rates during the identical season. The seasonal differences in the epinecrotic bacterium have also been demonstrated in the research of Dickson et al. (2011) and Benbow et al. (2015). They suggested that seasonal factors influence bacterial composition more than the decomposition process itself. Understanding the seasonal influences and other environmental factors on aquatic and epinecrotic communities is essential to accurately estimate the PMSI, especially for the long-term decomposition process across the different seasons.

In the present study we have mainly focused on the bacterial successions in the biofilm formed on the liquid–solid surfaces. The formation and maturation of the biofilm on different substrates were accompanied by the specific bacterial succession patterns. Several differential taxa between the epinecrotic and epilithic communities have been identified by LEfSe analysis (**Supplementary Tables S3, S4**). Significant different succession patterns have been observed (**Figure 3A**; **Supplementary Figure S9**). The alpha diversity on carcasses was lower than that on tiles (**Supplementary Figure S5**). The distances between the two sample types increased with the PMSI (**Figure 2A**). Similar to the aquatic communities during both trials, the epilithic samples clustered stably. The epinecrotic communities showed significant variations through time, and while being initially close to the epilithic samples, subsequently separated from the other dots on the PCoA plots. These results were consistent with our previous experiment done in the water containers (He et al., 2019). Several studies have already demonstrated that epilithic and epixylic (decaying plant material) biofilms differ in community composition, limiting nutrients, exoenzyme activity, and fungal biomass (Das et al., 2007; Sinsabaugh et al., 2010; Tank and Dodds, 2010). Lang et al. (2016) further found significant differences in the community composition between the two biofilm types (inorganic vs. carrion). Notably, dominating microorganisms in the epinecrotic community included a large portion of heterotrophs or detritivores, whereas the epilithic community was mostly represented by autotrophs. This point of view has been further testified by Hyun et al. (2019), who found that the microeukaryotic communities on carcasses were also significantly different from those in the abiotic control objects. Considering the influences of seasonal and other environmental factors, the succession patterns of epilithic biofilm which coexist with epinecrotic biofilm in most aquatic habitats can be used as a temporal control for PMSI estimation.

In conclusion, this study explored the bacterial community succession patterns associated with the vertebrate remains decomposing in the water for further use in the PMSI estimation. We successfully identified the epinecrotic and epilithic bacterial community structure variations throughout the decomposition processes of carcasses submerged in the non-flowing ponds during two seasons, providing preliminary support for potential use of the biofilm communities in the forensic investigations. The prominent bacterial succession patterns in the epinecrotic biofilms formed on the solid–liquid surfaces of the aquatic carcasses provide a new insight for the accurate PMSI estimation. The influencing mechanisms of the environmental factors and the aquatic bacterioplankton on the epinecrotic biofilm communities should be studied further before the biofilm could be considered as an accurate indicator of the PMSI.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The animal study was reviewed and approved by the Medical Ethics Committee of Xiangya Hospital, Central South University.

AUTHOR CONTRIBUTIONS

XFu and JC designed the study. YH, YL, and XFa collected the samples for all analyses. LZ performed DNA extraction for 16S rDNA sequencing. KJ performed raw data analysis. FD and JG performed the statistical analysis. All authors contributed, reviewed, and approved the manuscript.

FUNDING

This work was funded by the National Natural Science Foundation of China (81971791 and 82030058), Scientific Research Fund of Zhejiang Provincial Education Department (Y202146041), Basic Research Funds of Hangzhou Medical College (KYYB202009), Doctoral Scientific Research Foundation of Hangzhou Medical College (00004F1RCYJ2001), and Undergraduate Training Program for Innovation of Zhejiang Province (S202113023095 and S202113023093).

ACKNOWLEDGMENTS

The authors are grateful for the technical support provided by Personal Biotechnology Co., Ltd. (Shanghai, China).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.951707/full#supplementary-material

REFERENCES

Battin, T. J., Sloan, W. T., Kjelleberg, S., Daims, H., Head, I. M., and Curtis, T. P., et al. (2007). Microbial landscapes: new paths to biofilm research. *Nat. Rev. Microbiol.* 5, 76–81. doi: 10.1038/nrmicro1556

Benbow, M. E., Pechal, J. L., Lang, J. M., Erb, R., and Wallace, J. R. (2015). The potential of high-throughput metagenomic sequencing of aquatic bacterial communities to estimate the postmortem submersion interval. *J. For. Sci.* 60, 1500–1510. doi: 10.1111/1556-4029.12859

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., and Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276

Burcham, Z. M., Pechal, J. L., Schmidt, C. J., Bose, J. L., Rosch, J. W., and Benbow, M. E., et al. (2019). Bacterial community succession, transmigration, and differential gene transcription in a controlled vertebrate decomposition model. *Front. Microbiol.* 10, 745. doi: 10.3389/fmicb.2019.00745

Byard, R. W. (2018). Putrefaction: An additional complicating factor in the assessment of freshwater drownings in rivers. *J. For. Sci.* 63, 899–901. doi: 10.1111/1556-4029.13614

Cartozzo, C., Singh, B., Swall, J., and Simmons, T. (2021). Postmortem submersion interval (PMSI) estimation from the microbiome of sus scrofa bone in a freshwater lake. *J. For. Sci.* 66, 1334–1347. doi: 10.1111/1556-4029.14692

Claesson, M. J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J. R., and Ross, R. P., et al. (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* 38, e200. doi: 10.1093/nar/gkq873

Das, M., Royer, T. V., and Leff, L. G. (2007). Diversity of fungi, bacteria, and actinomycetes on leaves decomposing in a stream. *Appl. Environ. Microbiol.* 73, 756–767. doi: 10.1128/AEM.01170-06

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., and Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., and Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 45, W180–W188. doi: 10.1093/nar/gkx295

Dickson, G. C., Poulter, R. T., Maas, E. W., Probert, P. K., and Kieser, J. A. (2011). Marine bacterial succession as a potential indicator of postmortem submersion interval. *For. Sci. Int.* 209, 1–10. doi: 10.1016/j.forsciint.2010.10.016

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Fu, X., Guo, J., Finkelbergs, D., He, J., Zha, L., and Guo, Y., et al. (2019). Fungal succession during mammalian cadaver decomposition and potential forensic implications. *Sci. Rep*. 9, 12907. doi: 10.1038/s41598-019-49361-0

Guo, J., Fu, X., Liao, H., Hu, Z., Long, L., and Yan, W., et al. (2016). Potential use of bacterial community succession for estimating post-mortem interval as revealed by high-throughput sequencing. *Sci. Rep*. 6, 24197. doi: 10.1038/srep24197

Haefner, J. N., Wallace, J. R., and Merritt, R. W. (2004). Pig decomposition in lotic aquatic systems: the potential use of algal growth in establishing a postmortem submersion interval (PMSI). *J. For. Sci.* 49, 330–336 doi: 10.1520/JFS2003283

He, J., Guo, J., Fu, X., and Cai, J. (2019). Potential use of high-throughput sequencing of bacterial communities for postmortem submersion interval estimation. *Braz. J. Microbiol.* 50, 999–1010. doi: 10.1007/s42770-019-00119-w

Huang, R., Li, M., and Gregory, R. L. (2011). Bacterial interactions in dental biofilm. *Virulence* 2, 435–444. doi: 10.4161/viru.2.5.16140

Humphreys, M. K., Panacek, E., Green, W., and Albers, E. (2013). Comparison of protocols for measuring and calculating postmortem submersion intervals for human analogs in fresh water. *J. For. Sci.* 58, 513–517. doi: 10.1111/1556-4029.12033

Hyun, C. H., Kim, H., Ryu, S., and Kim, W. (2019). Preliminary study on microeukaryotic community analysis using NGS technology to determine postmortem submersion interval (PMSI) in the drowned pig. *J. Microbiol.* 57, 1003–1011. doi: 10.1007/s12275-019-9198-0

Johnson, C. H., Dejea, C. M., Edler, D., Hoang, L. T., Santidrian, A. F., and Felding, B. H., et al. (2015). Metabolism links bacterial biofilms and colon carcinogenesis. *Cell Metab.* 21, 891–897. doi: 10.1016/j.cmet.2015.04.011

Kaszubinski, S. F., Receveur, J. P., Wydra, B., Smiles, K., Wallace, J. R., and Babcock, N. J., et al. (2020). Cold case experiment demonstrates the potential utility of aquatic microbial community assembly in estimating a postmortem submersion interval. *J. For. Sci.* 65, 1210–1220. doi: 10.1111/1556-4029.14303

Lang, J. M., Erb, R., Pechal, J. L., Wallace, J. R., McEwan, R. W., and Benbow, M. E. (2016). Microbial biofilm community variation in flowing habitats: Potential utility as bioindicators of postmortem submersion intervals. *Microorganisms* 4, 1. doi: 10.3390/microorganisms4010001

Lauber, C. L., Metcalf, J. L., Keepers, K., Ackermann, G., Carter, D. O., and Knight, R. (2014). Vertebrate decomposition is accelerated by soil microbes. *Appl. Environ. Microbiol.* 80, 4920–4929. doi: 10.1128/AEM.00957-14

Li, C., Wang, K., and Xu, N. (2019). A survey for the applications of content-based microscopic image analysis in microorganism classification domains. *Artif. Intell. Rev.* 51, 577–646. doi: 10.1007/s10462-017-9572-4

Liu, Y. J., Xie, J., Zhao, L. J., Qian, Y. F., Zhao, Y., and Liu, X. (2015). Biofilm formation characteristics of *Pseudomonas lundensis* isolated from meat. *J. Food Sci.* 80, M2904–M2910. doi: 10.1111/1750-3841.13142

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

Mateus, M., and Vieira, V. (2014). Study on the postmortem submersion interval and accumulated degree days for a multiple drowning accident. *For. Sci. Int.* 238, e15–e19. doi: 10.1016/j.forsciint.2014.02.026

Metcalf, J. L., Wegener, P. L., Gonzalez, A., Lauber, C. L., Knights, D., and Ackermann, G., et al. (2013). A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *Elife* 2, e1104. doi: 10.7554/eLife.01104

Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., and Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646

Pakosh, C. M., and Rogers, T. L. (2009). Soft tissue decomposition of submerged, dismembered pig limbs enclosed in plastic bags. *J. For. Sci.* 54, 1223–1228. doi: 10.1111/j.1556-4029.2009.01161.x

Pechal, J. L., Crippen, T. L., Benbow, M. E., Tarone, A. M., Dowd, S., and Tomberlin, J. K. (2014). The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. *Int. J. Legal Med.* 128, 193–205. doi: 10.1007/s00414-013-0872-1

Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., and Wang, Q., et al. (2020). Identification of COVID-19 samples from chest X-Ray images using deep learning: a comparison of transfer learning approaches. *J. Xray Sci. Technol.* 28, 821–839. doi: 10.3233/XST-200715

Ram, R. J., Verberkmoes, N. C., Thelen, M. P., Tyson, G. W., Baker, B. J., and Blake, R. N., et al. (2005). Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920. doi: 10.1126/science.1109070

Randall, S., Cartozzo, C., Simmons, T., Swall, J. L., and Singh, B. (2021). Prediction of minimum postmortem submersion interval (PMSI(min)) based on eukaryotic community succession on skeletal remains recovered from a lentic environment. *For. Sci. Int.* 323, 110784. doi: 10.1016/j.forsciint.2021.110784

Schoenly, K. G., Haskell, N. H., Hall, R. D., and Gbur, J. R. (2007). Comparative performance and complementarity of four sampling methods and arthropod preference tests from human and porcine remains at the Forensic Anthropology Center in Knoxville, Tennessee. *J. Med. Entomol.* 44, 881–894. doi: 10.1093/jmedent/44.5.881

Sekkat, N., Kalia, Y. N., and Guy, R. H. (2002). Biophysical study of porcine ear skin *in vitro* and its comparison to human skin *in vivo*. *J. Pharm. Sci*. 91, 2376–2381. doi: 10.1002/jps.10220

Sinsabaugh, R. L., Golladay, S. W., and Linkins, A. E. (2010). Comparison of epilithic and epixylic biofilm development in a boreal river. *Freshwater Biol.* 25, 179–187. doi: 10.1111/j.1365-2427.1991.tb00483.x

Tank, J. L., and Dodds, W. K. (2010). Nutrient limitation of epilithic and epixylic biofilms in ten North American streams. *Freshwater Biol.* 48, 1031–1049. doi: 10.1046/j.1365-2427.2003.01067.x

Ueland, M., Breton, H. A., and Forbes, S. L. (2014). Bacterial populations associated with early-stage adipocere formation in lacustrine waters. *Int. J. Legal Med.* 128, 379–387. doi: 10.1007/s00414-013-0907-7

van Daalen, M. A., de Kat, D. S., Oude, G. B., de Leeuwe, R., Warnaar, J., and Oostra, R. J., et al. (2017). An aquatic decomposition scoring method to potentially predict the postmortem submersion interval of bodies recovered from the north sea. *J. For. Sci.* 62, 369–373. doi: 10.1111/1556-4029.13258

Wallace, J. R., Merritt, R. W., Kimbirauskas, R., Benbow, M. E., and McIntosh, M. (2008). Caddisflies assist with homicide case: determining a postmortem submersion interval using aquatic insects. *J. For. Sci.* 53, 219–221. doi: 10.1111/j.1556-4029.2007.00605.x

Widya, M., Moffatt, C., and Simmons, T. (2012). The formation of early stage adipocere in submerged remains: a preliminary experimental study. *J. For. Sci.* 57, 328–333. doi: 10.1111/j.1556-4029.2011.01980.x

Zimmerman, K. A., and Wallace, J. R. (2008). The potential to determine a postmortem submersion interval based on algal/diatom diversity on decomposing mammalian carcasses in brackish ponds in Delaware. *J. For. Sci.* 53, 935–941. doi: 10.1111/j.1556-4029.2008.00748.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Microbiology

TYPE Methods
PUBLISHED 19 August 2022
DOI 10.3389/fmicb.2022.963059

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Ping Huang,
Academy of Forensic Science, China
Dongri Li,
Southern Medical University, China
鹏 赵,
Northeastern University,
China

*CORRESPONDENCE
Yong Niu
niuyong770204@163.com
Chao Liu
liuchaogzf@163.com
Jian Zhao
zhaojian0721@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 07 June 2022
ACCEPTED 19 July 2022
PUBLISHED 19 August 2022

CITATION
Yu W, Xiang Q, Hu Y, Du Y, Kang X, Zheng D, Shi H, Xu Q, Li Z, Niu Y, Liu C and Zhao J (2022) An improved automated diatom detection method based on YOLOv5 framework and its preliminary study for taxonomy recognition in the forensic diatom test. *Front. Microbiol.* 13:963059. doi: 10.3389/fmicb.2022.963059

An improved automated diatom detection method based on YOLOv5 framework and its preliminary study for taxonomy recognition in the forensic diatom test

Weimin Yu[1†], Qingqing Xiang[2†], Yingchao Hu[3], Yukun Du[4], Xiaodong Kang[5], Dongyun Zheng[5], He Shi[5], Quyi Xu[5], Zhigang Li[5], Yong Niu[6*], Chao Liu[5*] and Jian Zhao[5*]

[1]Jiangsu JITRI Sioux Technologies Co., Ltd., Suzhou, China, [2]School of Forensic Medicine, Kunming Medical University, Kunming, China, [3]LabWorld (Suzhou) Intelligent Technology Co., Ltd., Suzhou, China, [4]School of Forensic Medicine, Southern Medical University, Guangzhou, China, [5]Key Laboratory of Forensic Pathology, Guangzhou Forensic Science Institute, Ministry of Public Security, Guangzhou, China, [6]Section of Forensic Sciences, Department of Criminal Investigation, Ministry of Public Security, Beijing, China

The diatom test is a forensic technique that can provide supportive evidence in the diagnosis of drowning but requires the laborious observation and counting of diatoms using a microscopy with too much effort, and therefore it is promising to introduce artificial intelligence (AI) to make the test process automatic. In this article, we propose an artificial intelligence solution based on the YOLOv5 framework for the automatic detection and recognition of the diatom genera. To evaluate the performance of this AI solution in different scenarios, we collected five lab-grown diatom genera and samples of some organic tissues from drowning cases to investigate the potential upper/lower limits of the capability in detecting the diatoms and recognizing their genera. Based on the study of the article, a recall score of 0.95 together with the corresponding precision score of 0.9 were achieved on the samples of the five lab-grown diatom genera *via* cross-validation, and the accuracy of the evaluation in the cases of kidney and liver is above 0.85 based on the precision and recall scores, which demonstrate the effectiveness of the AI solution to be used in drowning forensic routine.

KEYWORDS

forensic science, drowning, diatom test, artificial intelligence, YOLOv5 framework, microwave digestion-vacuum filtration-automated scanning electron microscopy

Introduction

In forensic sciences, it has been widely proved that the diatom test is an effective method for the diagnosis of drowning from other causes of death (Pollanen et al., 1997; Ludes et al., 1999; Zhao et al., 2017). As one of the unicellular algae, the diatoms exist in almost all water bodies, and naturally, they would go along with the inhaled water into the

lung of a drowning person, and these diatoms would appear in some other organs like the liver and kidney through the circulation of blood. However, a dead victim that was caused by other reasons but found in a water body would notpass the diatom test on his/her liver and kidney samples due to the end of the blood circulation (Kaushik et al., 2017). Even in drowning cases, there is only a small amount of diatoms in the closed organs which makes it difficult for forensic pathologists to detect them. In addition, there are hundreds of diatom genera living in the world, and the number of the dominant genera in a specific water region is countable, which allows for the construction of a diatom database to infer the drowning site of a drowned body (Zhang et al., 2021).

Either the diagnosis of drowning or the drowning site inference can resort to the diatom test by detecting the diatoms from the sediments in the tissue samples of multiple organs and then identifying their types for statistical analysis. To capture the diatoms varying from a few micrometers to a submillimeter, microscopy is required to scan the images at a magnification from a hundred to a thousand depending on optical microscopy or scanning electron microscopy (SEM). Traditionally, the diatom test always involves large numbers of laborious and tedious observation and search jobs on the scanned optical or SEM images, which have to be physically done by forensic pathologists. This situation is not friendly for practice and is apt to cause high false negative/positive rates due to fatigue and decreased concentration. It is of particular interest for academic research to explore the capability of automatically detecting the diatoms and/or recognizing the genera of the diatoms on optical microscope images (Bueno et al., 2018; Zhou et al., 2019, 2020; Kloster et al., 2020; Krause et al., 2020) or the SEM images (Deng et al., 2020; Yu et al., 2021). These studies are inspired by the development of artificial intelligence recently and especially the giant success of deep learning (LeCun et al., 2015) in image processing and analysis, such as image classification, object detection, and region-of-interest (ROI) segmentation, which then makes it possible to build our own intelligent diatom test solution.

Deep learning is a category of machine learning (Jordan and Mitchell, 2015) that is within the scope of artificial intelligence, and artificial intelligence allows machines to work efficiently and solve problems automatically based on the technologies of machine learning and pattern recognition which is another domain. For machine learning, there is a long history of development and prosperity, and conventionally the machine learning methods always contain a key step called feature engineering to design high-dimensional hand-crafted descriptors for downstream tasks like classification. In Safavian and Landgrebe (1991), Fischer and Bunke (2001), Jalba et al. (2001), and Gloria et al. (2017), a few studies were conducted on the taxonomy of the diatoms on the microscopic images based on machine learning. Various features were proposed to effectively distinguish the diatoms from other objects and these features were generally computed from statistical, textural, and morphological information. Then, a classifier such as a decision tree was trained on the feature data extracted from the given training images to infer the genera of the diatoms. However, conventional machine learning is not very suitable for the detection of diatoms due to the difficulty in encoding the position of diatoms to a high-dimensional feature representation. In Paul and Jones (2001), this challenging work was first and preliminarily evaluated with different visual descriptors and classifiers based on the Viola-Jones object detection framework.

In general, optical microscopy is not that powerful for zooming in on the features of diatoms when compared to scanning electron microscopy; however, the former with the advantage of much lower cost is enough for the classification of the images about if they contain the diatoms or not. In Zhou et al. (2020), 58 sample slides were scanned by a Leica scanner at a 40× magnification, and each slide image was split into a group of 255 × 255 non-overlapping small patches. The deep learning classification model Inception-v3 proposed by Google (Szegedy et al., 2016) was learned on the given training patches for binarily predicting if one test patch includes at least one diatom or not. By sliding window, the location of the diatoms can be coarsely determined. Similarly, the study of taxonomically identifying the morphologically diverse microalgal group of diatoms was reported on Kloster et al. (2020). The images for the study were acquired by an optical scanner with a pixel resolution of 0.1 μm. The classical model VGG16 (Karen and Zisserman, 2014) was adapted for the evaluation and a high F1 score of 0.97 was achieved.

Object detection is not well tackled until the introduction of deep learning on this diatom image processing task (Deng et al., 2009). In this pioneering work, feature engineering is replaced by a deep neural network called R-CNN to automatically learn the representation of a high-dimensional latent space on a large-scale image database ImageNet (Girshick et al., 2014), including 14 million images with rich morphological and textural features, and thus it provides the potential to build a strong capability of generalization. Faster R-CNN (Ren et al., 2017), as the third generation of the R-CNN, is a robust object detection framework that has been used for the detection of the diatoms on the SEM images (Deng et al., 2020) for the diatom test. They compared the results achieved by the faster R-CNN model and three conventional machine learning methods which demonstrated the superiority of deep learning. This is a preliminary investigation on the automatic diatom detection issue, while some detailed information like the magnification of image acquisition and the false negative/positive rates are not mentioned. In Yu et al. (2021), we assessed the performance of detecting the diatoms on an 800× image set and a 1,500× image set with another well-known object detection model RetinaNet (Lin et al., 2017). Both image sets were scanned by a Phenom XL desktop scanning electron microscopy, and we set

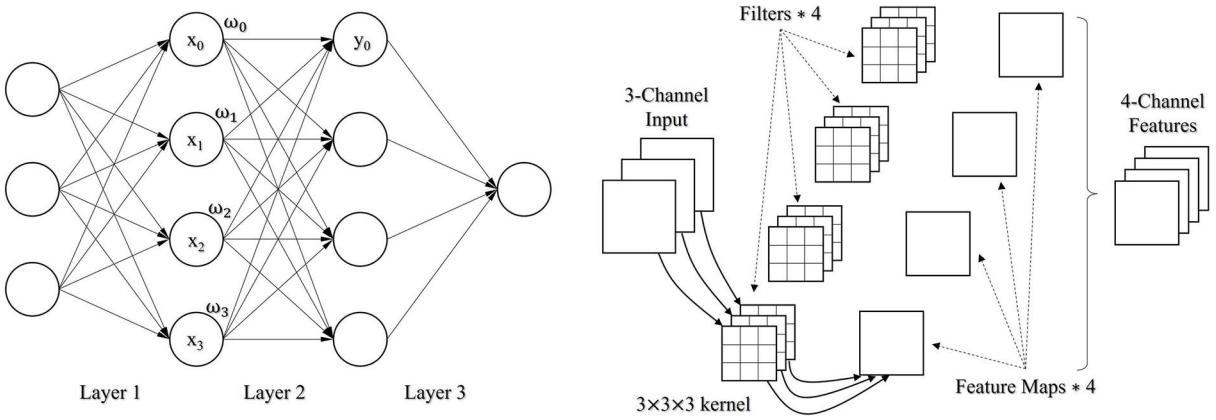


FIGURE 1
A fully connected network with 3 FC layers (Left); An illustration of how a convolutional layer works (Right).

the magnification to a low-medium level to substantially save the time of scanning which is routinely quite needed. Consequently, a 12% false negative rate and a corresponding 18% false positive rate were achieved. In Krause et al. (2020), the evaluation was performed on a group of two-channel (fluorescence and phase contrast) microscopic images, and the F1 score of 0.82 was achieved on another 600 test images.

In our previous work (Yu et al., 2021), we adapted the deep learning object detection framework RetinaNet for a preliminary evaluation of the SEM-based diatom detection. In Yu et al. (2021), considering the quantity limit of the collected SEM images, we applied a strategy of data augmentation by randomly cropping a single 1,024 × 1,024 SEM image to a local 512 × 512 region that contains at least one diatom for training, and splitting one test image to four 512 × 512 image patches for inference. In this study, it is not necessary because we collected much more images for training and testing. We adopt another AI-based object detection framework YOLOv5 (YOLOv5 GitHub Repository)[1] which is the latest version of the deep learning architectures YOLO (Redmon et al., 2016). One prominent difference is that the RetinaNet-101 model has more than 5.532×10^7 parameters while a medium YOLOv5 model is more compact with only about 2.104×10^7 parameters, which means there is less computation and faster. In the meanwhile, the YOLOv5 has been proved to be superior to the RetinaNet model for accuracy.

No matter the RetinaNet or the YOLOv5 object detection method, both have a structure of convolution neural network (CNN) belonging to the scope of deep neural network. A conventional neural network, that is, a fully connected (FC) network is completely based on the connection of adjacent neurons along the direction of propagation (Figure 1, Left). The mathematical form of a fully connected network can be represented as $y = F\left(\sum_{i=0}^{N} \omega_i \cdot x_i + b\right)$, where $\{\omega_i\}$ are the learnable weight parameters, b is either a constant value or a learnable parameter as a bias factor, and F is an activation function like sigmoid or softmax function to involve nonlinearity in the network. In practice, the fully connected network has some issues with handling the tasks like image classification, detection, and segmentation. Particularly, when a fully connected network is a little deep, it is prone to overfitting due to too intensive computation. By contrast, the convolution neural network is the combination of multiple types of neural computing layers, including the convolutional layer, pooling layer, and the mentioned fully connected layer. A convolutional layer convolves an input and passes its result to the next layer. The input is filtered by a set of convolution kernels with a limited number of learnable parameters (Figure 1, Right) compared to the fully connected layer. Therefore, the adaptation of the convolutional layers allows building a deep neural network with a better capability of the fitting. Besides, the pooling layer is used to downsample a feature map by voting on a local, for example, 2 × 2 feature. There are some pooling methods, such as average pooling and max pooling.

Following the previous evaluation (Yu et al., 2021), we continue the work by investigating new deep learning technologies to achieve better performance and developing a practical SEM-based diatom detection technology with an artificial intelligence engine. Moreover, we conducted a more comprehensive study trying to approach the potential upper and lower limits of our proposed method which will be introduced in the following section.

Materials and methods

Figure 2 illustrates the workflow of our proposed SEM-based diatom detection and recognition solution which can be broken down into multiple modules. The workflow begins with

1 https://github.com/ultralytics/yolov5

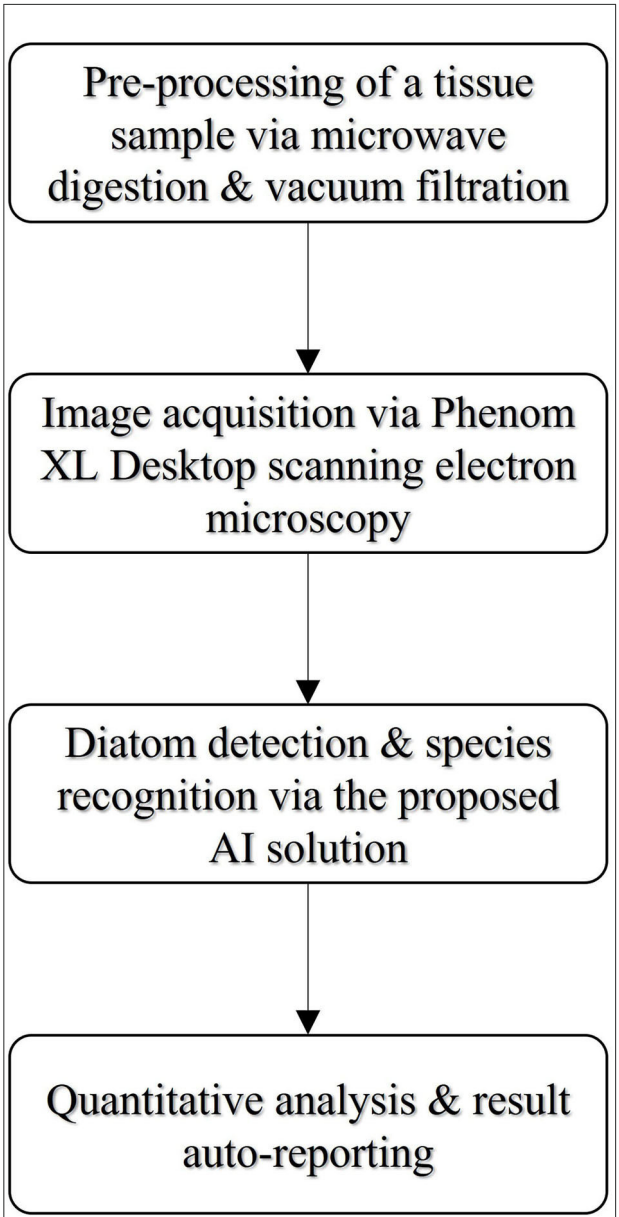


FIGURE 2
The workflow of the proposed diatom detection and recognition solution.

a hierarchical pre-processing module combining microwave digestion and vacuum filtration (Zhao et al., 2013, 2017) developed by Guangzhou forensic science institute, and followed by the image acquisition using a Phenom XL desktop SEM at a certain magnification. This workflow has been proved to be a sensitive method for the forensic diatom test (Zhao et al., 2013) compared to the conventional acid digestion method. We scanned the pre-processed tissue samples using back-scatter electron mode (BSE) as images and fed them into our developed detection and recognition AI solution, which is composed of a bunch of automatic functions like the AI-based diatom detection and recognition, quantitative analysis and report generation, and the function of training your own models for some specific sample cases from end-users.

Data acquisition

To comprehensively evaluate the accuracy of the diatom detection and genus recognition models, we collected three types of sample data as follows:

A. Samples of five lab-grown diatom genera with the names of *Coscinodiscus, Cymbella, Navicula, Nitzschia, and Synedra,* respectively;
B. Samples collected from lung tissues;
C. Samples collected from liver and kidney tissues.

The samples of *Coscinodiscus, Cymbella, Navicula, Nitzschia,* and *Synedra* were provided by the Institute of Hydrobiology, Chinese Academy of Sciences (Volume: 13–15 ml, Concentration: $>10^6$, Culture Condition: 25°C). These samples were processed by the Microwave Digestion-Vacuum Filtration-Automated Scanning Electron Microscopy as a Sensitive Method (Zhao et al., 2013).

The SEM membrane samples of lung, liver, and kidney tissues from nine cases that have been involved here are confirmed drowning by the eyewitness and autopsy findings of drowning signs and the exclusion of other injuries, drug, intoxication, alcohol, and medication-related. In this study, we assess the performance of the trained AI models on the given liver and kidney samples *via* cross-validation and demonstrate the efficiency of the solution in the general cases of drowning forensic diatom test. On the other hand, the samples extracted from lung tissue contain various impurities. Although the pre-processing steps of microwave digestion followed by vacuum filtration are applied in our workflow to remove those impurities, there are still many remaining impurities. Therefore, it is a real challenge to well detect and recognize the diatoms located in the SEM images of the lung samples, and herewith, we test the lung samples collected from those drowning cases for the evaluation on some extreme conditions with numerous different sediments which make the background of the images very complicated.

In addition, the samples of the five lab-grown diatom genera were collected from a laboratory environment, and an apparent difference between these samples and the samples collected from the lung, liver, and kidney tissues of drowning corpses is that the acquired SEM images from the lab-grown diatoms suffer less from the interference of impurities. Thus, the given samples are quite appropriate for the quantitative analysis of the potential upper limit performance on both the diatom detection and the genus classification.

The diatom test method that combines microwave digestion (MD) and vacuum filtration (VF) was proposed to replace the conventional pre-processing method based on acid digestion and centrifugation, and the former has a higher time efficiency and a better filtration quality (Zhao et al., 2013). We acquired the SEM images on these processed samples using a Phenom XL desktop SEM at the magnification of 1,500× with a pixel resolution of 0.33 μm and a field of view (FOV) of 336 μm. Each scanned image has a unified size of 1,024 × 1,024 pixels, and the positions and genera labeling of the diatoms were done by two senior forensic pathologists experienced in diatom tests.

For the samples of each lab-grown diatom genera, there are around 2,000 images scanned, and not all of them contain the diatoms (about 46%). Table 1 is the summary of the scanned SEM images of the standard samples evaluated in our study.

For the lung tissue samples, we mixed all the scanned images for training a robust diatom detection AI model based on a large dataset. In detail, there are totally 2,343 images while 1,783 images contain at least one diatom, and the total number of all the diatoms is 5,899. In addition, there are totally 11 samples collected from the liver and kidney tissues which are described in Table 2. Note that we inherited two sets of images from our previous work (Yu et al., 2021). The first set was scanned at the magnification of 800× (#01) and the second one was acquired under the setting of 1,500× magnification (#02). Besides, there are images from liver samples (#03→ #05) and kidney samples (#06→ #11) randomly selected from the nine drowning cases.

In comparison to the standard samples, the samples of the lung, liver, and kidney tissues are extracted from the real cases, and the number distribution of different diatom genera is not uniformly distributed for the training of a multi-class recognition AI model that can work well on the inference of all the target genera. The label information of both the lung data and the liver and kidney data is illustrated in Figure 3, and we notice that there are two interference labels named "debris" and "other." The label "debris" means the incomplete diatoms and the second label "other" denotes those uncommon diatom genera in forensic practice. Therefore, we only conduct the study of assessing the performance of the diatom detection based on the current samples.

TABLE 1 Summary of the SEM images scanned from the standard samples.

Genera	Scanned images	Images with diatoms	Diatom count
Coscinodiscus	2,018	630	812
Cymbella	2,084	672	921
Navicula	1,966	930	1,356
Nitzschia	1,999	1,476	6,515
Synedra	1,875	1,622	5,741
Total	9,942	5,330	15,345

TABLE 2 Summary of the SEM images scanned from the liver and kidney samples.

Image set	Images with diatoms	Diatom count
#01	904	2,789
#02	938	1,168
#03	8	8
#04	509	597
#05	108	113
#06	3	3
#07	1,687	2,125
#08	54	56
#09	69	72
#10	58	60
#11	35	39
Total	4,373	7,030

YOLOv5

Same as the RetinaNet framework (Lin et al., 2017), the YOLOv5 is also a one-stage object detection framework that takes a batch of the resized 3-channel SEM images as input and can directly predict the location of the diatoms and optionally their genera. The localization of each diatom candidate is predicted by a sub-regression model as a part of the YOLOv5 detection solution. On the other hand, it has a sub-classification model trained to recognize if the candidate is a real diatom with a confidence prediction. More specifically, the YOLOv5 has four model structures depending on the number of the model layers and parameters, ranging from small to super large, and we picked a medium model YOLOv5m for training and testing to evaluate the performance on the given image data. The YOLOv5m model has the network architecture illustrated in Figure 4, and the architecture can be broken down into a backbone network followed by a neck structure connecting to the section for detection prediction. In detail, the construction of the backbone network is based on the Focus module and CSP module (see Figure 4), and the neck structure is the enhancement of the FPN structure (Lin et al., 2017) appeared in the RetinaNet by adding a structure called PAN for bottom-up path aggregation. The combination of the FPN structure and the PAN structure was originally proposed for image segmentation (Liu et al., 2018) to shorten the information path between lower and topmost features. It was first introduced into YOLOv4 and then in the YOLOv5 framework, this structure was slightly modified with the replacement of some CBL modules by the CSP modules which are constructed based on CBL.

The prediction from the neck structure consists of three outputs with different feature sizes (e.g., 80). The prediction from the neck structure consists of three outputs with different

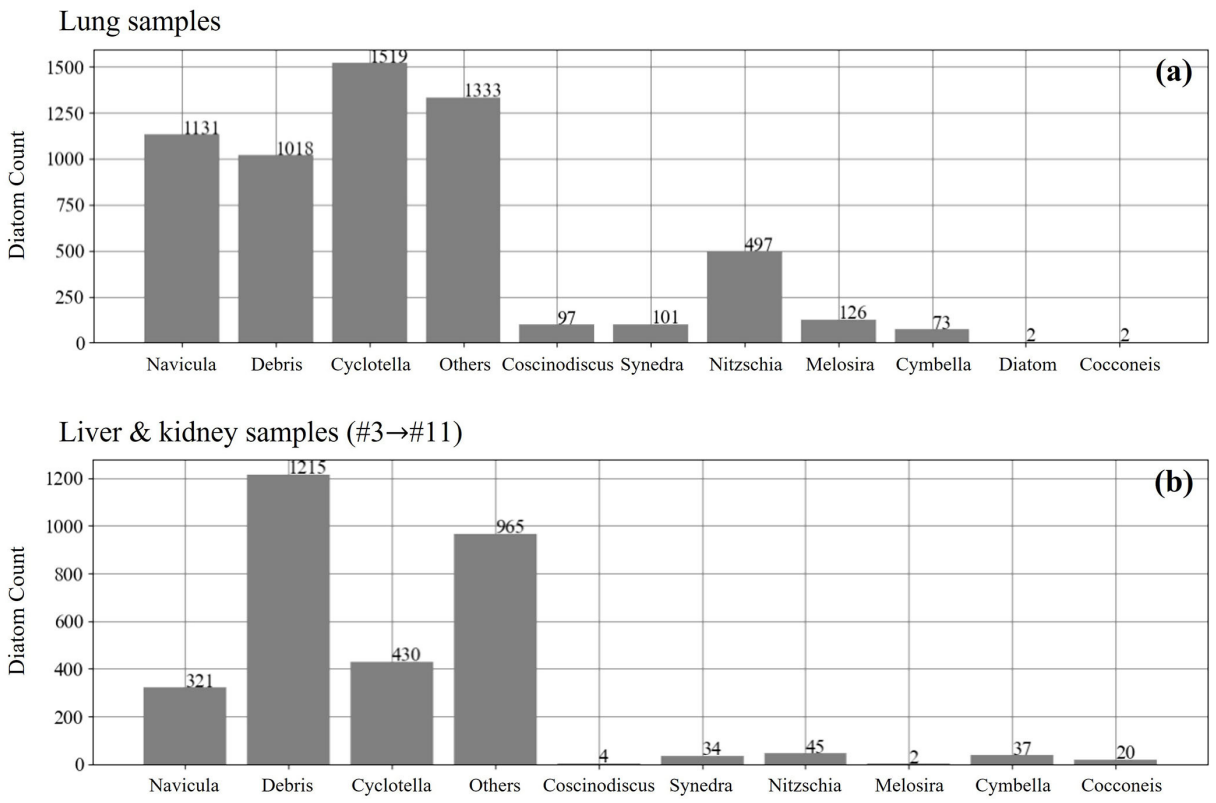


FIGURE 3
The number distribution of different diatom genus in the lung **(a)**, liver and kidney samples **(b)**.

feature sizes (e.g., 80 × 80, 40 × 40, and 20 × 20) and receptive fields. The 3rd dimension of every output is composed of four coordinates of a bounding box, one confidence score, and K probability values for the genus recognition (Redmon et al., 2016), and $K = 1$ if we only consider the diatom detection for counting the diatoms without the genus recognition. In addition, corresponding to the diatom classification and the diatom localization regression, the loss function used for the YOLOv5m model training can be divided into classification loss and bounding box regression loss. The classification loss is calculated *via* binary cross entropy (BCE) and the bounding box regression loss is calculated by a novel metric $\mathrm{CIoU}_{\mathrm{Loss}}$ which takes overlapping area, center distance, and aspect ratio into consideration. The formula of the $\mathrm{CIoU}_{\mathrm{Loss}}$ loss is defined in Equations (1–3).

$$CIoU_{Loss} = 1-(IoU-\frac{{Dist_2}^2}{{Dist_C}^2}-\frac{v^2}{(1-IoU)+v}) \quad (1)$$

$$IoU = \frac{O^{gt}\cap O^{p}}{O^{gt}\cup O^{p}} \quad (2)$$

$$v = \frac{4}{\pi^2}\left(arctan\frac{w^{gt}}{h^{gt}}-arctan\frac{w^{p}}{h^{p}}\right)^2 \quad (3)$$

IoU is the intersection over the union between the candidate localization prediction o^p and the ground truth localization o^{gt} (rectangle). $Dist_2$ is the center distance between the prediction and the ground truth, and $Dist_C$ is the diagonal distance of the ground truth. In addition, $\{w^{gt}, h^{gt}\}$ and $\{w^p, h^p\}$ are the width and height sizes of the ground truth and the prediction, respectively.

Moreover, the training of a deep neural network model is generally based on the back-propagation strategy which follows the chain rule to iteratively update the learnable parameters of the model. Gradient descent is used to optimize the training process, and specifically, we adapt the stochastic gradient descent optimization which can be formulated as Equations (4–6).

$$g_k = \nabla w_k + w_k * wd \quad (4)$$

$$v_{k+1} = v_k * mu + g_k \quad (5)$$

$$w_{k+1} = w_k - v_{k+1} * lr \quad (6)$$

Here, w_k is a parameter to be estimated by training, and the parameters wd, mu, and lr denote weight decaying, momentum, and learning rate, respectively.

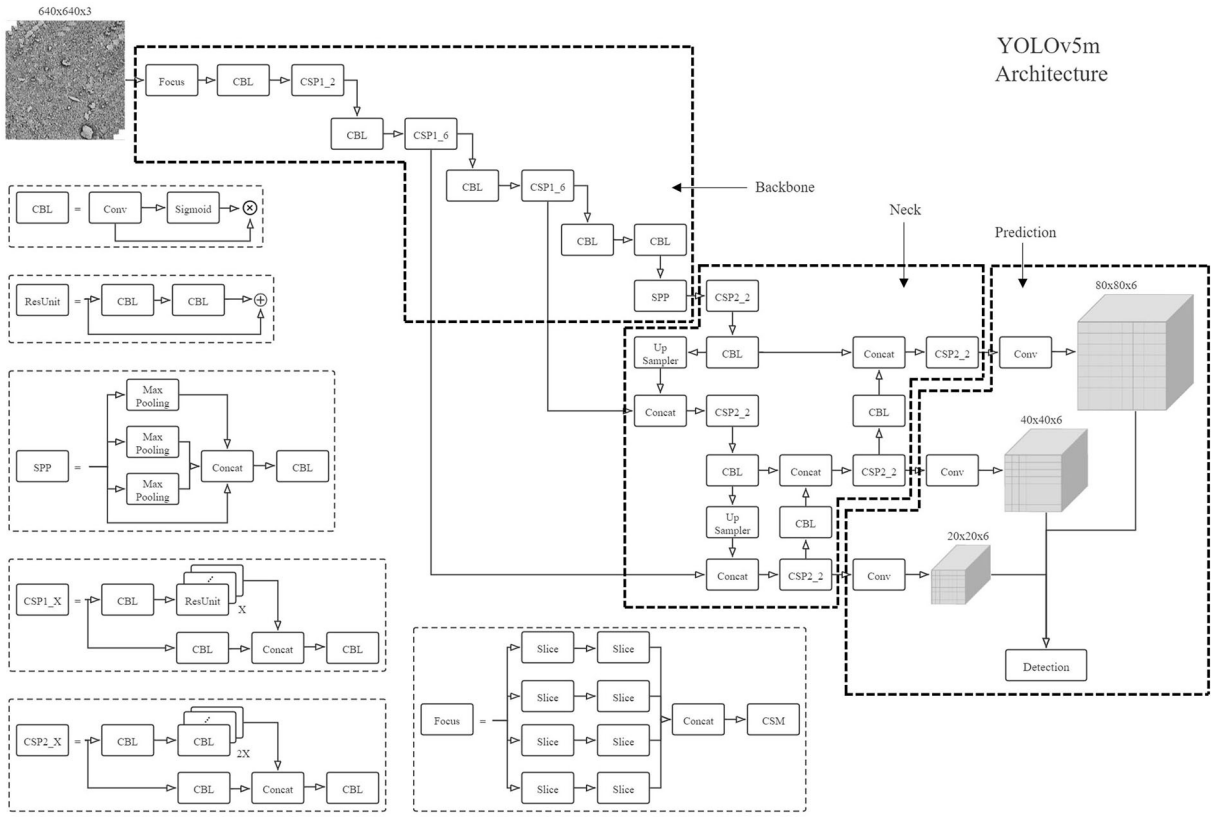


FIGURE 4
The YOLOv5m architecture.

Evaluation metric

In accordance with our previous work, we also calculate the recall and the precision given by Equations (6, 7) to evaluate the performance of the trained YOLOv5m models on the image data mentioned in the section of "materials and methods—data acquisition." Here, the terms *TP*, *FN*, and *FP* are the numbers of true positives, false negatives, and false positives, respectively. The recall metric is to reflect the proportion of the actual positives identified correctly and the precision metric is to answer the question: what is the proportion of the correct positive identification? This metric couple is very significant for the quantitative assessment of the false negative rate and the false positive rate. For instance, if there are a total of 100 diatoms for detection, a recall score of 0.95 means only five objects are not recognized. On the other hand, a precision score of 0.95 represents that 5% of all the detected objects are not diatoms.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

The precision-recall curve is another metric to reflect the overall performance of the trained models on a given data. Previously (Yu et al., 2021), we obtained a set of precision-recall measurements by manually setting different confidence threshold values for inference and plotted them as a precision-recall curve. Hereby, we propose a more elaborate definition of the precision-recall curve that is based on the following steps:

1. Sort the confidence scores (i.e., probabilities) of all the diatom candidates outputted from a trained AI model in descending order;
2. Iteratively update the confidence threshold from 0 to 1 with the change of a small step like 0.01. The threshold is the lower limit of accepting a detected object as a diatom according to its confidence score. For each threshold, we can calculate a couple of precision-recall values, and then a group of precision-recall values can be obtained by changing the confidence threshold;
3. Plot these precision-recall points as a curve and optionally smooth them if necessary.

The area under a given precision-recall curve (AUC), also known as average precision (AP), is a metric for assessing the overall accuracy of a model. In general, a higher AUC score

indicates potentially better performance on the same test dataset, and ideally a perfect case would be subject to the AUC score of 1.

The prediction on an SEM image *via* the trained YOLOv5m model depends on the inference settings of not only the already-mentioned confidence threshold but also the IoU threshold as the lower boundary of the overlapping level between a diatom candidate and the ground truth to decide whether the candidate can be accepted as a diatom. There are two more evaluation metrics associated with the IoU threshold: AP@0.5 and AP@0.5:0.95 that are involved as part of the measurements in our study. The metric AP@0.5 is the average precision at the IoU threshold of 0.5 and the metric AP@0.5:0.95 is the mean value of all the APs corresponding to the IoU threshold setting from 0.5 to 0.95 with an interval of 0.05.

To evaluate the performance of the multi-class diatom recognition on the lab-grown diatom samples, we introduce another two evaluation methods called mAP and confusion matrix into this study. For each diatom genus, there is an AP@0.5/AP@0.5:0.95 score and the mAP is essential to calculate the mean value of the average precisions in terms of all the classes. Therefore, in the case of the multi-class diatom recognition, we can also achieve the mAP@0.5 and mAP@0.5:0.95 scores other than the previous AP@0.5 and AP@0.5:0.95 for each genus. On the other hand, the confusion matrix in our scenario is a way of observing the implicit correlation among different diatom genera. Specifically, it demonstrates the relations in a matrix where the sum of each row is the actual number of one genus and each column includes the prediction results of each genus for a specific diatom class, in such a way that the number of correct and incorrect predictions are summarized with their counts and are broken down by each class. This can help us to find out which classes are hard to be differentiated and further can guide us to design more reasonable algorithms for distinguishing them. Note that the recall score of each class can be directly computed from the confusion matrix according to the definition of Equation (6).

Settings

The following studies were conducted on the hardware and software environments summarized in Table 3. All the scanned SEM images have the same image size 1,024 × 1,024 and considering the trade-off between the available computation resource of the Nvidia RTX 2080Ti GPU in Table 3 and a reasonable batch size, we resized each SEM image to either 800 × 800 or 640 × 640 before feeding it into the YOLOv5m model. For these two input image sizes, the corresponding batch sizes are 16 and 28, so that they are not too small and are not prone to cause the oscillation of training. The training parameters *wd*, *mu*, and *lr* for the stochastic gradient descent optimization are set to be 5e-4, 0.937, and 0.01. Moreover, we define the complete pass of a training dataset as an epoch, and the epoch number of every single training is set to be 100 for all the proposed studies. For a single training, each model instance after an epoch was used to test the given test data set, and we denote the model achieving the best AP@0.5 score as Best-640/Best-800 and the finally obtained model after 100 epochs as Last-640/Last-800.

TABLE 3 The configuration of hardware and software environment for evaluation.

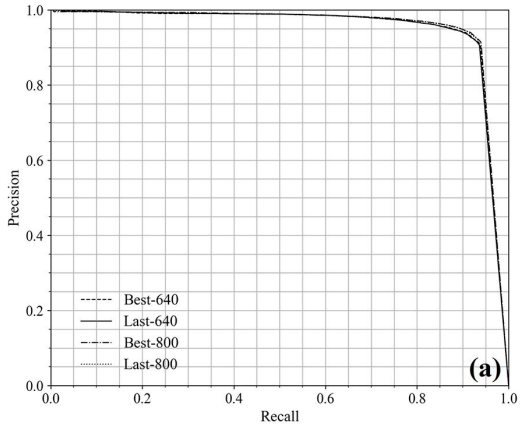
Hardware	CPU	Intel Xeon CPU E5-1620 v2 @ 3.70GHz
	RAM	24GB
	GPU	NVIDIA GeForce RTX 2080 Ti (×1)
	Video Memory	12GB
	Hard Disk	500GB
Software	OS	Windows 10
	Programming Toolkit	Python 3.9 + PyTorch 1.9 + CUDA 11.1
	IDE	PyCharm Professional

Results

Study on the samples of five lab-grown diatom genera

According to the summary of Table 1, we collected five lab-grown diatom genera that are *Coscinodiscus, Cymbella, Navicula, Nitzschia*, and *Synedra*. There are around 2,000 images prepared for each diatom genus, and about 46% of all the images have at least one diatom. We applied 4-fold cross-validation on the available images to evaluate both the single-class detection of the diatoms and the multi-class diatom recognition. Specifically, we partitioned the images of each genus into two categories depending on the criterion if one image contains at least one diatom or not. Furthermore, we uniformly divided the images of every category into four groups, and then all the images labeled with the same group index were mixed for the 4-fold cross-validation. In each fold, one group was picked for validation, and the rest three groups were used for training. All the models were initialized by the pre-trained parameters learned on the image dataset ImageNet (Girshick et al., 2014) before training.

For the single-class diatom detection, all the results achieved at the confidence threshold of 0.5 are outlined as a table in Figure 5. We can find that the input image size of 800 × 800 is slightly superior to 640 × 640 in terms of precision, and the recall score achieved by the Best-800 model has reached 0.94 while the corresponding precision score is 0.914. Besides, no matter whichever model, the AP@0.5 score is always around 0.95, which demonstrates the capability of the trained YOLOv5m models in handling the standard samples.



(b)

Model	Precision	Recall	AP@0.5	AP@0.5:0.95
Last-640	0.905	0.936	0.949	0.836
Last-800	0.915	0.935	0.95	0.843
Best-640	0.906	0.939	0.950	0.836
Best-800	0.914	0.94	0.952	0.844

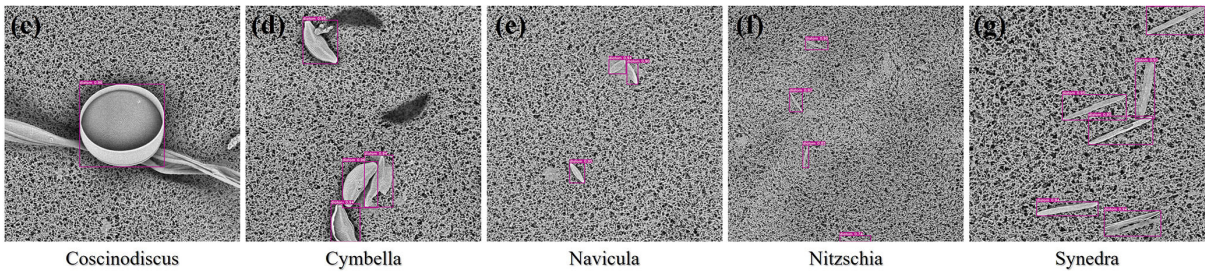


FIGURE 5
(a) The precision-recall curves of the single-class diatom detection under the confidence threshold 0.5 and the IoU threshold 0.5. **(b)** The precisions and recalls at the confidence threshold 0.5 achieved by different models. **(c–g)** The qualitative demonstration of the detection cases of the five test genus.

The precision-recall curves are plotted in the left-upper corner of Figure 5, and we also exhibited the detection cases of all five genera. Here, it is noticeable that the sizes of the diatom genera are quite different, which proves that the YOLOv5m architecture enables to capture the objects on a large scale.

In a forensic diatom test, the recall should be more important than the precision, in that the false positives can be possibly corrected *via* some post-processing strategies, such as an individual classification after the current detection. Since the precision and the recall are commonly a couple of measurements standing by the false negative rate and the false positive rate, respectively, we modulated the confidence threshold and achieved different results. Especially, when the confidence threshold is set to be 0.4, the recall score achieved by the Best-800 model is slightly higher than 0.95, while the associated precision score is 0.9.

For the multi-class diatom recognition which includes the diatom detection and the classification of every diatom candidate with a genus label, we computed the mAP@0.5 score and mAP@0.5:0.95 score from the AP results of each genus and summarized them in Figure 6. In comparison to the previous single-class diatom detection test, there are no remarkable differences between the Best-640 model and the Best-800 model, while the mean recall of the last-640 model is 1% higher than the one of the last-800 model. To get a perception of the model performance in each genus, we plot the precision-recall curves of the five genera in Figure 6. The AP@0.5 scores of the two diatom genera *Nitzschia* and *Synedra* achieved by the Last-640 model are considerably better than the Last-800 model, leading to the overall AP of the Last-640 model being superior to that of the Last-800 model. In addition, we notice that the performances of the Best-800 model on every genus are similar with a smaller variance of the AP@0.5 scores than that achieved by the Best-640 model.

The normalized confusion matrices in terms of the Last-640 and Best-640 models are summarized in Table 4, where we can find some hidden correlations among different genera. For instance, there is a 45% probability of misrecognizing *Nitzschia* as *Synedra* by the Last-640 model and indeed the two genera look rather similar in shape. Also, the size of the Synedra is very small in our standard samples and this genus is easy to be recognized as background. Moreover, among the false positives, more than 80% of them are identified as *Nitzschia* obtained by the Best-640 model. As well, more than 30% and 65% of the false positives detected by the Best-640 model are regarded as *Nitzschia* and *Synedra* individually. Overall, the confusion matrix is a useful tool to indicate the potential intra-class confusion for solution improvement.

Model	Mean Precision	Mean Recall	mAP@0.5	mAP@0.5:0.95
Last-640	0.87	0.879	0.896	0.828
Last-800	0.874	0.869	0.885	0.824
Best-640	0.902	0.922	0.942	0.859
Best-800	0.901	0.921	0.942	0.861

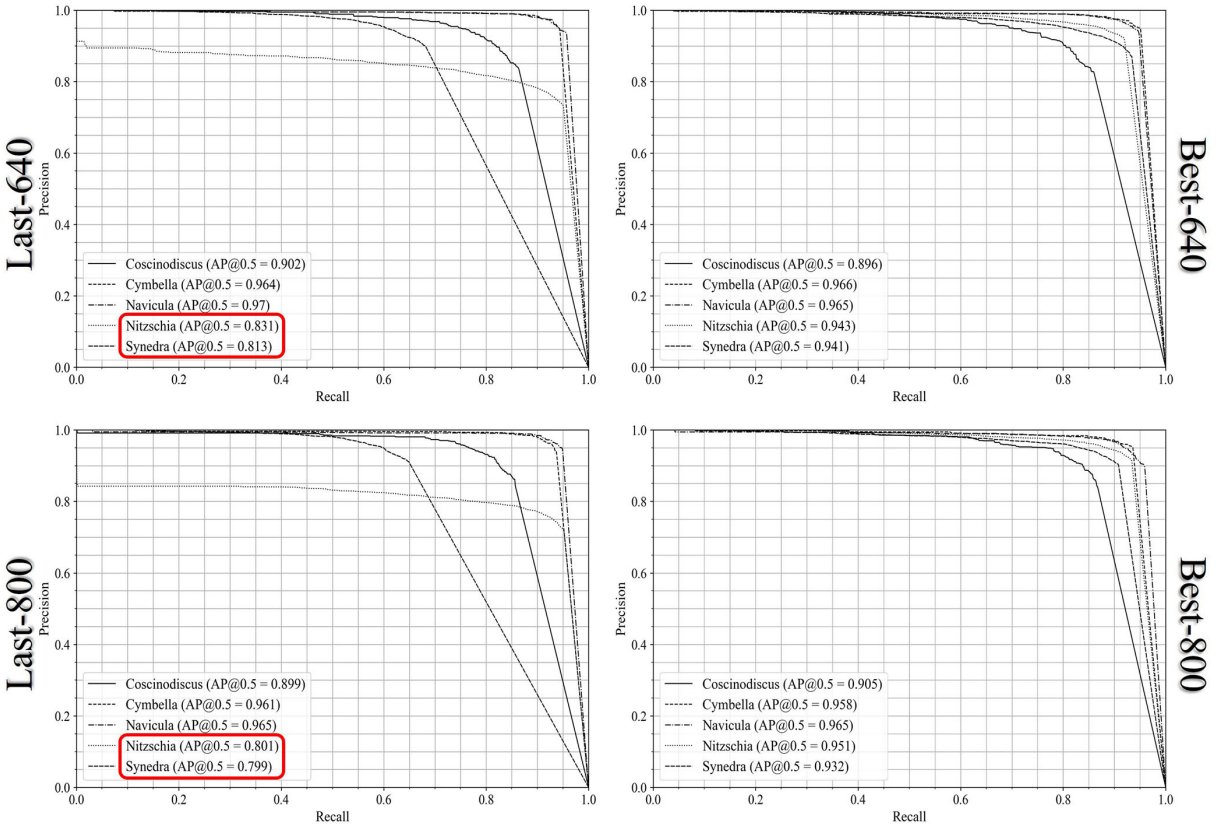


FIGURE 6
The precision-recall curves of all the diatom genus achieved by the Last-640, Last-800, Best-640, and Best-800 models.

Study on the samples of lung tissue

We applied the evaluation to the lung samples from the drowning cases. Since there are many sediments in the lung of a drowning corpse, it is very challenging to effectively distinguish the diatoms and the non-relevant objects in the test images. Therefore, this study can be regarded as a performance evaluation on the worst cases of the drowning forensic diatom test.

Again, we evenly split the SEM images scanned on the given lung samples into four groups and conducted a cross-validation. As already mentioned, due to the existence of two interference labels "debris" and "other," as well as the count imbalance among different genera, it is not suitable to launch a multi-class diatom recognition study, instead, we only focus on the search of all the diatoms in the images. Following the denotation of the models trained on the samples of the lab-grown diatom genera, we also compared the results achieved by the Last-640, Best-640, Last-800, and Best-800 models. Note that each training of the cross-validation begins with an initialization by the YOLOv5m model pre-trained on the SEM images of the lab-grown diatom samples, which already have learned some general features of diatoms.

In Figures 7a,b, the precision-recall curves are plotted in the left corner, and in the right corner, there is a summary of the achieved results at the confidence threshold of 0.5. In accordance with the summary, the recall score is getting higher by increasing the size of the images fed into the YOLOv5m from 640 × 640 to 800 × 800. As a result, the Last-800/Best-800 models perform better than the corresponding Last-640/Best-640 models because the textural and morphological information is more abundant. Quantitatively, the best recall score is above

10.3389/fmicb.2022.963059

TABLE 4 The confusion matrices derived from the multi-class recognition of the lab-grown diatoms with the Last-640 and Best-640 models.

Last-640		**Predicted**					
		Coscinodiscus	***Cymbella***	***Navicula***	***Nitzschia***	***Synedra***	**Background**
Actual	*Coscinodiscus*	0.94	0	0	0	0	0.06
	Cymbella	0	1	0	0	0	0
	Navicula	0	0	0.94	0	0.01	0.05
	Nitzschia	0	0	0	0.51	0.45	0.04
	Synedra	0	0	0	0	0.71	0.29
	Background	0.01	0.01	0.02	0.81	0.15	0
Best-640		**Predicted**					
		Coscinodiscus	***Cymbella***	***Navicula***	***Nitzschia***	***Synedra***	**Background**
Actual	*Coscinodiscus*	0.94	0	0	0	0	0.06
	Cymbella	0	1	0	0	0	0
	Navicula	0	0	0.94	0	0	0.06
	Nitzschia	0	0	0	0.88	0	0.12
	Synedra	0	0	0	0.05	0.88	0.07
	Background	0.01	0.02	0.01	0.31	0.66	0

The boxed values indicate high mis-recognition cases among some diatom genera and background.

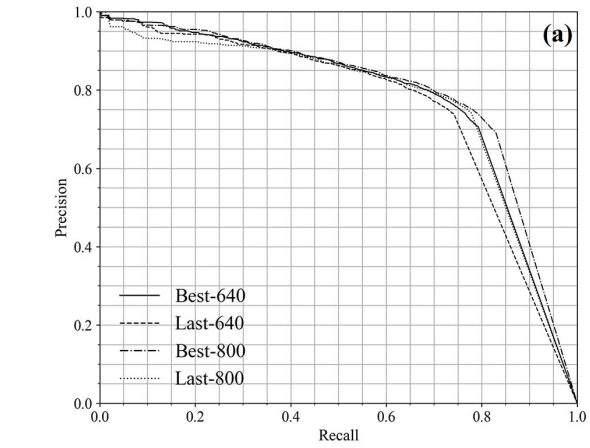
0.8 by the Best-800 model and the corresponding AP@0.5 score is also closed to 0.8. In Figures 7c–f, we qualitatively illustrated several detection cases achieved by the Best-800 model and imposed the confidence score of every diatom candidate on the test images.

Study on the samples of liver and kidney tissues

In this study, we evaluated the image data scanned from the liver and kidney tissues of some drowning corpses. Once more, the available SEM images cannot meet the requirement for a multi-class diatom recognition study due to the already mentioned reasons of the uneven count distribution of each genus, as well as a large portion of diatoms labeled as "other" and "debris." Hence, we took an evaluation of the single-class diatom detection with a 4-fold cross-validation, while only the input image size 800 × 800 is taken into consideration this time. We initialized each training of the cross-validation with the weights pre-trained on the dataset of ImageNet (Girshick et al., 2014) to reduce the influence of transfer learning. Note that the image dataset used in this study is composed of 11 samples with an obvious variation in the dirty level of the image background which can be found in Figure 8. The image quality of some samples is as poor as that of the previous lung samples, while in the best cases, there are only the diatoms left after the MD-VF pre-processing steps, therefore this study takes the general situation into account for a fair evaluation of the simulation of routine cases.

The assessment was conducted on the prepared image data and all the quantitative results are shown in Figures 9a,b. In comparison to the RetinaNet-101 architecture, the YOLOv5m achieved a balance between the precision score of 0.84 and the recall score above 0.86 at the confidence threshold of 0.5. For the same threshold, the RetinaNet-101-Last-800 model is tilted to the precision side, while the false negative rate is therefore much higher than that achieved from the YOLOv5m model. In Figures 9c–f, there are two couples of the diatom detection results predicted by the YOLOv5m-Last-800 model and the RetinaNet-101-Last-800 model.

Since the diatom candidates predicted from a YOLOv5m/RetinaNet-101 model will be filtered both by the IoU threshold and by the confidence threshold defined empirically, we would like to explore the impact of the two threshold parameters on both the precision and the recall to guide our practice. In detail, we kept one threshold at 0.5 and changed the other threshold from 0.1 to 0.5 with a step of 0.1 to observe the trend of performance. All the precision and recall scores are outlined in Tables 5, 6 where the precision and recall scores maintain stable when the IoU threshold is $\leq$0.5, and when the threshold is above 0.5, the average precision drops down according to the AP@0.5 score and the AP@0.5:0.95 score from the same model as shown in Figure 9. The invariant property of the precision and recall scores shows that most of the true positives and the corresponding ground truth are well overlapped with each other. On the other hand, when progressively changing the confidence threshold from 0.1 to 0.5, the precision score increases while the corresponding recall score decreases, and



(b)

Model	Precision	Recall	AP@0.5	AP@0.5:0.95
Last-640	0.739	0.741	0.756	0.551
Last-800	0.742	0.778	0.767	0.559
Best-640	0.703	0.794	0.777	0.566
Best-800	0.686	0.831	0.792	0.577

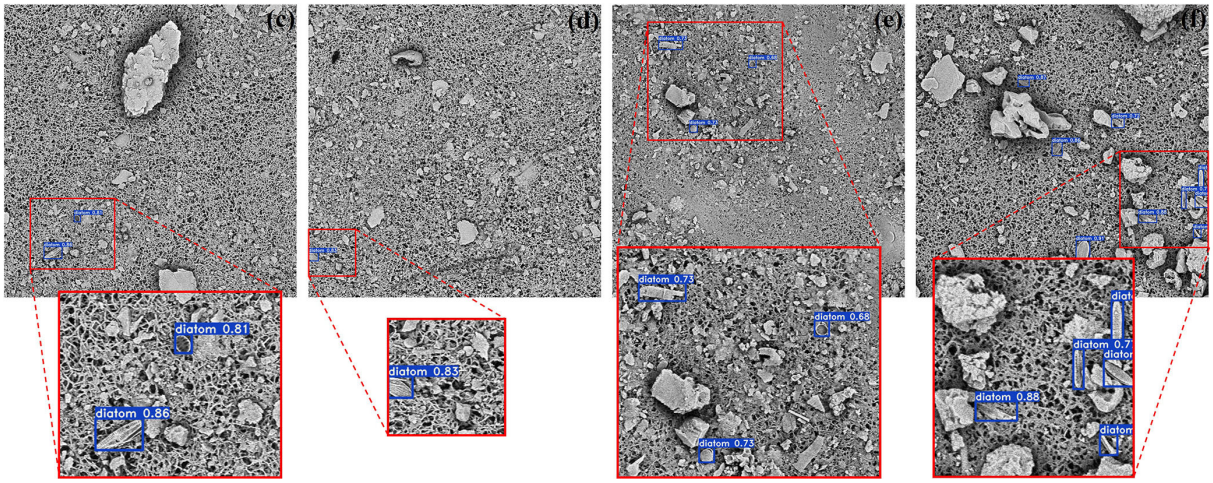


FIGURE 7
The demonstration of the quantitative results of the lung samples **(a,b)**. Several qualitative cases achieved by the Best-800 model **(c–f)**.

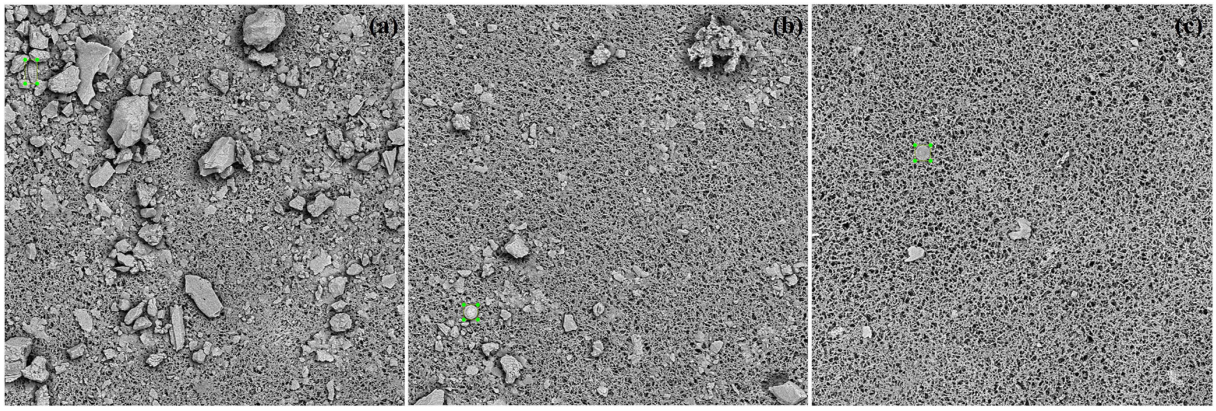
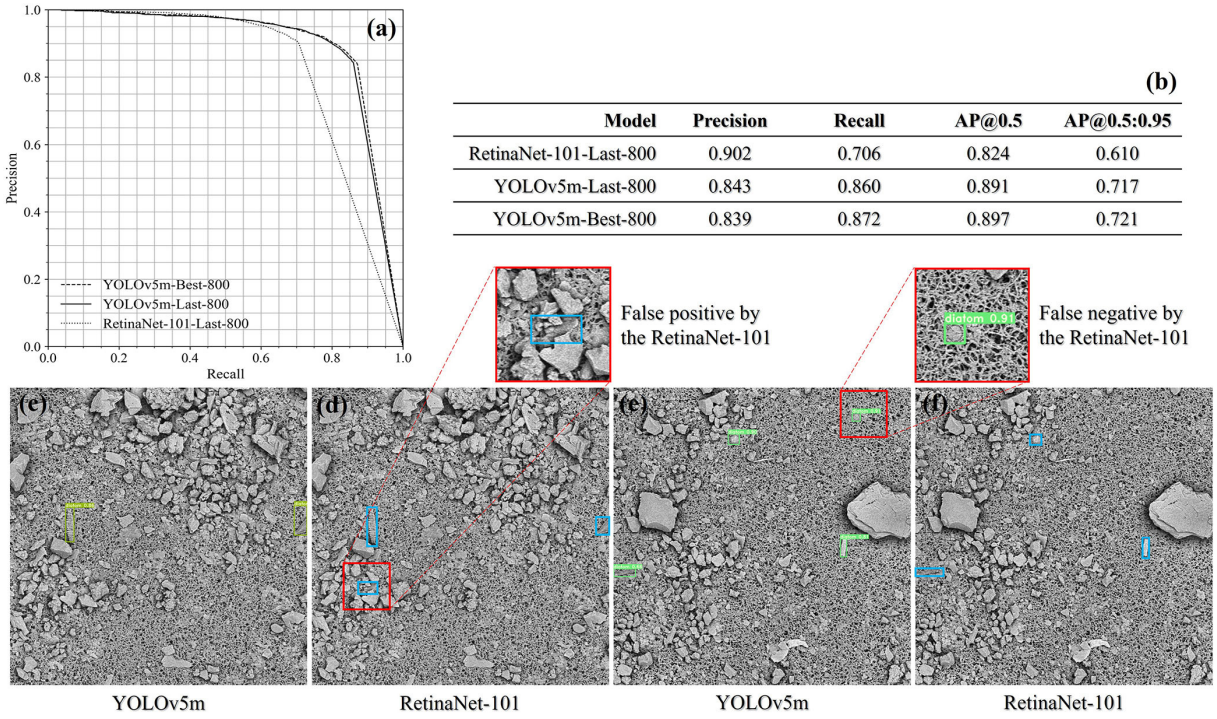


FIGURE 8
(a–c) Three annotated images acquired from the liver and kidney samples with different situations on background.

we notice that the precision and the recall get closed at a low confidence threshold of 0.1 for the RetinaNet-101-Last-800 model, while the YOLOv5m-Last-800 model approaches the balance at the threshold 0.5, which is preferable for practice. In conclusion, the results in Tables 5, 6 indicate that the training in distinguishing the diatoms from the other sediments in the



Model	Precision	Recall	AP@0.5	AP@0.5:0.95
RetinaNet-101-Last-800	0.902	0.706	0.824	0.610
YOLOv5m-Last-800	0.843	0.860	0.891	0.717
YOLOv5m-Best-800	0.839	0.872	0.897	0.721

FIGURE 9
The evaluation result summary of the liver and kidney samples **(a,b)** and two drowning cases qualitatively compared between the YOLOv5m-Last-800 model and the RetinaNet-101-Last-800 model **(c–f)**.

images is much more challenging than learning of predicting the diatom locations.

Discussion

In this article, we proposed an AI solution to assist the diatom test for searching drowning forensic evidence. Inspired by the huge success of deep learning in various domains (LeCun et al., 2015), we collected the samples from different sources and generated a large image dataset with the microwave digestion and vacuum filtration pre-processing steps (Zhao et al., 2013, 2017) and the image acquisition by a desktop scanning electron microscopy to train our deep learning-based diatom detection models. We adapted the YOLOv5 which is an engineering-optimized version of a well-known object detection architecture (Redmon et al., 2016). If the image data used for training is accompanied by the bounding box annotation, as well as the genus label for every diatom, we can train the multi-class diatom recognition model to predict not only the location of a diatom candidate but also its most possible genus.

As discussed in the "materials and methods—data acquisition" section, the collected samples include three groups for different evaluation purposes *via* cross-validation. All the scanned images have the same size 1,024 × 1,024 and almost all of them were acquired at 1,500× magnification except a liver sample (800× magnification) inherited from our previous work (Yu et al., 2021). For the lab-grown samples of the five specific diatom genera, we evaluated the capabilities of both single-class diatom detection and multi-class diatom recognition. For the former, we tried to achieve the upper limit of the YOLOv5m model considering that the images in this group suffer less from the pollution of impurities. As a result, a recall score of 0.95 together with a precision score of 0.9 are achieved by setting the IoU threshold at 0.5, the confidence threshold at 0.4, and the AP@0.5 score around 0.95. For the latter, it is more challenging due to the extra diatom taxonomy. In conformity with the results reported in the last section, we achieved the best recall score of about 0.92 when the corresponding precision score is 0.9. The difficulty in recognizing the genus of each diatom is not the same, and we observed that the *Nitzschia* and *Synedra* are easy to be misidentified with each other while almost all the false positives are from these two genera.

Both the lung samples and the samples of the liver and kidney tissues were extracted from the drowning cases. We conducted experiments on these samples to estimate the performance of our AI solution in the general situations encountered in the drowning forensic routine. Some interference labels and the distribution of the diatom genera from both groups make the multi-class diatom recognition evaluation not applicable, we therefore care only about the diatom detection issue. Especially, the experiments on the lung

TABLE 5 The precision and recall scores when changing the IoU confidence threshold from 0.1 to 0.5 and the confidence score is always 0.5.

IoU threshold	YOLOv5m-Last-800		RetinaNet-101-Last-800	
	Precision	Recall	Precision	Recall
0.1	0.843	0.858	0.905	0.705
0.2	0.843	0.859	0.905	0.705
0.3	0.843	0.860	0.905	0.705
0.4	0.843	0.860	0.905	0.706
0.5	0.843	0.860	0.902	0.706

TABLE 6 The precision and recall scores by changing the confidence threshold from 0.1 to 0.5 while the IoU score is fixed at 0.5.

Confidence threshold	YOLOv5m-Last-800		RetinaNet-101-Last-800	
	Precision	Recall	Precision	Recall
0.1	0.707	0.914	0.778	0.800
0.2	0.764	0.899	0.847	0.764
0.3	0.796	0.887	0.874	0.737
0.4	0.819	0.875	0.891	0.723
0.5	0.843	0.860	0.902	0.706

samples are designed to evaluate the worst cases due to the existence of various sediments, which indicates the lower limit of accuracy we can potentially achieve in those real cases. In the controlled study of resizing the original image to 800 × 800 and 640 × 640, respectively, as input for training and testing, the best recall score is above 0.83 at the confidence threshold of 0.5 and the corresponding precision score is around 0.7 when the input size is 800 × 800. Also, the AP@0.5 score can reach 0.8. The precision score of 0.7 indicates that there are many false positives that are common for the cases of lung samples, and the candidates predicted from the current model can be further refined by an individual AI model.

On the other hand, we tested the performance of the trained YOLOv5m models on the given liver and kidney images. In comparison to the RetinaNet-101 architecture (Lin et al., 2017) adopted in our previous work, we conducted a 4-fold cross-validation for both frameworks under the same threshold settings. When the confidence threshold and the IoU threshold are both 0.5, the precision score achieved by the RetinaNet-101 (∼0.9) is higher than the score from the YOLOv5m (∼0.84), while the recall score achieved by the RetinaNet-101 (∼0.71) is much lower than the score from the YOLOv5m (∼0.87), and the AP@0.5 score of the YOLOv5m can be almost 0.9. Moreover, a balance between the precision score and the recall score was achieved at the confidence threshold of 0.5 and 0.1 corresponding to the YOLOv5m and the RetinaNet-101, which demonstrates the superiority of the YOLOv5m since the RetinaNet-101 is prone to be tilted to the precision side. Besides the mentioned experiments, we also trained a YOLOv5m model using all the images from the liver and kidney samples and deployed it in application software for forensic practice.

In future, we will aim at the completion of our diatom detection and recognition solution by integrating the function of multi-class diatom recognition into it. Since an even distribution of the diatom genera is a prerequisite for training, the annotation on the newly scanned images from more samples is required while there is another possible way for the same purpose by generating many synthetic training images. The multi-class diatom recognition function can be built on either an end-to-end method or a hierarchical strategy, which has been on schedule to be explored.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

WY and QXi presented concepts, methods and models, wrote the original draft, and wrote and edited the revised manuscript. YH conducted data curation, implemented computer code, and supported algorithms. YD, XK, DZ, HS, QXu, and ZL conducted the experiment and analyzed the data. CL, JZ, and YN offered resources, supervised, and wrote-reviewed and edited the article. All authors read and approved the final manuscript.

Funding

This study was financially supported by the Grant-in Aids for Scientific Research from the Ministry of Public Security of the People's Republic of China (2020GABJC38, CL), and a grant from the Guangzhou Municipal Science and Technology Project, 2019030001, JZ. Guangzhou Municipal Science and Technology Project, 2019030011, ZL. Guangzhou Municipal Science and Technology Project, 2019030012, CL.

Conflict of interest

Author WY is employed by Jiangsu JITRI Sioux Technologies Co., Ltd. Author YH was employed by LabWorld (Suzhou) Intelligent Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer DL declared a shared affiliation with one of the author YD to the handling editor.

Publisher's note

References

Bueno, G., Déniz, O., Ruiz, S. J., Olenici, A., and Borrego, R. M. (2018). "Lights and pitfalls of convolutional neural networks for diatom identification," in *Optics, Photonics and Digital Technologies for Imaging Applications*, 10679.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Li, F. F., et al. (2009). "ImageNet: a large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition* (Miami, FL: IEEE), 248–255.

Deng, J. H., He, D. D., Zhuo, J. H., Zhao, J., Xiao, C., Kang, X. D., et al. (2020). Deep learning network-based recognition and localization of diatom images against complex background. *Nan Fang Yi Ke Da Xue Xue Bao* 40, 183–189. doi: 10.12122/j.issn.1673-4254.2020.02.08

Fischer, S., and Bunke, H. (2001). "Automatic identification of diatoms using decision forests," in *Machine Learning and Data Mining in Pattern Recognition. MLDM 2001. Lecture Notes in Computer Science, Vol. 2123*, ed P. Perner (Berlin; Heidelberg: Springer).

Girshick, R., Donahue, J. T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition* (Columbus, OH: IEEE), 580–587.

Gloria, B., Oscar, D., Anibal, P., Jesús, R. S., Jesús, S., Gabriel, C., et al. (2017). Automated diatom classification (Part A): handcrafted feature approaches. *Appl. Sci.* 7, 753. doi: 10.3390/app7080753

Jalba, A. C., Wilkinson, M. H. F., Roerdink, J. B. T. M., Bayer, M. M., and Juggins, S. (2001). Automatic diatom identification using contour analysis by morphological curvature scale spaces. *Mach. Vis. Appl.* 16, 217–228. doi: 10.1007/s00138-005-0175-8

Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. doi: 10.1126/science.aaa8415

Karen, S., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *preprint arXiv 1409–1556*. doi: 10.48550/arXiv.1409.1556

Kaushik, N., Pal, S. K., Sharma, A., and Thakur, G. (2017). Role of diatoms in diagnosis of death due to drowning: case studies. *Int. J. Med. Toxicol. Forensic Med.* 7, 59–65. doi: 10.22037/ijmtfm.v7i1(Winter).14047

Kloster, M., Langenkämper, D., Zurowietz, M., Beszteri, B., and Nattkemper, T. W. (2020). Deep learning-based diatom taxonomy on virtual slides. *Sci. Rep.* 10, 14416. doi: 10.1038/s41598-020-71165-w

Krause, L. M. K., Koc, J., and Rosenhahn, B. A. (2020). Fully convolutional neural network for detection and counting of diatoms on coatings after short-term field exposure. *Environ. Sci. Technol.* 54, 10022–10030. doi: 10.1021/acs.est.0c01982

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lin, T. Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *IEEE International Conference on Computer Vision* (Venice: IEEE), 2999–3007.

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8759–8768.

Ludes, B., Coste, M., North, N., Doray, S., Tracqui, A., Kintz, P., et al. (1999). Diatom analysis in victim's tissues as an indicator of the site of drowning. *Int. J. Legal Med.* 112, 163–166. doi: 10.1007/s004140050224

Paul, V., and Jones, M. (2001). Robust real-time object detection. *Int. J. Comput. Vis.* 57, 137–154. doi: 10.1023/B:VISI.0000013087.49260.fb

Pollanen, M. S., Cheung, C., and Chiasson, D. A. (1997). The diagnostic value of the diatom test for drowning, I. Utility: a retrospective analysis of 771 cases of drowning in Ontario. Canada. *J. Forensic Sci.* 42, 281–285. doi: 10.1520/JFS14111J

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 37, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Safavian, S. R., and Landgrebe, D. A. (1991). survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* 21, 660–674. doi: 10.1109/21.97458

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 2818–2826.

Yu, W. M., Xue, Y., Knoops, R., Yu, D. Y., Balmashnova, E., Kang, X. D., et al. (2021). Automated diatom searching in the digital scanning electron microscopy images of drowning cases using the deep neural networks. *Int. J. Legal Med.* 135, 497–508. doi: 10.1007/s00414-020-02392-z

Zhang, J., Zhou, Y. Y., Vieira, D. N., Cao, Y. J., Deng, K. F., Cheng, Q., et al. (2021). An efficient method for building a database of diatom populations for drowning site inference using a deep learning algorithm. *Int. J. Legal Med.* 135, 817–827. doi: 10.1007/s00414-020-02497-5

Zhao, J., Liu, C., Bardeesi, A. S. A., Wu, Y. D., Ma, Y. B., Hu, S. L., et al. (2017). The diagnostic value of quantitative assessment of diatom test for drowning: an analysis of 128 water-related death cases using microwave digestion-vacuum filtration-automated scanning electron microscopy. *J. Forensic Sci.* 62, 1638–1642. doi: 10.1111/1556-4029.13455

Zhao, J., Liu, C., Hu, S. L., He, S. W., and Lu, S. Y. (2013). Microwave digestion-vacuum filtration-automated scanning electron microscopy as a sensitive method for forensic diatom test. *Int. J. Legal Med.* 127, 459–463. doi: 10.1007/s00414-012-0756-9

Zhou, Y. Y., Cao, Y. J., Huang, J., Deng, K. F., Ma, K. J., Zhang, K. Y., et al. (2020). Research advances in forensic diatom testing. *Forensic Sci. Res.* 5, 98–105. doi: 10.1080/20961790.2020.1718901

Zhou, Y. Y., Zhang, J. J., Deng, K. F., Zhang, J. H., Qin, Z. Q., et al. (2019). Digital whole-slide image analysis for automated diatom test in forensic cases of drowning using a convolutional neural network algorithm. *Forensic Sci. Int.* 302, 109922. doi: 10.1016/j.forsciint.2019.109922

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 21 September 2022
DOI 10.3389/fmicb.2022.971437

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Jinghua Zhang,
National University of Defense Technology,
China
Stefano Amalfitano,
National Research Council (CNR), Italy

*CORRESPONDENCE
Qiuwen Chen
qwchen@nhri.cn

SPECIALTY SECTION
This article was submitted to
Aquatic Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 June 2022
ACCEPTED 15 August 2022
PUBLISHED 21 September 2022

CITATION
Pan B, Liu X, Chen Q, Sun H, Zhao X and Huang Z (2022) Hydrological connectivity promotes coalescence of bacterial communities in a floodplain.
Front. Microbiol. 13:971437.
doi: 10.3389/fmicb.2022.971437

Hydrological connectivity promotes coalescence of bacterial communities in a floodplain

Baozhu Pan[1], Xinyuan Liu[1], Qiuwen Chen[2]*, He Sun[3], Xiaohui Zhao[1] and Zhenyu Huang[1]

[1]State Key Laboratory of Eco-hydraulic in Northwest Arid Region of China, Xi'an University of Technology, Xi'an, Shaanxi, China, [2]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, China, [3]State Key Laboratory of Crop Stress Biology in Arid Areas, Shaanxi Key Laboratory of Agricultural and Environmental Microbiology, College of Life Sciences, Northwest A&F University, Yangling, Shaanxi, China

Floodplains play essential roles in the ecological functions of regional environments. The merging and coalescence of bacterial communities in aquatic environments results in periodic patterns driven by regular hydrological activities, which may, in turn, influence ecological activities. However, the degree of bacterial community coalescence in the lateral and vertical directions as well as the underlying hydrological mechanism of floodplain ecosystems is poorly understood. Therefore, we investigated the spatiotemporal patterns and coalescence processes of planktonic and sedimentary bacterial communities during normal and high-water periods in a floodplain ecosystem of the Yellow River source region. We classified bacterial operational taxonomic units (OTUs) based on 16S rRNA gene sequencing, and quantified community coalescence by calculating the proportions of overlapping OTUs, the contributions of upstream sources to downstream sinks, and positive/negative cohesion. The results revealed major differences in the composition and diversity of planktonic and sedimentary bacterial communities. Bacterial community diversity in the high-water period was higher than in the normal period. Laterally, hydrological connectivity promoted the immigration and coalescence of bacterial communities to oxbow lakes in both the mainstream and tributaries, with the coalescence degree of planktonic bacteria (2.9%) higher than that of sedimentary bacteria (1.7%). Vertically, the coalescence degree of mainstream planktonic and sedimentary bacterial communities was highest, reaching 2.9%. Co-occurrence network analysis revealed that hydrological connectivity increased the complexity of the bacterial network and enhanced the coalescence of keystone species to oxbow lakes. Furthermore, community coalescence improved the competitiveness and dispersal of bacterial communities. This study demonstrated that coalescence of bacterial communities is driven by hydrological connectivity in a floodplain ecosystem. Further studies should investigate the processes of bacterial community coalescence in floodplains in more detail, which could provide new approaches for environmental protection and ecological function preservation.

KEYWORDS

community coalescence, co-occurrence network, hydrologic connectivity, river floodplain, planktonic bacteria, sedimentary bacteria

Introduction

Floodplains are alluvial complexes comprising interconnected biota and ecological gradients. Being extremely vital ecosystems (Argiroff et al., 2017; Wang et al., 2020), floodplains contribute to preserving biodiversity (Naiman et al., 1993), maintaining water quality (Mitsch et al., 2001; Tockner et al., 2010) and handling flood surges (Kousky and Walls, 2014). Floodplains contain a vertical tree-like network of mainstream and tributaries (Mansour et al., 2018). In addition, numerous ecological niches are present as oxbow lakes, formed by shore erosion and overflow floods, which are seasonally separated from the original rivers (Durkin et al., 2015; Wang et al., 2020). During the normal period, oxbow lakes are partitioned from the mainstream, exhibiting a high spatial heterogeneity (Mayora et al., 2020). However, during the high-water period, the rising water of the mainstream will flood the floodplain between the mainstream and the oxbow lake, and the mainstream and the oxbow lake are connected. As a result, most environment of mainstream and oxbow lake displays typical equilibrium effects (Mayora et al., 2013, 2020). Hydrological connectivity not only drives matter and energy flows laterally and vertically, but also maintains the spatiotemporal heterogeneity of microbial community structure in riverine networks (Mansour et al., 2018).

Bacteria constitute a substantial part of microbial communities and play a paramount role in biogeochemical processes and nutrient cycling in aquatic ecosystems (Findlay, 2010; Madsen, 2011; Zhang et al., 2022a,b). According to their habitat preferences in rivers and lakes, bacterial communities can be divided into planktonic and sedimentary. The planktonic bacterial community is the sum of the sources of upstream bacteria, including rainfall, lake water, groundwater, and soil water, and it is susceptible to compositional and structural variations (Liu et al., 2018). The sedimentary bacterial community is formed through long-term sediment erosion and accumulation (Qian et al., 1987), and it is sensitive to environmental disturbances (Labbate et al., 2016; Zeng et al., 2019). There are also diversity and compositional differences between the two different bacterial communities (Jiang et al., 2006; Liu et al., 2018; Zeng et al., 2019). Planktonic bacteria can flow to the downstream and benthic zone, where they coalesce with sedimentary bacteria (Mansour et al., 2018; Gao et al., 2021). However, the extent to which these bacterial communities merge and coalesce in aquatic ecosystems is still not fully understood (Mansour et al., 2018; Langenheder and Lindström, 2019). This question includes the merging and coalescence of the same bacterial community in different aquatic environments, and the merging and coalescence of different communities in the same environment.

From an ecological perspective, a community coalescence event is more than just a part of a dispersal process, and it results in interactions between the whole community and its environment (Rillig et al., 2015). Meanwhile, community coalescence is an exchange event among communities (and the surrounding environments); that is, individual communities coalesce with a new entity under mixing of relatively large environments (Rillig et al., 2015). By contrast, bacterial dispersal encompasses the immigration and establishment of individuals (Hanson et al., 2012). In recent years, coalescence of bacterial communities has received increasing attention, including the construction of theoretical frameworks (Rillig et al., 2015; Mansour et al., 2018), verification by microcosmic experiments or mathematical algorithms (Livingston et al., 2013; Rillig and Mansour, 2017), significance in biological evolution (Castledine et al., 2020), and quantitative extent of community coalescence (Zhou and Ning, 2017; Mei and Liu, 2019). However, comprehensive studies elucidate the distribution patterns and ecological significance of bacterial community coalescence in natural habitats are still limited. Consequently, how hydrological connectivity influences bacterial community coalescence in floodplain ecosystems remains an open question.

Coalescence is a community assembly process involving settlement and interactions of species (Castledine et al., 2020). Co-occurrence network analysis is commonly used to explore interactions among species and to ascertain the importance of certain species (Röttjers and Faust, 2018). Co-occurrence networks cannot always illustrate a real biological connection (Freilich et al., 2018; Qiu et al., 2021). Nonetheless, co-occurrence network analysis can visualise the complexity of bacterial communities, and identify which taxa are more important than others for maintaining the network structure (Qiu et al., 2021; Yuan et al., 2021). The network structure of bacterial communities in rivers is influenced by water environmental factors (Peng et al., 2017). Community coalescence is bound to affect the complexity of the bacterial network, the number of keystone species, and the connectivity among species. Thus, network analysis can be used as a tool to determine the possible influence of bacterial community coalescence on interspecies interactions under variable hydrological connectivity.

The present study was conducted in a floodplain ecosystem in the source region of the Yellow River, China. We analyzed the merging and coalescence of planktonic and sedimentary bacterial communities in vertical and lateral directions during different hydrological periods. We hypothesized that coalescence of planktonic and sedimentary bacterial communities occurs in the floodplain during the normal period,

and would be enhanced in vertical and lateral directions by increased hydrological connectivity during the high-water period; in this way, hydrological connectivity positively influences the network complexity of bacterial communities and community coalescence, ecologically. To verify the hypothesis, we studied the distribution patterns of different bacterial communities at multiple spatiotemporal scales, quantified the extent of community coalescence, and investigated the influence of community coalescence on bacterial networks. This study was designed to explore the following: (1) Why are there differences in the spatial distribution patterns of bacterial communities between normal and high-water periods? (2) How does hydrological connectivity influence lateral and vertical coalescence of bacterial communities, in addition to the community structure and keystone species? (3) How does community coalescence improve the stability of bacterial communities in the floodplain?

Materials and methods

Study area and sampling

The study area (102°00′–103°00′E, 33°00′–33°30′N) is located in the Baihe River Basin in Hongyuan County (Yellow River source region), Aba Tibetan and Qiang Autonomous Prefecture, Sichuan Province, Southwest China. The Baihe River has a large number of tributaries and oxbow lakes, providing a natural observation window for this study. We classified the Baihe River into three types of water bodies (i.e., mainstream, tributaries, and oxbow lakes) based on their connectivity. A total of 36 sampling sites were selected along the Baihe River, with 10 in the mainstream, 14 in the tributaries, and 12 in the oxbow lakes (Figure 1).

Considering the influence of river connectivity on the migration and spread of bacterial communities, paired samples of surface water (0.5 m depth) and surface sediment (0.05 m depth) were collected in September 2019 (normal period: the runoff of the Baihe River is 41.81 ± 23.49 m^3/s) and June 2020 (high-water period: the runoff is 164.17 ± 136.03 m^3/s). In each season, sampling was completed within a 5-day period. At each sampling site, water samples (10 L each) were collected using two 5 L sterile polyethylene terephthalate bottles and kept at a low temperature of 0°C–4°C. Meanwhile, three sediment samples were collected near the water sampling site and mixed to form a composite sample, which was sealed in 50 ml sterile polypropylene tubes and kept in liquid nitrogen. All samples were immediately transported to the laboratory, where water samples were filtered through a 0.22 μm polycarbonate membrane (47 mm diameter; Millipore, Billerica, MA, United States). The filtered membranes and sediment samples were stored at-80°C until DNA extraction. A total of 140 samples (72 water and 68 sediment) were collected.

Environmental information

A total of 23 environmental variables were measured or collected (Supplementary File 2; Supplementary Table S1). Nine of the environmental variables were measured in the field. Specifically, flow velocity (V) was measured using an FP211 direct-reading flow meter (Global Water Instrumentation, Sunnyvale, CA, United States). Water quality parameters, namely water temperature (WT), electrical conductivity (EC), dissolved oxygen (DO), pH, oxidation–reduction potential (ORP), and total dissolved solids (TDS), were measured using a portable multi-parameter analyzer (YSI Corp., Yellow Springs, OH, United States). Turbidity (Tur) was measured with a 2100Q portable turbidity meter (Hach, Loveland, CO, United States), and mud temperature (MT) was measured using a DS600T mud thermometer (EDKORS, Changzhou, Jiangsu Province, China).

Another 14 environmental variables were determined in the laboratory. For water samples, chemical oxygen demand (COD) was determined by fast-digestion spectrophotometry based on the Chinese Environmental Protection Industry Standard for Water Quality (HJ/T 399-2007), and dissolved organic carbon (DOC) was determined by combustion based on the International Standard for Water Quality (ISO 8245-1987). Total phosphorus (TP), total nitrogen (TN), ammonium-nitrogen (NH_4-N) and nitrate-nitrogen (NO_3-N) of water samples were determined by spectrophotometry according to standard methods described in "Water and Wastewater Monitoring and Analysis Methods" (Third Edition). Levels of chlorophyll *a* (Chl-*a*) in water samples were determined by spectrophotometry after extraction with 95% ethanol according to "Specifications for Lake Eutrophication Investigation" (Second Edition). Soil total nitrogen (STN), total phosphorus (STP) and organic carbon (SOC) were also determined based on the Chinese Environmental Protection Industry Standards for Soil Quality (HJ 717-2014, HJ 632-2011 and HJ 615-2011, respectively). Sediment particle size was measured using a Mastersizer 2000 Laser Particle Sizer (Malvern Instruments Ltd., Worcestershire, United Kingdom) with a working range of 0.02–2,000 μm and relative error vc < 1%. Median particle size (D_{50}) was obtained after drawing a gradation curve. Sediment type with a grain size was classified as clay (particle size <4 μm, 8Φ), silt (4–63 μm, 4–8Φ) and sand (>63 μm, 4Φ; Huang et al., 2010).

Illumina sequencing and bioinformatics analysis

Genomic DNA was extracted in duplicate using a FastDNA SPIN Kit (MP Biomedicals, Santa Ana, CA, United States) according to the manufacturer's protocols. Duplicate DNA extracts were pooled for subsequent PCR amplification on a BioRad S1000 (Bio-Rad Laboratory, Hercules, CA, United States), targeting the hypervariable V4 region of the bacterial 16S ribosomal RNA (rRNA) gene. Each DNA sample was amplified

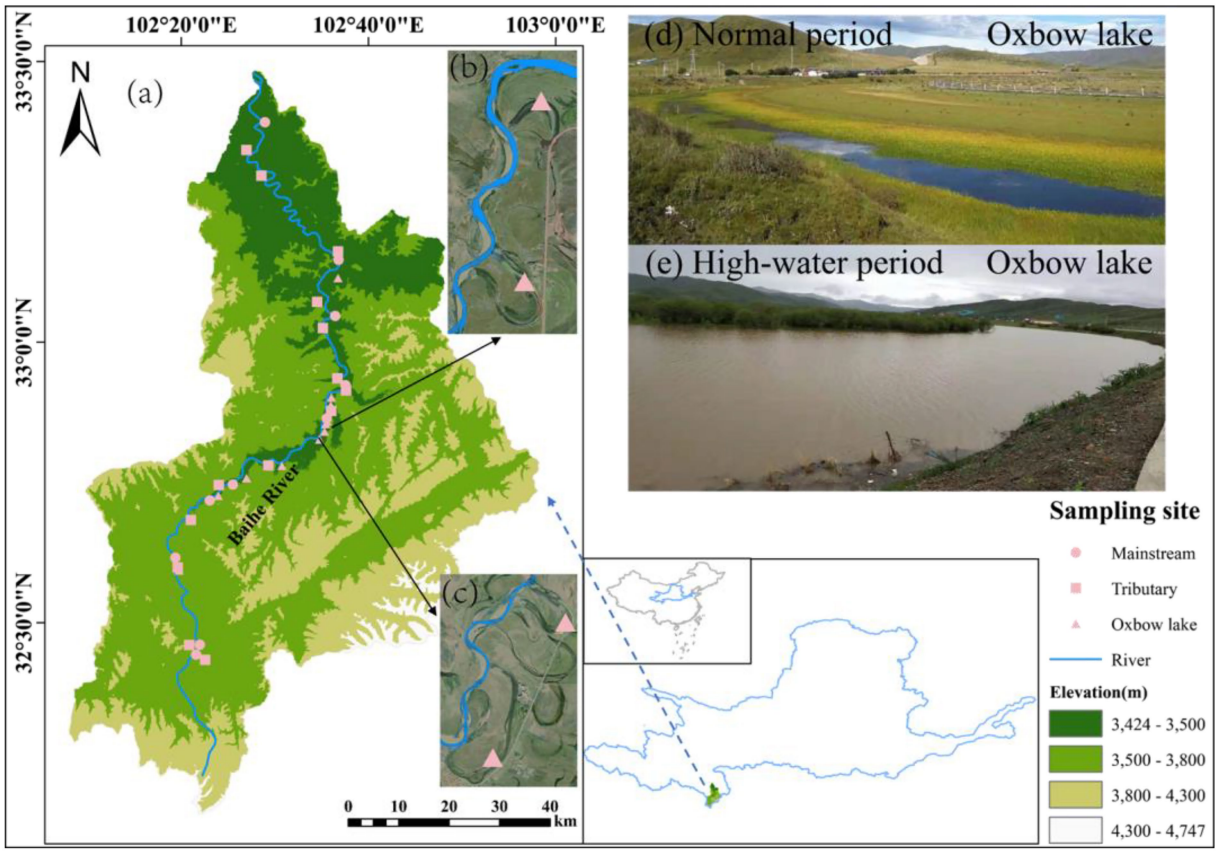


FIGURE 1
Locations of sampling sites in Baihe River in the source region of the Yellow River, China. **(A)** Baihe River Basin. **(B)** Sampling sites at oxbow lakes 5 and 6. **(C)** Sampling sites at oxbow lakes 7 and 8. **(D)** Oxbow lake in normal period. **(E)** Oxbow lake in high-water period.

using primers 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACNVGGGTWTCTAAT-3′; Invitrogen, Carlsbad, CA, United States). PCR amplifications contained 25 μl of 2 × Premix Taq (Takara Biotechnology, Dalian, Liaoning Province, China), 1 μl of each primer (10 mM) and 3 μl of sample DNA (20 ng/μl). Thermal cycling included an initial denaturation at 94°C for 5 min, followed by 30 cycles of 30 s at 98°C, 30 s at 52°C, and 30 s at 72°C, and a final extension step of 10 min at 72°C. Triplicate PCR products for each of the 140 samples were purified using an AxyPrep DNA Gel Extraction Kit (Axygen Biosciences, Union City, CA, United States). All libraries were sequenced on an Illumina MiSeq platform (Illumina Inc., San Diego, CA, United States) using a paired-end (2 × 250 bp) approach. The raw 16S rDNA sequence data have been stored in a public National Center for Biotechnology Information (NCBI) database (accession number: PRJNA853875).

Sequences of bacterial 16S rRNA gene amplicons were quality-filtered using QIIME v2.0 (Li et al., 2019) following the official suggestions, and detailed processes can be found elsewhere (Gao et al., 2020). High-quality sequence data were checked and corrected using DADA2 to obtain operational taxonomic units (OTUs) with a sequence similarity of 100% (Gao et al., 2021). Taxonomic annotation of OTUs was assigned using the Naive Bayes classifier trained by the Silva (SSU132) 16S rRNA database (Hoyningen-Huene et al., 2019). In order to prevent sequencing errors in subsequent analyses, all sequences classified as chloroplasts, mitochondria, archaea, or eukaryotes were removed (Mo et al., 2018). Furthermore, to minimize the influence of unequal sequencing efforts, random sampling was conducted on an ESV table to equalize the number of sequences in each sample ($n = 9{,}315$).

Data analysis

Alpha-and beta-diversity

We calculated alpha-diversity (i.e., OTU richness, Chao1 and Shannon-Wiener indices) of bacterial communities for each sample using vegan version 2.5-7 with R program version 4.1.0 (Chen et al., 2019). One-way analysis of variance (ANOVA) and Student's *t*-test were used to compare alpha-diversity between groups in SPSS version 25.0 (IBM Corp., Armonk, NY, United States). For beta-diversity, bacterial community composition was visualized using non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarities, and analysis of similarity (ANOSIM) was used to evaluate differences in

bacterial communities between groups (Mo et al., 2021). These were implemented using the R program (version 4.1.0) with vegan (version 2.5-7), ggplot2 (version 3.3.5), and RColorBrewer (version 1.1-2) packages.

Community coalescence

We used three standard methods to evaluate the immigration and coalescence of bacterial communities. First, R version 4.1.0 was used to calculate the overlap of species (proportion of shared species, or number of reads of common OTUs) between adjacent communities (Gao et al., 2021). Second, the Bayesian classifier SourceTracker was used to predict the contributions of different types of upstream sources to different types of downstream sinks (Knights et al., 2011). Finally, the helperfunctions.r and calcCohesion.r packages were used to quantify the connectivity between communities (Herren and McMahon, 2017).

Habitat niche breadth

We calculated the Levins' niche breadth (B) index for bacterial communities using the following formula:

$$B_j = \frac{1}{\sum_{i=1}^{N} P_{ij}^2}$$

where B_j indicates the habitat niche breadth of OTU j in a metacommunity, N represents the total number of communities in each metacommunity, and P_{ij} is the proportion of OTU j in community i (Wu et al., 2018). A high B value represents a wide habitat niche breadth. It is generally believed that at the community level, the wider the niche, the broader the distribution and the larger the number of species, and vice versa (Jiao et al., 2020). The calculation was implemented using the R spaa package (version 0.2.2; Zhang, 2016).

Co-occurrence network

The OTU distribution patterns in samples of normal and high-water periods were displayed across the taxonomic tree by directed networks using the *prefuse* layout algorithm in CYTOSCAPE v3.7.1 (Faust and Raes, 2016). We selected prevalent OTUs (present in ≥20% of samples) among samples in the same habitat type as nodes to prevent inconsistent trends caused by transient OTUs (Liu et al., 2018). The network topology of each sample was characterized using the *subgraph* function *via* the R igraph package (Ma et al., 2016), in terms of node number (the number of OTUs), edge number (the number of connections among all nodes), average path length (APL, average shortest path length between any two nodes in the network), and betweenness (the number of times a node acts as a bridge along the shortest path between two other nodes). Higher node number, edge number and APL and lower betweenness represent greater network complexity (Gao et al., 2021; Qiu et al., 2021). Identification of keystone species was based on calculation of within-module connectivity (Z_i) and among-module connectivity (P_i) in the co-occurrence network (Guimerà et al., 2005). Excluding peripherals ($Z_i < 2.5$, $P_i < 0.62$), the other three types of nodes (module hubs, connectors and network hubs) were classified as keystone species (Deng et al., 2012; Shi et al., 2016). Visualisation of the co-occurrence network was performed using Gephi version 0.9.2.[1]

Results

Comparison of environmental factors between normal and high-water periods

Approximately half of the 23 environmental variables in floodplain showed a significant difference between normal and high-water periods ($p < 0.05$; Supplementary File 2; Supplementary Table S2). In the mainstream, the mean values of EC, TDS, NH_4-N, TN, and TP were all significantly higher in the normal period than the high-water period, while the opposite was true for ORP and Tur. In tributaries, the mean values of EC, DO, TDS, Si, NH_4-N and STN were significantly higher in the normal period than the high-water period, in contrast to the trends of ORP, Chl-*a* and Tur. In oxbow lakes, EC, TDS, NH_4-N, TN and TP displayed similar trends to those in the mainstream, with significantly higher mean values in the normal period than the high-water period. On the contrary, ORP, Tur, STN and STP exhibited higher mean values in the high-water period than the normal period. Overall, the mean values of EC, TDS and NH_4-N were significantly higher in the normal period than in the high-water period, while only ORP had higher mean values in the high-water period.

Relative abundances of bacterial communities

A total of 181,778 OTUs were retrieved from the 140 samples by high-throughput sequencing. The rarefaction curves revealed that the bacterial OTUs obtained from the applied sequencing depth were sufficient to represent the bacterial communities in water and sediment samples. In addition, the number of OTUs observed in different times and spaces were highly variable; the number of sedimentary bacterial OTUs was greater than that of planktonic bacterial OTUs, while the number of OTUs in the high-water period was greater than that in the normal period. Specifically, the number of OTUs in different groups were ordered sediment in the high-water period (HS) > water in the high-water period (HW) > sediment in the normal period (NS) > water in the normal period (NW; Supplementary File 1; Supplementary Figure S1).

1 https://gephi.org/

Regarding planktonic bacteria, irrespective of the season, the number of OTUs in tributaries was the highest, and the number of OTUs in oxbow lakes was the lowest. In the high-water period, the mean number of OTUs in the three different water environments was 2.4 times that in the normal period. Regarding sedimentary bacteria, in both periods, tributaries harboured the largest number of OTUs, with the fewest found in the mainstream. Similar to planktonic bacteria, the number of OTUs in the three different sedimentary environments was higher in the high-water period than in the normal period. However, the magnitude of the increase in OTUs varied in different water body types, by 2.3 times in tributaries, 1.7 times in oxbow lakes, and 1.4 times in the mainstream (Supplementary File 1; Supplementary Figure S2).

With respect to the relative abundance of major bacterial phyla, Proteobacteria accounted for the largest proportions of planktonic and sedimentary bacterial communities in the two periods, and the proportions in water were slightly larger than those in sediment. Bacteroidetes was the second dominant phylum in all samples. In addition, Chlamydiae and Cyanobacteria only existed in water, while Latescibacteria and Rokubacteria only occurred in sediment. The major bacterial phyla also shifted with season. For example, Armatimonadetes and Cyanobacteria only appeared in the normal period (water), while Nitrospirae only emerged in the high-water period (water and sediment). Compared with the planktonic bacterial community, variations in the sedimentary bacterial community were minimal between the two study periods (Figure 2).

Diversity of bacterial communities

With the exception of Good's coverage, the other five alpha-diversity indices of planktonic and sedimentary bacterial communities were all significantly higher in the high-water period than in the normal period (Supplementary File 1; Supplementary Figure S3). For the planktonic bacterial community, the five alpha-diversity indices were highest in tributaries, followed by the mainstream, and lowest in oxbow lakes in both periods. However, the five alpha-diversity indices of the sedimentary bacterial community in the three types of water bodies exhibited distinctively different trends between the two periods. In the normal period, there was little difference among the three sedimentary environments, despite slightly higher bacterial diversity in oxbow lakes and slightly lower bacterial diversity in the mainstream. In the high-water period, tributaries harboured the highest bacterial diversity, while the mainstream showed the lowest bacterial diversity, and there was a significant difference between the mainstream and the other two water body types.

The NMDS biplot shows that the bacterial communities of water samples were significantly different from those of the corresponding sediment samples in the normal period, while only partial community differences were observed in the high-water period (Figure 3). The bacterial communities of sediment samples displayed distinct seasonal variations, but the bacterial communities of water samples did not form two separated clusters for the two seasons. The consistency of the results was corroborated by ANOSIM (Supplementary File 1; Supplementary Figure S4). Both the planktonic and sedimentary

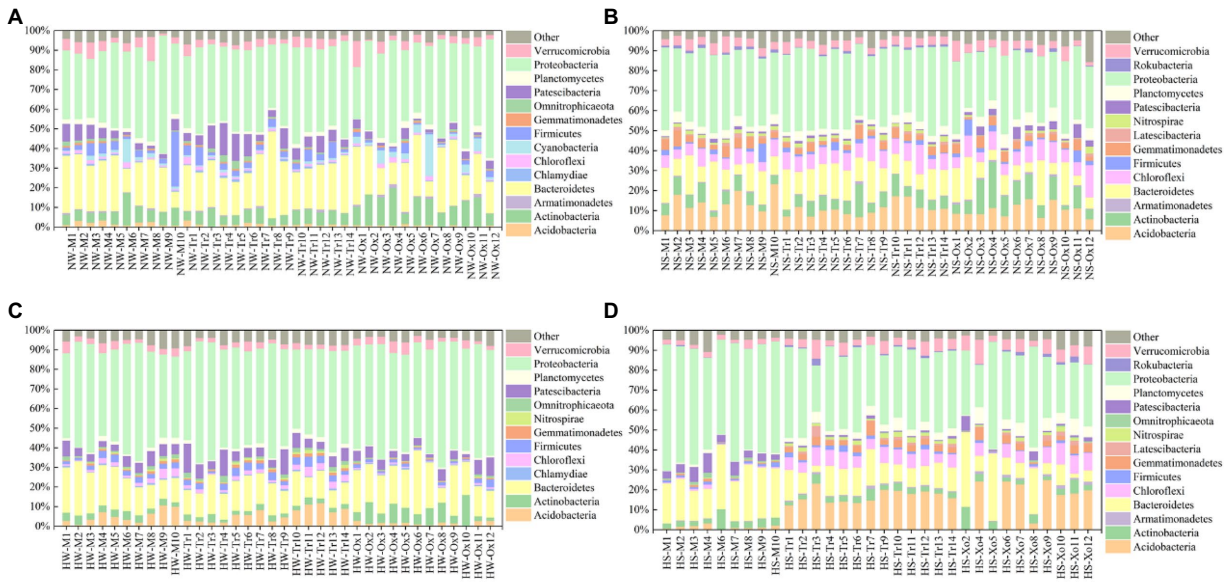


FIGURE 2
Relative abundances of major bacterial phyla in water and sediment samples of different periods. **(A)** Water in the normal period. **(B)** Sediment in the normal period. **(C)** Water in the high-water period. **(D)** Sediment in the high-water period.

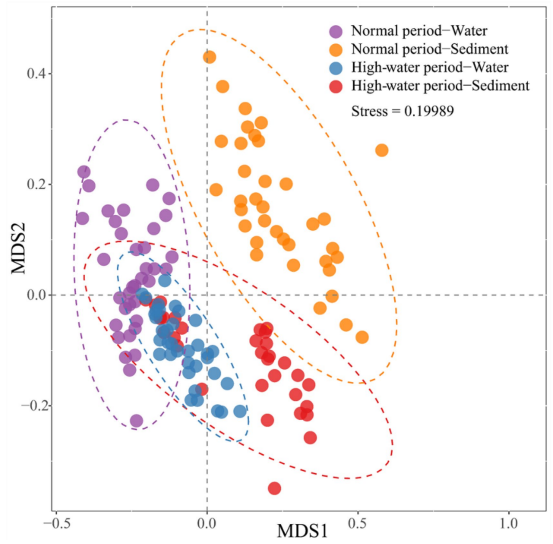


FIGURE 3
Non-metric multidimensional scaling (NMDS) biplot showing differences in bacterial community composition in water and sediment samples of Baihe River across normal and high-water periods.

bacterial communities were significantly different between normal and high-water periods (planktonic, global $r = 0.203$, $p = 0.001$; sedimentary, global $r = 0.263$, $p = 0.001$).

Furthermore, the NMDS and ANOSIM results demonstrated a clear separation of different bacterial communities (NW, NS, HW, and HS) in samples based on water body type (Supplementary File 1; Supplementary Figures S5, S6). For the normal water period, separation of the sedimentary bacterial community was clearer than that of the planktonic bacterial community; however, irrespective of the planktonic or sedimentary bacterial community, oxbow lake samples were markedly different from mainstream and tributary samples (Supplementary File 1; Supplementary Figures S5a,b). For the high-water period, separation of the planktonic bacterial community was clearer than that of the sediment bacterial community, and there were significant differences in the sedimentary bacterial community between the mainstream and tributaries (Supplementary File 1; Supplementary Figures S5c,d). In the high-water period, the planktonic bacterial community in the oxbow lakes and mainstream clustered more closely, while the sedimentary bacterial community of the mainstream and tributaries tended to be separated more clearly, compared with those in the normal period. In addition, the sedimentary bacterial community in different oxbow lake samples showed significant differences in the high-water period (Supplementary File 1; Supplementary Figure S5).

Coalescence of bacterial communities

Water and sediment from adjacent sampling sites were regarded as sources and sinks for the coalescence of bacterial communities, and default flow directions (from west to east, and from tributaries to mainstream to oxbow lakes) were taken into consideration to obtain more general and meaningful results. The detailed pairs of tributaries–mainstream–oxbow lake samples that met the upstream–downstream requirements are listed in Supplementary File 2; Supplementary Table S3.

Based on this hypothesis, the relative abundance of overlapping (shared) OTUs was calculated for each bacterial community and its neighbored upstream communities (Figure 4). Following merging of upstream–downstream bacterial communities in pairs, OTUs in water were more preserved than those in sediment, irrespective of the season. This indicates greater coalescence of planktonic bacteria than for sedimentary bacteria across different periods. Compared with the normal period, preservation of OTUs in both water and sediment was higher in the high-water period. Accordingly, there was increased connectivity between the tributaries, mainstream, and oxbow lakes in the high-water period, which promoted the integration of bacterial communities.

The same trends were found based on correlation analysis between Bray–Curtis similarity matrices of bacterial communities and cumulative dendritic distances. The planktonic bacterial community displayed a distance attenuation pattern during the high-water period ($p < 0.01$; Figure 5). In the vertical direction, the coalescence of bacterial communities in water and sediment also showed temporal and spatial differences; with increasing water level, bacterial communities in the mainstream water and sediment merged most strongly (normal period, $0.23\% \pm 0.07\%$; high-water period, $2.94\% \pm 0.35\%$; Supplementary File 1; Supplementary Figure S7).

The coalescence patterns of bacterial communities were corroborated by SourceTracker estimates (Figure 6). In the normal period, when the sink was set as the mainstream water or sediment, the planktonic and sedimentary bacterial communities in tributaries made larger contributions, respectively. Under the influence of connectivity, the bacterial communities in both the mainstream and tributaries were major contributors when the sink was set as oxbow lake water or sediment. When compared between water and sediment, there was greater coalescence between planktonic bacterial communities than between sedimentary bacterial communities. In the high-water period, the source of the bacterial community changed, and the contribution of the bacterial community in the mainstream increased for both planktonic and sedimentary bacteria.

Next, we calculated the cohesion of bacterial communities across time and space. The absolute values of both positive and negative cohesions of bacterial communities were higher in the high-water period than in the normal period, irrespective of water body type (Figure 7). The results of positive cohesion were consistent with the changes in the mean habitat niche breadth of bacterial communities (Supplementary File 1; Supplementary Figure S8). In the high-water period, the cohesion of both planktonic and sedimentary bacterial communities showed spatial differences. For example, the

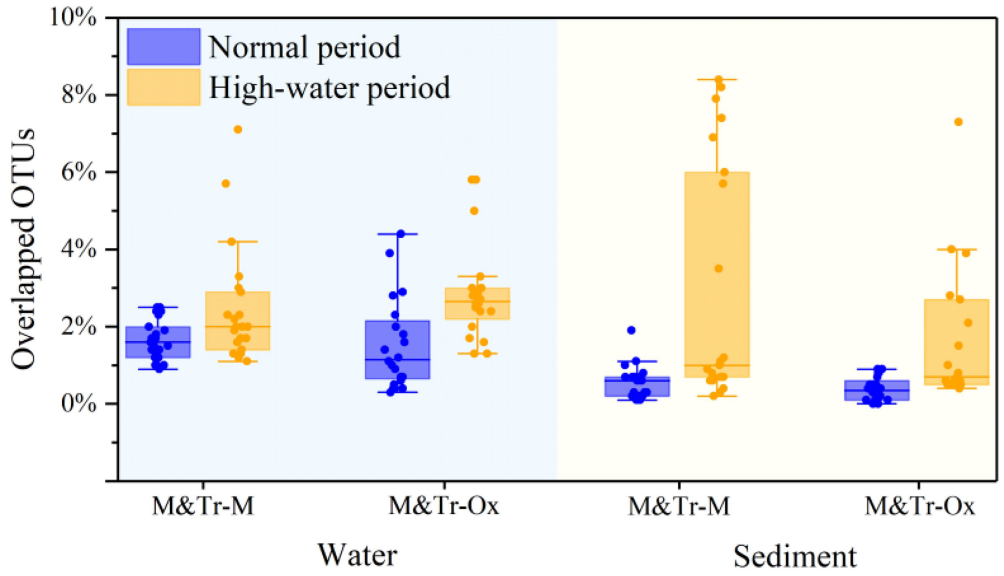


FIGURE 4
Proportions of overlapping operational taxonomic units (OTUs) between adjacent sampling sites in all OTUs of both sites in different study periods. The proportions of overlapping OTUs were used to quantify community coalescence between upstream and downstream sites. M&Tr-M represents the contribution of the upstream mainstream and tributaries to the downstream mainstream, and M&Tr-Ox represents the contribution of the upstream mainstream and tributaries to the downstream oxbow lakes. Data are means±standard deviation.

cohesion of the planktonic bacterial community in oxbow lakes reached the highest level, and the cohesion of the sedimentary bacterial community in the mainstream was much higher than that in the tributaries and oxbow lakes after flooding.

Co-occurrence patterns and keystone species of bacterial communities

The networks of bacterial communities constructed for the two different periods demonstrated distinct co-occurrence patterns (Figures 8A,B). In both periods, the betweenness of bacterial networks in water was lower than that in sediment, while their node number, edge number, and APL were all higher than those in sediment (Figures 8C–F). In the normal period, planktonic bacteria dominated the network, which played greater roles in the tributaries than in the mainstream and oxbow lakes. Compared with the normal period, the complexity of the bacterial network increased in the high-water period, and the role of sedimentary bacteria was enhanced, especially in the mainstream.

Network analysis identified 39 OTUs and 367 edges in the bacterial network of the normal period, compared with 159 OTUs and 400 edges in the bacterial network of the high-water period. In both periods, the top three phyla with the largest proportions were Proteobacteria (normal period 64.1%, high-water period 58.49%), Bacteroidetes (25.64, 25.79%) and Actinobacteria (10.26, 6.29%; Figures 8A,B). In the two networks, 39 keystone OTUs were identified for the normal period compared with 112 for the high-water period (Supplementary File 2; Supplementary Table S4). After screening, Proteobacteria, Bacteroidetes and Actinobacteria were the three most abundant phyla, regardless of the water body type, habitat environment, and season. In the normal period, the abundance of keystone OTUs in sediment was extremely low, while the keystone OTUs in water were almost twice as abundant in tributaries than in the mainstream and oxbow lakes. The abundance of keystone OTUs increased dramatically in the high-water period compared with the normal period, but the magnitude of the increase was variable across different water bodies and environments due to distinctive connectivity. The most prominent increases were observed in water of oxbow lakes and sediment of the mainstream (Supplementary File 1; Supplementary Figure S9).

Discussion

Hydrological connectivity is defined as the amount of water-mediated transfer of matter, energy and organisms within or between elements of the hydrologic cycle (Michaelides and Chappell, 2010). Although hydrological connectivity is one of the main non-biological factors driving ecological processes and organism distribution, its influence on bacterial community coalescence in floodplain ecosystems is largely unknown. In this work, we found that the extent of enrichment and the composition of planktonic and sedimentary bacterial communities vary in different water bodies of a floodplain ecosystem over normal and high-water periods, with hydrological connectivity being the crucial factor driving bacterial community coalescence.

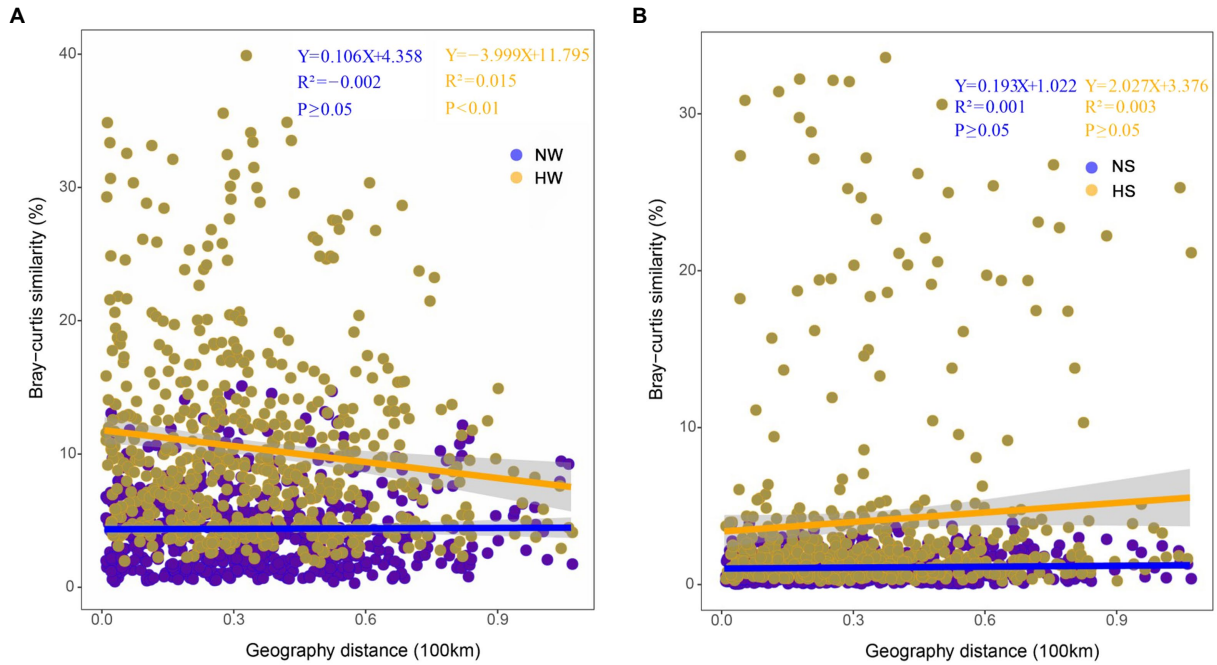


FIGURE 5
Distance-decay patterns based on the Bray–Curtis similarity of bacterial community composition and cumulative dendritic distance in different periods. **(A)** Comparison of planktonic bacterial community between the two periods. NW, water in the normal period; HW, water in the high-water period. **(B)** Comparison of sedimentary bacterial community between the two periods. NS, sediment in the normal period; HS, sediment in the high-water period.

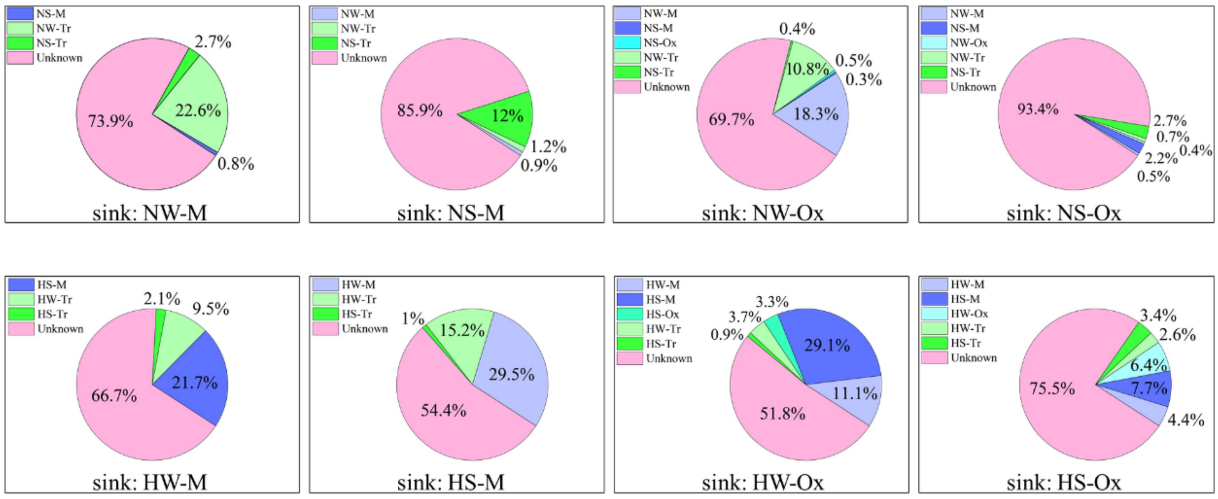


FIGURE 6
SourceTracker estimates of the contributions of source communities to sink communities of planktonic and sedimentary bacteria in the mainstream and oxbow lakes of Baihe River across different periods. NW, water in the normal period; NS, sediment in the normal period; HW, water in the high-water period; HS, sediment in the high-water period; M, mainstream; Tr, tributaries; Ox, oxbow lakes.

Bacterial communities display spatiotemporal patterns in the floodplain ecosystem

As expected, NMDS analysis revealed a separate clustering of planktonic and sedimentary bacterial communities in the floodplain of Baihe River in the normal period, with community intersection in the high-water period (Figure 3). This result is contradictory to the findings of Liu et al. (2018) showing that planktonic and sediment bacterial communities did not intersect in the Yangtze River due to seasonal changes. The differences in bacterial communities may be attributable to variations in

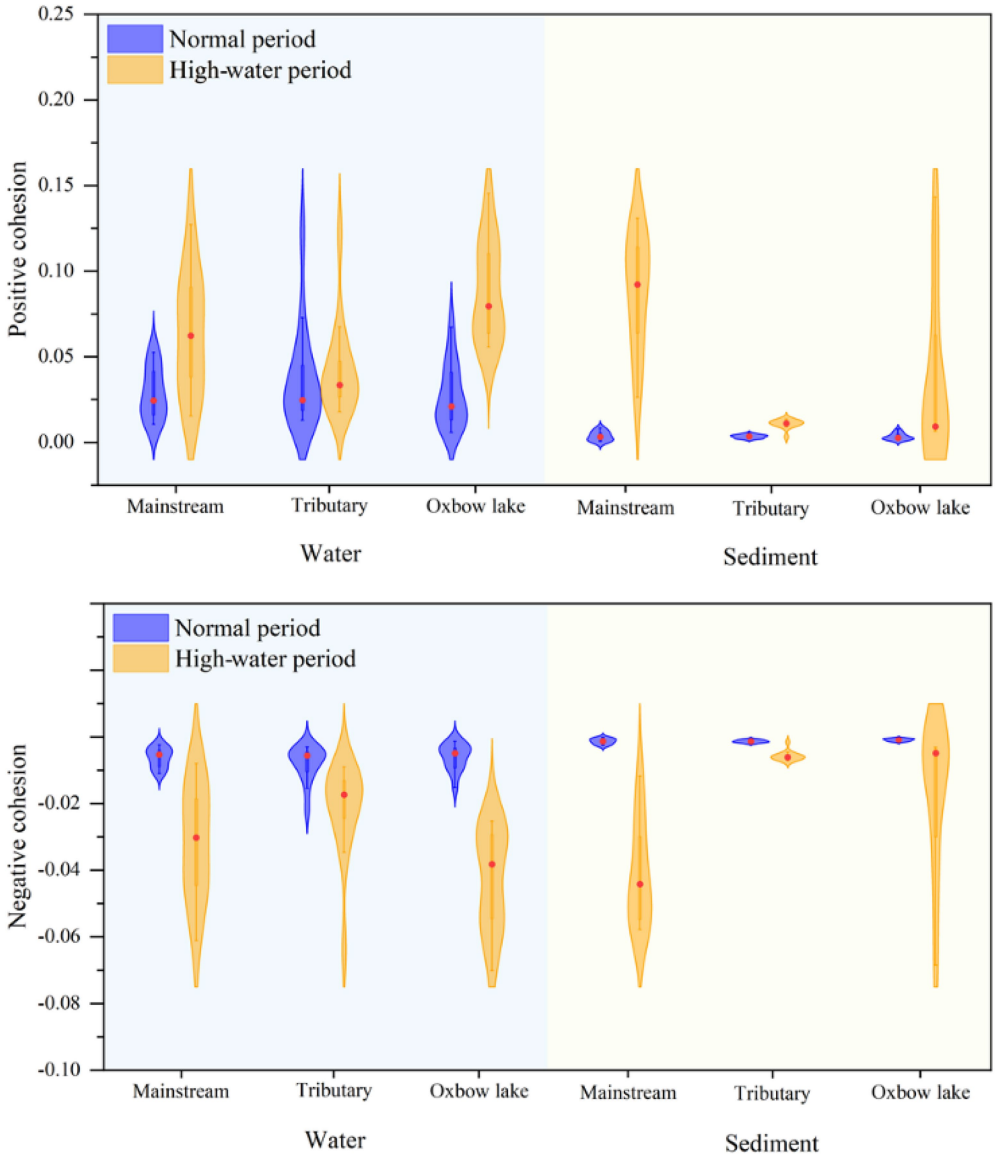


FIGURE 7
Cohesion metric of bacterial communities in water and sediment samples of the mainstream, tributaries, and oxbow lakes between two different periods.

environmental and hydrographic conditions. The Baihe River mainstream in the present study area was shallow (mean depth 30 ± 15 cm in the high-water period); however, compared with that of the normal period, the mainstream flow velocity drastically increased during the high-water period (Supplementary File 2; Supplementary Table S2), thus contributing to sediment disturbance and hence the coalescence between planktonic and sedimentary bacterial communities. In addition, the number of bacterial OTUs, relative abundances of major taxa (phylum level; Figure 2), and alpha-diversity of bacterial communities were all higher in the high-water period than the normal period (Supplementary File 1; Supplementary Figures S2, S3). These seasonal patterns could be explained by several reasons. First, the bacterial communities may have experienced seasonal succession. Second, in the high-water period, sediment disturbance and the coalescence between planktonic and sedimentary bacterial communities could lead to an increased number of bacterial OTUs and improved diversity. Third, rainfall events might wash out bacteria from the surroundings, increasing species richness and shifting the bacterial communities in the rainy season (Chen et al., 2019).

Among the three typical water bodies (i.e., mainstream, tributaries, and oxbow lakes) of Baihe River, the distribution, number of OTUs, and diversity of planktonic and sedimentary bacterial communities all changed in the study periods (Supplementary File 1; Supplementary Figures S2, S3, S5). The

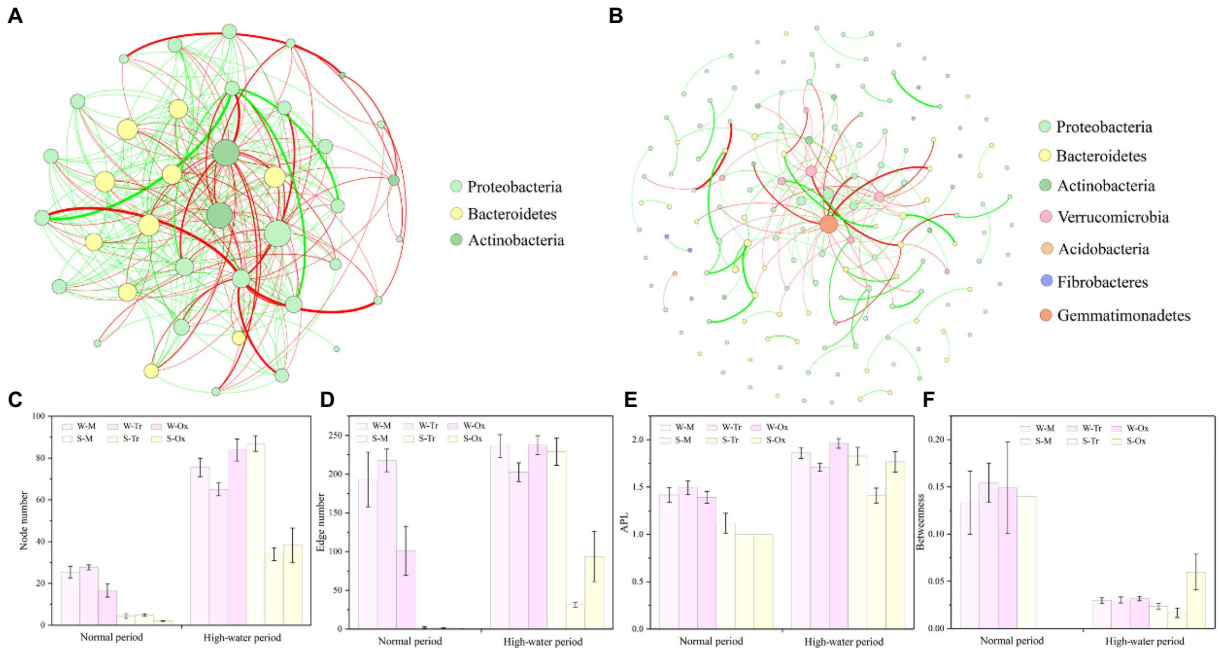


FIGURE 8
Co-occurrence networks of bacterial communities and topological features of sub-networks. Bacterial networks in the **(A)** normal and **(B)** high-water periods were constructed at the phylum level. The network topology was characterized using **(C)** node number, **(D)** edge number, **(E)** average path length (APL) and **(F)** betweenness. W-M, water-mainstream; W-Tr, water-tributaries; W-Ox, water-oxbow lakes; S-M, sediment-mainstream; S-Tr, sediment-tributaries; and S-Ox, sediment-oxbow lakes.

community differences among these water bodies may be partially caused by environmental changes across seasons (Chen et al., 2019; Luo et al., 2019; Gao et al., 2020), as distinctive differences in some environmental variables (e.g., flow velocity, pH, nutrients) were detected (Supplementary File 2; Supplementary Table S2). Another possible reason is the potential influence of hydrological connectivity, because the mainstream, tributaries and oxbow lakes would be connected with each other, and the exchange of matter, energy and species in water would be higher than that of sediment during the high-water period. Furthermore, biological differences (e.g., planktonic and sedimentary bacteria) and interactions between external and internal factors (e.g., physicochemical factors and bacterial species) can enhance bacterial community dynamics (Sommer et al., 2012). Overall, complex interactions among aquatic environments, biological conditions, and spatial factors result in the distinctive patterns of bacterial community diversity and composition in the floodplain ecosystem.

Hydrological connectivity facilitates bacterial immigration and community coalescence

Similar to previous findings for large rivers (e.g., Liu et al., 2018), the immigration ability of the planktonic bacterial community was higher than that of the sedimentary bacterial community in the floodplain of Baihe River. This phenomenon depends not only on the living habits of bacterial species themselves, but also on the influence of surrounding environments (Liu et al., 2018; Gao et al., 2020). During the high-water period, the connections between the mainstream, tributaries and oxbow lakes would be enhanced with the rising water level. Consequently, the Bray–Curtis similarity of the planktonic bacterial community increased (Figure 5), and planktonic bacteria aggregated in oxbow lakes with increasing flow (Figures 4, 6). However, immigration and coalescence of the sedimentary bacterial community showed different patterns compared with to those of the planktonic bacterial community, consistent with results reported for the Yangtze River (Gao et al., 2021). Accordingly, increased hydrological connectivity in the high-water period can promote the immigration and coalescence of the planktonic rather than the sedimentary bacterial community in the lateral direction of the floodplain.

In the vertical direction, there were spatiotemporal variations in the proportions of overlapping OTUs between planktonic and sedimentary bacterial communities (Supplementary File 1; Supplementary Figure S7). Irrespective of the water body type (i.e., mainstream, tributaries or oxbow lakes), both the planktonic and sedimentary bacterial communities exhibited minimal coalescence in the normal period, in agreement with results for the Yangtze River and other places (e.g., Liu et al., 2018). However, upon the arrival of the high-water period, there was increased coalescence of both planktonic and sedimentary bacterial communities, especially in the mainstream, compared with

that of the normal period (Supplementary File 1; Supplementary Figure S7). In addition, the results of microbial source tracing indicated the coalescence of the planktonic bacterial community with the sedimentary bacterial community in the mainstream during the high-water period (Figure 6). These results are mainly attributable to the influence of flood tides, increased mainstream velocity, and suspension of clay and silt in sediment during the high-water period (Padding and Louis, 2004; Zeng et al., 2015).

Furthermore, in the high-water period, suspended sedimentary bacteria would immigrate to the oxbow lakes with flow laterally, while the contribution of tributary bacterial communities to mainstream and oxbow lake bacterial communities showed a downward trend (Figure 6). In summary, hydrological connectivity can facilitate the coalescence of planktonic and sedimentary bacterial communities in the mainstream vertically, and increase the probability of sedimentary bacterial community immigrating from the mainstream to oxbow lakes. As a result of community immigration and coalescence, a more alike community and more homogeneous environment would be formed in the mainstream, tributaries, and oxbow lakes, leading to the convergence of environmental conditions in the floodplain ecosystem.

Hydrological connectivity influences bacterial network complexity and keystone species

Co-occurrence network analysis can be used to explore interactions between microbial species (Röttjers and Faust, 2018). Compared with that of the normal period, the network of planktonic and sedimentary bacterial communities in the high-water period was more complex mainly because of the increased complexity of sub-networks (Figure 8). A plausible mechanism is that source limitation played a reduced role in the high-water period (e.g., increased availability of nutrients in water and sediment); consequently, the diversity of bacterial species and the complexity of the bacterial network increased (Barberán et al., 2011; Hu et al., 2018; Banerjee et al., 2019). This mechanism is supported by previous observations in rivers showing that microbial network complexity is positively correlated with sediment organic matter (Fagervold et al., 2014) and negatively correlated with water pollution level (Wu et al., 2019). Some researchers have reported that during high-water periods, matter, energy and organic substances within the hydrological cycle can readily transfer between each other, increasing the utilization efficiency of resources by living creatures (Tang et al., 2020; Xie et al., 2020). Therefore, our hypothesis proves that the complexity of the bacterial network increases in the floodplain as a result of increased hydrological connectivity.

Based on the connectivity within and among modules, we identified highly connected bacteria, known as keystone species, in the sub-networks. Keystone species play a key role in the overall structure of the microbiota, and they can be used as indicators of environmental changes (Berry and Widder, 2014; Gao et al., 2021). Therefore, we also investigated the relationships between keystone species and hydrological connectivity in the floodplain ecosystem. Across different habitat environments and seasons, the top three most abundant keystone species both in planktonic and sedimentary bacterial communities were always identified as Proteobacteria, Bacteroidetes and Actinobacteria (Supplementary File 1; Supplementary Figure S9). This result suggests that the keystone bacterial species did not shift with hydrological connectivity in the study area. Previous studies also showed that external factors, including the environment (Wu et al., 2019; Tang et al., 2020) and altitude (Lee et al., 2012) had a profound influence on riverine bacterial species. With respect to different water body types, the keystone species abundances of both planktonic and sedimentary bacterial communities increased in the high-water period compared with those of the normal period (Supplementary File 1; Supplementary Figure S9). The difference is related not only to the seasonal succession of bacterial communities themselves, but also their immigration and coalescence driven by hydrological connectivity. Moreover, a stronger coalescence of keystone species in different habitats (water and sediment) could be supported by the drastic increase in the keystone species abundance of mainstream sediment (Gao et al., 2021). Oxbow lakes, located at the end of the mainstream and tributaries in the lateral flow during the high-water period, are the sites of pooling of keystone species.

We found that some other keystone species in the bacterial sub-networks had higher abundances during the high-water period than the normal period. These keystone species were classified as Verrucomicrobia, Acidobacteria, Fibrobacteres and Gemmatimonadetes, all found in terrestrial habitats including farmland, forest and woodland (Ludwig et al., 2015). Indeed, these land use types were observed around the sampling sites, suggesting the possibility of bacterial community coalescence between aquatic and terrestrial habitats (Mansour et al., 2018), particularly during flood events.

Ecological implications of bacterial community coalescence promoted by hydrological connectivity

Positive cohesion indicates the extent of cooperative behavior between microbial communities in samples, while negative cohesion reflects competitive behavior among community members (Herren and McMahon, 2017). The results of the present study showed that during the normal period, positive cohesion of the planktonic bacterial community was higher than that of the sedimentary bacterial community in different water body types (Figure 7), consistent with the

findings reported for China's Three Gorges Reservoir (Gao et al., 2021). During the high-water period, there was a higher positive cohesion for bacterial communities in the mainstream and oxbow lakes compared with tributaries (Figure 7), because the mainstream and oxbow lakes were the areas where bacterial communities coalesced. This demonstrates that in the mainstream and oxbow lakes, the coalescence of bacterial communities led to an increase in their positive cohesion, while community complexity and stability increased simultaneously.

The results of bacterial community connectivity quantified using cohesion were corroborated through calculations of niche breadth. During the high-water period, we observed the largest increase in niche breadth for the sedimentary bacterial community in the mainstream and the planktonic bacterial community in oxbow lakes, compared with those of the normal water period (Supplementary File 1; Supplementary Figure S8). Coalescence of bacterial communities could play a positive role in improving competitiveness, expanding the distribution area, and increasing biomass. Furthermore, stronger community coalescence could result in more similar bacterial communities and associated environments (Rillig et al., 2015), with minor changes in community structure and species turnover (Hengeveld, 2002). Our study demonstrates that hydrological connectivity in the floodplain ecosystem facilitates the coalescence of planktonic and sedimentary bacterial communities, and thereby drives homogenous selection, reaching a balance in competition, dispersal, coalescence and selection. Consequently, coalescence of bacterial communities could increase community complexity and stability, thereby enhancing their competition and dispersal capacity.

Conclusion

We analyzed the spatiotemporal patterns and coalescence processes of planktonic and sedimentary bacterial communities in a floodplain ecosystem of the Yellow River source region. The results highlighted the importance of hydrological connectivity in bacterial community coalescence in the mainstream, tributaries and oxbow lakes. Hydrological connectivity promoted the lateral immigration and coalescence of planktonic bacterial community, and increased its vertical coalescence with sedimentary bacterial community, with plenty of keystone species enriched in the oxbow lakes after coalescence. Furthermore, the coalescence of bacterial communities enhanced the community complexity and stability, thereby improving their competitiveness and dispersal capacity. The findings shed light on the ecological significance of bacterial community coalescence driven by hydrological connectivity in the floodplain ecosystem.

Despite being successful in demonstrating the role of hydrological connectivity in promoting bacterial community coalescence, we did not further explore its influence based on the strength of hydrological connectivity in the oxbow lakes and mainstream. In addition, the shifts in bacterial functions as a result of community coalescence were not taken into consideration. To gain a full understanding of the ecological role of hydrological connectivity in bacterial community coalescence and after coalescence, future studies should quantify the strength of hydrological connectivity in different water body types, and determine how bacterial community coalescence influences bacterial functions in the floodplain ecosystem.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BP: methodology, writing–original draft preparation, and formal analysis. QC: visualization and investigation. XL: conceptualization, validation, data curation, and funding acquisition. HS: supervision. XZ: software. ZH: writing–reviewing and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (51939009; 52121006; 51622901; and 92047303).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.971437/full#supplementary-material

References

Argiroff, W. A., Zak, D. R., Lanser, C. M., and Wiley, M. J. (2017). Microbial community functional potential and composition are shaped by hydrologic connectivity in riverine flood-plain soils. *Microb. Ecol.* 73, 630–644. doi: 10.1007/s00248-016-0883-9

Banerjee, S., Walder, F., Büchi, L., Meyer, M., Held, A. Y., Gattinger, A., et al. (2019). Agricultural intensification reduces microbial network complexity and the abundance of keystone taxa in roots. *ISME J.* 13, 1722–1736. doi: 10.1038/s41396-019-0383-2

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2011). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbil.* 5:219. doi: 10.3389/fmicb.2014.00219

Castledine, M., Sierocinski, P., Padfield, D., and Buckling, A. (2020). Community coalescence: an eco-evolutionary perspective. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 375:20190252. doi: 10.1098/rstb.2019.0252

Chen, W. D., Ren, K. X., Isabwe, A., Liu, M., and Yang, J. (2019). Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* 7:138. doi: 10.1186/s40168-019-0749-8

Deng, Y., Jiang, Y. H., Yang, Y., He, Z., Luo, F., and Zhou, J. (2012). Molecular ecological network analyses. *BMC Bioinformatics* 13:113. doi: 10.1186/1471-2105-13-113

Durkin, P. R., Hubbard, S. M., Boyd, R. L., and Leckie, D. A. (2015). Stratigraphic expression of intra-point-bar erosion and rotation. *J. Sediment. Res.* 85, 1238–1257. doi: 10.2110/jsr.2015.78

Fagervold, S. K., Bourgeois, S., Pruski, A. M., Charles, F., Kerhervé, P., Vétion, G., et al. (2014). River organic matter shapes microbial communities in the sediment of the Rhône prodelta. *ISME J.* 8, 2327–2338. doi: 10.1038/ismej.2014.86

Faust, K., and Raes, J. (2016). CoNet app: inference of biological association networks using Cytoscape. *F1000Res.* 5:1519. doi: 10.12688/f1000research.9050.2

Findlay, S. (2010). Stream microbial ecology. *J. North Am. Benthol. Soc.* 29, 170–181. doi: 10.1899/09-023.1

Freilich, M. A., Wieters, E., Broitman, B. R., Marquet, P. A., and Navarrete, S. A. (2018). Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 99, 690–699. doi: 10.1002/ecy.2142

Gao, Y., Zhang, W., and Li, Y. (2021). Microbial community coalescence: does it matter in the three gorges reservoir? *Water Res.* 205:117638. doi: 10.1016/j.watres.2021.117638

Gao, Y., Zhang, W. L., Li, Y., Wu, H. N., Yang, N., and Hui, C. Z. (2020). Dams shift microbial community assembly and imprint nitrogen transformation along the Yangtze River–ScienceDirect. *Water Res.* 189:116579. doi: 10.1016/j.watres.2020.116579

Guimerà, R., Nunes, A., and Luís, A. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895–900. doi: 10.1038/nature03288

Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C., and Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* 10, 497–506. doi: 10.1038/nrmicro2795

Hengeveld, R. (2002). The theory of island biogeography. *Acta Biotheor.* 50, 133–136. doi: 10.1023/a:1016393430551

Herren, C. M., and McMahon, K. D. (2017). Cohesion: a method for quantifying the connectivity of microbial communities. *ISME J.* 11, 2426–2438. doi: 10.1101/112391

Hoyningen-Huene, A. J. E. V., Schneider, D., Fussmann, D., Reimer, A., Arp, G., and Daniel, R. (2019). Bacterial succession along a sediment porewater gradient at Lake Neusiedl in Austria. *Scientific Data.* 6, 163–167. doi: 10.1038/s41597-019-0172-9

Hu, Y., Bai, C. R., Jian, C., Dai, J. Y., Shao, K. Q., Tang, X. M., et al. (2018). Co-occurrence network reveals the higher fragmentation of the bacterial Community in Kaidu River Than its Tributaries in northwestern China. *Microbes Environ.* 33, 127–134. doi: 10.1264/jsme2.ME17170

Huang, Z., Wang, J. L., and Wang, Y. (2010). Grain-size features of quaternary sediments in Changjiang three gorge reservoir of the Wushan area. *Trop. Geogr.* 1, 30–33. doi: 10.13284/j.cnki.rddl.001323 (in Chinese).

Jiang, H. C., Dong, H. L., Zhang, G. X., Yu, B. S., Chapman, L. R., and Fields, M. W. (2006). Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. *Appl. Environ. Microbiol.* 72, 3832–3845. doi: 10.1128/AEM.02869-05

Jiao, S., Yang, Y. F., Xu, Y. Q., Zhang, J., and Lu, Y. H. (2020). Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J.* 14, 202–216. doi: 10.1038/s41396-019-0522-9

Knights, D., Kuczynski, J., Charlson, E. S., Zaneveld, J., Mozer, M. C., Collman, R. G., et al. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nat. Methods* 8, 761–763. doi: 10.1038/nmeth.1650

Kousky, C., and Walls, M. (2014). Floodplain conservation as a flood mitigation strategy: examining costs and benefits. *Ecol. Econ.* 104, 119–128. doi: 10.1016/j.ecolecon.2014.05.001

Labbate, M., Seymour, J. R., Lauro, F., and Brown, M. V. (2016). Anthropogenic impacts on the microbial ecology and function of aquatic environments. *Front. Microbiol.* 7:1044. doi: 10.3389/fmicb.2016.01044

Langenheder, S., and Lindström, E. S. (2019). Factors influencing aquatic and terrestrial bacterial community assembly. *Environ. Microbiol. Rep.* 11, 306–315. doi: 10.1111/1758-2229.12731

Lee, C. K., Barbier, A. B., Bottos, E. M., McDonald, L. R., and Cary, S. C. (2012). The Inter-Valley soil comparative survey: the ecology of Dry Valley edaphic microbial communities. *ISME J.* 6, 1046–1057. doi: 10.1038/ismej.2011.170

Li, Y., Gao, Y., Zhang, W. L., Wang, C., Wang, P. F., Niu, L. H., et al. (2019). Homogeneous selection dominates the microbial community assembly in the sediment of the three gorges reservoir. *Sci. Total Environ.* 690, 50–60. doi: 10.1016/j.scitotenv.2019.07.014

Liu, T., Zhang, A. N., Wang, J., Liu, S. F., Jiang, X. T., Dang, C. Y., et al. (2018). Integrated biogeography of planktonic and sedimentary bacterial communities in the Yangtze River. *Microbiome* 6:16. doi: 10.1186/s40168-017-0388-x

Livingston, G., Jiang, Y., Fox, J. W., and Leibold, M. A. (2013). The dynamics of community assembly under sudden mixing in experimental microcosms. *Ecology* 94, 2898–2906. doi: 10.1890/12-1993.1

Ludwig, W., Euzéby, J., and Whitman, W. B. (2015). *Taxonomic outlines of the phyla Bacteroidetes, Spirochaetes, Tenericutes (Mollicutes), Acidobacteria, Fibrobacteres, Fusobacteria, Dictyoglomi, Gemmatimonadetes, Lentisphaerae, Verrucomicrobia, Chlamydiae, and Planctomycetes*. John Wiley & Sons, Ltd., 21–24.

Luo, X., Zhang, W. L., Xu, M., Zhang, Y., Jiang, Y. F., He, L., et al. (2019). Correlation between antibiotic resistance genes and microbial Community in Winter Rivers. *Environ. Sci. Technol.* 42, 20–26. doi: 10.19672/j.cnki.1003-6504.2019.05.004 (in Chinese).

Ma, B., Wang, H. Z., Dsouza, M., Lou, J., He, Y., Dai, Z. M., et al. (2016). Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J.* 10, 1891–1901. doi: 10.1038/ismej.2015.261

Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* 22, 456–464. doi: 10.1016/j.copbio.2011.01.008

Mansour, I., Heppell, C. M., Ryo, M., and Rillig, M. C. (2018). Application of the microbial community coalescence concept to riverine networks. *Biol. Rev.* 93, 1832–1845. doi: 10.1111/brv.12422

Mayora, G., Devercelli, M., and Giri, F. (2013). Spatial variability of chlorophyll-a and abiotic variables in a river-floodplain system during different hydrological phases. *Hydrobiologia* 717, 51–63. doi: 10.1007/s10750-013-1566-x

Mayora, G., Scarabotti, P., Schneider, B., Alvarenga, P., and Marchese, M. (2020). Multiscale environmental heterogeneity in a large river-floodplain system. *J. S. Am. Earth Sci.* 100:102546. doi: 10.1016/j.jsames.2020.102546

Mei, R., and Liu, W. T. (2019). Quantifying the contribution of microbial immigration in engineered water systems. *Microbiome* 7:144. doi: 10.1186/s40168-019-0760-0

Michaelides, K., and Chappell, A. (2010). Connectivity as a concept for characterising hydrological behaviour. *Hydrol. Process.* 23, 517–522. doi: 10.1002/hyp.7214

Mitsch, W. J., Day, J. W., Wendell, G., and Peter, M. G. (2001). Reducing nitrogen loading to the Gulf of Mexico from the Mississippi River basin: strategies to counter a persistent ecological problem. *Bioscience* 51, 373–388. doi: 10.1641/0006-3568(2001)051[0373:RNLTTG]2.0.CO;2

Mo, Y. Y., Peng, F., Gao, X. F., Xiao, P., Logares, R., Jeppesen, E., et al. (2021). Low shifts in salinity determined assembly processes and network stability of microeukaryotic plankton communities in a subtropical urban reservoir. *Microbiome* 9:128. doi: 10.1186/s40168-021-01079-w

Mo, Y. Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018). Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J.* 12, 2198–2210. doi: 10.1038/s41396-018-0153-6

Naiman, R. J., Decamps, H., and Pollock, M. (1993). The role of riparian corridors in maintaining regional biodiversity. *Ecol. Appl.* 3, 209–212. doi: 10.2307/1941822

Padding, J. T., and Louis, A. A. (2004). Hydrodynamic and brownian fluctuations in Sedimenting suspensions. *Phys. Rev. Lett.* 93:220601.1–220601.4. doi: 10.1103/PhysRevLett.93.220601

Peng, Y. K., Yue, D. M., Xiao, L., and Qian, X. (2017). Temporal variation and co-occurrence patterns of bacterial communities in eutrophic Lake Taihu, China. *Geomicrobiology* 35, 186–197. doi: 10.1080/01490451.2017.1348406

Qian, N., Zhang, R., and Zhou, Z. D. (1987). *The Fluvial Process*. Beijing: Science Press, (in Chinese).

Qiu, L. P., Zhang, Q., Zhu, H. S., Reich, P. B., Banerjee, S., Heijden, M. G. A., et al. (2021). Erosion reduces soil microbial diversity, network complexity and multifunctionality. *ISME J.* 15, 2474–2489. doi: 10.1038/s41396-021-00913-1

Rillig, M. C., Antonovics, J., Caruso, T., Lehmann, A., Powell, J. R., Veresoglou, S. D., et al. (2015). Interchange of entire communities: microbial community coalescence. *Trends Ecol. Evol.* 30, 470–476. doi: 10.1016/j.tree.2015.06.004

Rillig, M. C., and Mansour, I. (2017). Microbial ecology: community coalescence stirs things up. *Curr. Biol.* 27, R1280–R1282. doi: 10.1016/j.cub.2017.10.027

Röttjers, L., and Faust, K. (2018). From hairballs to hypotheses–biological insights from microbial networks. *FEMS Microbiol. Rev.* 42, 761–780. doi: 10.1093/femsre/fuy030

Shi, S. J., Nuccio, E. E., Shi, Z. J., and Firestone, M. K. (2016). The interconnected rhizosphere: high network complexity dominates rhizosphere assemblages. *Ecol. Lett.* 19, 926–936. doi: 10.1111/ele.12630

Sommer, U., Adrian, R., De, L., Domis, S., and Winder, M. (2012). Beyond the plankton ecology group(PEG) model: mechanisms driving plankton succession. *Annu. Rev. Ecol. Evol. S.* 43, 429–448. doi: 10.1146/annurev-ecolsys-110411-160251

Tang, X. M., Xie, G. J., Shao, K. Q., Hu, Y., Cai, J., Bai, C. R., et al. (2020). Contrast diversity patterns and processes of microbial community assembly in a river-lake continuum across a catchment scale in northwestern China. *Environ. Microbiome* 15:10. doi: 10.1186/s40793-020-00356-9

Tockner, K., Pennetzdorfer, D., Reiner, N., Schiemer, F., and Ward, J. V. (2010). Hydrological connectivity, and the exchange of organic matter and nutrients in a dynamic river-floodplain system (Danube, Austria). *Freshw. Biol.* 41, 521–535. doi: 10.1046/j.1365-2427.1999.00399.x

Wang, D. Y., Li, Z. W., Li, Z. W., Pan, B. Z., Tian, S. M., and Nie, X. D. (2020). Environmental gradient relative to oxbow lake-meandering river connectivity in Zoige Basin of the Tibetan plateau. *Ecol. Eng.* 156:105983. doi: 10.1016/j.ecoleng.2020.105983

Wu, H. N., Li, Y., Zhang, W. L., Wang, C., Wang, P. F., Niu, L. H., et al. (2019). Bacterial community composition and function shift with the aggravation of water quality in a heavily polluted river. *J. Environ. Manag.* 237, 433–441. doi: 10.1016/j.jenvman.2019.02.101

Wu, W. X., Lu, H. P., Sastri, A., Yeh, Y. C., Gong, G. C., and Chou, W. C. (2018). Contrasting the relative importance of species sorting and dispersal limitation in shaping marine bacterial versus protist communities. *ISME J.* 12, 485–494. doi: 10.1038/ismej.2017.183

Xie, C. J., Tian, X., Liu, Z. Z., Ning, Z. H., Yu, S. L., Wang, Q., et al. (2020). Effects of hydrological connection on wetland organisms. *Environ. Ecol.* 2, 26–34. 2096-6830 (2020) 01-0026-09 (in Chinese).

Yuan, M. M., Guo, X., Wu, L. W., Zhang, Y., Xiao, N. J., Ning, D. L., et al. (2021). Climate warming enhances microbial network complexity and stability. *Nat. Clim. Chang.* 11, 343–348. doi: 10.1038/s41558-021-00989-9

Zeng, X. M., He, R. Y., Xue, Z., Wang, H. J., Wang, Y., Yao, Z. G., et al. (2015). River-derived sediment suspension and transport in the Bohai, yellow, and East China seas: a preliminary modeling study. *Cont. Shelf Res.* 111, 112–125. doi: 10.1016/j.csr.2015.08.015

Zeng, J., Jiao, C. C., Zhao, D. Y., and Xu, H. M. (2019). Patterns and assembly processes of planktonic and sedimentary bacterial community differ along a trophic gradient in freshwater lakes. *Ecol. Indic.* 106:105491. doi: 10.1016/j.ecolind.2019.105491

Zhang, J. Y.. (2016). Spaa: species association analysis. R package version 0.2.2. Available at: http://CRAN.R-project.org/package=spaa

Zhang, J. W., Li, C., Rahaman, M. M., Yao, Y. D., Ma, P. L., Zhang, J. H., et al. (2022a). A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approaches. *Artif. Intell. Rev.* 55, 2875–2944. doi: 10.1007/s10462-021-10082-4

Zhang, J. H., Li, C., Yin, Y. M., Zhang, J. W., Zhao, X., and Grzegorzek, M. (2022b). Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artif. Intell. Rev.* 1–58. doi: 10.1007/s10462-022-10192-7

Zhou, J., and Ning, D. (2017). Stochastic community assembly: does it matter in microbial ecology? *Microbiol. Mol. Biol. Rev.* 81:e00002–e00017. doi: 10.1128/MMBR.00002-17

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 04 October 2022
DOI 10.3389/fmicb.2022.1034051

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Fanglin Guan,
Xi'an Jiaotong University Health Science Center, China
Feng Chen,
Nanjing Medical University,
China
Lagabaiyila Zha,
Central South University,
China

*CORRESPONDENCE
Dawei Guan
dwguan@cmu.edu.cn
Rui Zhao
rzhao@cmu.edu.cn

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 01 September 2022
ACCEPTED 15 September 2022
PUBLISHED 04 October 2022

CITATION
Wang Z, Zhang F, Wang L, Yuan H, Guan D and Zhao R (2022) Advances in artificial intelligence-based microbiome for PMI estimation.
Front. Microbiol. 13:1034051.
doi: 10.3389/fmicb.2022.1034051

Advances in artificial intelligence-based microbiome for PMI estimation

Ziwei Wang[1], Fuyuan Zhang[1], Linlin Wang[1,2], Huiya Yuan[1,2], Dawei Guan[1,2]* and Rui Zhao[1,2]*

[1]Department of Forensic Pathology, China Medical University School of Forensic Medicine, Shenyang, China, [2]Liaoning Province Key Laboratory of Forensic Bio-evidence Science, Shenyang, China

Postmortem interval (PMI) estimation has always been a major challenge in forensic science. Conventional methods for predicting PMI are based on postmortem phenomena, metabolite or biochemical changes, and insect succession. Because postmortem microbial succession follows a certain temporal regularity, the microbiome has been shown to be a potentially effective tool for PMI estimation in the last decade. Recently, artificial intelligence (AI) technologies shed new lights on forensic medicine through analyzing big data, establishing prediction models, assisting in decision-making, etc. With the application of next-generation sequencing (NGS) and AI techniques, it is possible for forensic practitioners to improve the dataset of microbial communities and obtain detailed information on the inventory of specific ecosystems, quantifications of community diversity, descriptions of their ecological function, and even their application in legal medicine. This review describes the postmortem succession of the microbiome in cadavers and their surroundings, and summarizes the application, advantages, problems, and future strategies of AI-based microbiome analysis for PMI estimation.

KEYWORDS

postmortem submersion interval, forensic medicine, microbial community, artificial intelligence, microbial succession

Introduction

Postmortem interval (PMI) is the time between the discovery and examination of the body and the occurrence of death. Relatively accurate estimation of PMI has always been an important issue in the field of forensic medicine. PMI estimation based on postmortem phenomena is still the common and feasible way in forensic practice. Owing to the inference of PMI being highly susceptible to the individual's physical condition, cause of death, and environmental conditions, the predicted accuracy of PMI cannot meet the requirements of the actual work. Microbial communities are involved in the decomposition of deceased bodies and present a certain regular succession on the host, making it possible to predict PMI based on the microbial communities (Diez Lopez et al., 2022). In the last decade, postmortem microbiome has been applied to predict PMI, and technologies for

microorganisms cover the shortfall of traditional morphological methods. Traditional methods using microbial cultivation of target-specific strains are highly dependent on culture conditions and have limitations for the analysis of the component and function of microbial communities (Cecchini et al., 2012; Zhou and Bian, 2018). Next-generation sequencing (NGS) has brought revolutionary progress to the study of microorganisms in forensic medicine. NGS can quickly and accurately analyze the entire microbial community, including many species that cannot be cultured in the laboratory (Kuiper, 2016). Meanwhile, the use of NGS brings a huge amount of microbial data, which requires an efficient data analysis method to process. Recently, artificial intelligence (AI) technologies shed new lights on forensic medicine through analyzing big data, establishing prediction models, assisting in decision-making, etc. (Geradts, 2018). Importantly, the development of AI techniques has facilitated forensic practitioners to improve understanding of microbial communities through analysis of the postmortem changes of microorganisms in different organs/tissues at various taxonomic levels (Speruda et al., 2021).

This review summarizes the succession patterns of postmortem microbial communities both on cadavers and their surrounding environment, and analyzes the advances of AI techniques on PMI estimation and their potential application on PMI prediction in the future.

Postmortem microbial succession in cadavers

Microorganisms predominantly colonize five parts of cadavers: the gastrointestinal tract, the oral cavity, skin, the respiratory tract, and the genitourinary tract. Due to the convenience of sampling from living individuals, most studies have focused on the gastrointestinal tract, the oral cavity, and skin (Dash and Das, 2022). In recent years, numerous studies have been conducted on the succession pattern of microbial communities and PMI prediction based on different organs in both human remains and animal models. There are dramatic postmortem changes of microbial community succession in different organs (Pechal et al., 2014; DeBruyn and Hauther, 2017; Dash and Das, 2022) The diversity of most microorganisms shows similar decreasing trends with PMI, presenting a significant negative linear correlation (Pechal et al., 2014; DeBruyn and Hauther, 2017; Li et al., 2020). At the phylum level, *Proteobacteria* and *Firmicutes* dominate the microbial communities in different postmortem organs in both terrestrial and water environments, making them potential markers for PMI or postmortem submersion interval (PMSI) prediction (Benbow et al., 2015; He et al., 2019; Javan et al., 2019; Yuan et al., 2020; Dash and Das, 2022). The detailed taxonomy on families or genus levels of *Proteobacteria* and *Firmicutes* would undoubtedly enhance understanding of postmortem microbial community succession in different samples. For instance, in terrestrial conditions, Tuomisto et al. (2013) found that the pericardial fluid and liver remain sterile within 5 days postmortem, while the highest abundances of *Bifidobacteria*, *Bacteroides*, *Enterobacter*, and *Clostridia* are detected in the liver, mesenteric lymph node, pericardial fluid of cadavers within 7 days, providing a short-term PMI fetch reference. Some studies revealed changes of dominant microorganisms in different human organs and blood specimens after death. *Bacillus* and *Lactobacillus* predominated in the short-term after death followed by an exponential decrease with the extension of PMI, while parthenogenic anaerobic bacteria, such as *Clostridium*, were predominant in the late phase of PMI (Can et al., 2014; Hauther et al., 2015; Javan et al., 2016; DeBruyn and Hauther, 2017). This accounted for the phenomenon of Postmortem Clostridium Effect (PCE) at decomposition stage (Javan et al., 2017). In addition, the alterations in several species of *Clostridium* may provide more information on different stages of PMI, for example, *C. novyi* was relatively more abundant in late PMI; however, an unknown member of the genus *Clostridium* was found to be more abundant in early PMI (Javan et al., 2016). In consistent with the data in terrestrial conditions, some studies demonstrated that *Enterococcus* and *Clostridium* were predominated on the skin and bones of water-dead pigs in the late stage of PMI (Benbow et al., 2015; Cartozzo et al., 2021). Our recent studies also showed that *Clostridium* in the lung and cecum were associated with PMSI in the fresh water environment (Wang et al., 2020; Zhang et al., 2022).

Numerous studies revealed the influence of different factors on PMI estimation, such as sample type (Javan et al., 2019; Lutz et al., 2020) and environmental factors (Iancu et al., 2018; Diez Lopez et al., 2022). Furthermore, the effects of gender on the analysis of microbial communities cannot be ignored, with evidence that the genera *Rothia* and *Streptococcus* were only present in the visceral organs of men, while an abundance of the genera *Clostridium* and *Pseudomonas* were found in a higher proportion of heart tissues from women compared with those from men (Javan et al., 2016; Bell et al., 2018). The study of sex-specific microbial communities could help to improve the precise of PMI estimation. Considering so many factors that affect postmortem microbial community succession and PMI estimation, exploration of an effective detection method and sufficient microbial datasets should be undertaken in future work.

Postmortem microorganisms changes in the surrounding environment

Microbial communities of cadavers interact with the surrounding environment. Although microbial community succession in carcasses placed on different soil types tends to be consistent postmortem (Metcalf et al., 2016), the microbial community in the environment does affect the process of decomposition. For example, mice that interact with normal soil decompose faster than the cadavers placed on sterile soil (Lauber et al., 2014). In addition, decomposed cadavers release various

adipose tissues, volatile fatty acids, organic acids, organic nitrogen, and bacterial flora—such as obligately anaerobic *Bacteroides*—into the soil (Vass et al., 1992). This is followed by changes in the microbial community in the soil after death, which make it possible to predict PMI based on soil microbes (Cobaugh et al., 2015).

Terrestrial soil microbes related to forensic research can be broadly divided into surface soil and buried soil (Carter et al., 2007). Surface soil microbial communities exhibit decreased trends in abundance, diversity, and evenness during decomposition, with a sharp increase in the abundance of *Firmicutes* and *Proteobacteria* and a decrease in the abundance of *Acidobacteria* in soils around cadavers (Cobaugh et al., 2015; Adserias-Garriga et al., 2017a,b; Procopio et al., 2019). In contrast, a study found that buried soil microbial communities showed the trends of increasing abundance, decreasing evenness, and consistent diversity, and the microbial composition remained unchanged throughout the decomposition process, with *Proteobacteria* being the most abundant phylum (Finley et al., 2016). According to the microbial community succession of soils surrounding cadavers, Procopio et al. (2019) revealed that *Bacteroides* spp., specific mammal-derived taxa, could be detected in the buried soil 6 months after PMI. However, soil microbial communities are easily influenced by environmental factors (Chernov and Zhelezova, 2020), such as temperature, moisture, vegetation cover, and insect activity. Habtom et al. (2019) analyzed different soils in five different rainfall zones and found significant differences in bacterial population structure among soil types in the same geographic location. Yang et al. (2021) studied the variation of microbial community composition in 529 soil samples from 61 urban districts of 10 cities in China at a large spatial scale and showed that the similarity of urban soil bacterial communities decreased significantly with increasing geographical distance. Although the population structure of soil bacteria within the same city was relatively similar, the identification accuracy of random soil samples was 90.0% at the city level and 66.7% at the district level within the city. However, the use of distinguished soil microorganisms in forensic science needs to be confirmed in further studies. Owing to the inherent microbial communities in different soils, it is difficult to compare the microbial databases from numerous studies using different soils for PMI prediction. Hence, predicting PMI according to the soil microbial community succession alone is inadequate; a better option would be to combine the soil microbial community with that in cadavers and consider the influence of entomology and ecology.

Application of artificial intelligence for PMI prediction based on microbial data

Improvements in sequencing technology, especially NGS technology, provide sufficient genomic information for analyzing entire microbial communities (Kuiper, 2016). However, owing to the massive amount of data generated and statistical validity, an effective analysis method aligned to digging deeper is needed. AI has the advantages of effective assessment models by comprehensively examining and mining multidimensional big data, evaluating weights, and identifying patterns of data changes to establish an effective "time fingerprint" mathematical model (Zou et al., 2020). Hereafter, the presented studies on PMI prediction using NGS technology are predominantly based on AI.

Postmortem microbiome analysis for PMI estimation has been improved to a relatively accurate stage using AI. At present, Machine Learning (ML) is the main AI technology used in forensic studies, ML is one type of artificial intelligence that develops algorithms to enable computers to learn from existing data without explicit programming (Zaharchuk et al., 2018). Such ML methods include k-nearest neighbor (KNN), Partial Least Squares (PLS), random forest (RF), support vector machine (SVM), and artificial neural network (ANN; Table 1). For instance, Johnson et al. (2016) and colleagues constructed a KNN model ($k=4$) for PMI estimation using microbial communities from skin in the nasal cavity and ear canal, which developed an error of only 55 accumulated degree hours (ADD) over a time period of 800 ADD. Cao et al. (2021) used segmented cecum microbial community data from rats to construct PLS models and found that the PLS model was effective in the first 9 days after death. RF is the most common ML algorithm in microbial community studies for PMI prediction and has the advantages of strong learning ability, robustness, and feasibility of the hypothesis space (Ao et al., 2019). In the terrestrial environment, Metcalf et al. (2013) established a RF regression model for the first time based on the microbial community in mouse cadaver skin and abdominal cavity samples, and this model predicted PMI with a mean absolute error (MAE) of 3.30 ± 2.52 days within the first 34 days and further provided the concept of "microbial clock." Subsequently, RF regression models were constructed using microbial communities from dead pig skin and oral swabs for PMI predictions, and the accuracy was up to 94.4% within 5 days postmortem (Pechal et al., 2014). Zhang et al. (2021) compared the separate RF regression models using microbial communities from different organs and buried soils and found that the lowest MAE value was for buried soils within 60 days after death. Zhao et al. (2022) and colleagues used rat oral microorganisms to construct a RF model, and the R^2 of the model within 59 days was 93.94%. In the aquatic environment, our recent studies provide evidence that RF regression models were effective for predicting PMSI based on the microbiota succession of the mouse cecum, with a MAE of 0.818 days within the 14 days postmortem (Zhang et al., 2022). For long-term aquatic environmental decomposition (>1 year), different researchers constructed RF regression models to predict PMSI using microbial communities of porcine ribs and scapula. The model using rib microbiota performed best within 353 days, with a root mean square error (RMSE) of ±27 days, while the model using scapula microbiota performed best within 579 days with a RMSE of ±63 days (Cartozzo et al., 2021; Randall et al., 2021). Kaszubinski et al. (2022) constructed a RF regression

TABLE 1 Application of AI on microbiome for predicting PMI.

Animal model	Experimental environment	PMI/PMSI	AI model	Model performance	Sampling location	References
Human	Temperate forest	800ADD	KNN	MAE ±55ADD	Nasal cavity, Ear canal	Johnson et al. (2016)
Rat	Artificial climate chamber	30d	PLS	RMSE within 9d: 1.96d RMSE 12d later: 5.37d RMSE within 30d: 6.57d	Cecum	Cao et al. (2021)
Mice	Laboratory	48d	RF	MAE 3.30+/−2.52d	Skin	Metcalf et al. (2013)
Pig	Temperate forest	5d	RF	94.4% accuracy rate	Skin, Oral cavity	Pechal et al. (2014)
Rat	Gravesoil	60d	RF	MAE 1.82d MAE 2.06d MAE 2.13d	Gravesoil Rectum Skin	Zhang et al. (2021)
Rat	Sterile anti-scavenging cages	59d	RF	R2 93.94%	Oral cavity	Zhao et al. (2022)
Porcine bones	Natural fresh river	353d	RF	RMSE±27d RMSE±29d	Rib Scapulae	Cartozzo et al. (2021)
Porcine bones	Freshwater lake	579d	RF	RMSE±104d RMSE±63d	Rib Scapulae	Randall et al. (2021)
Sus scrofa	Freshwater pond	547d	RF	>80% variation explained	Bone	Kaszubinski et al. (2022)
Mice	Artificial climate chamber	15d	RF	MAE 20.01 h	Cecum	Liu et al. (2021)
			ANN	MAE Within 24 h: 1.5 ± 0.8 h， Within 15d: 14.5 ± 4.4 h		Liu et al. (2020)

model using microbial communities from pig bone within 547 days, and the model exhibited high accuracy, explaining more than 80% of the variation in PMSI. Recently, Liu et al. (2020, 2021) compared the performance of RF, SVM, and ANN models using microbial communities in cecum and concluded that the ANN model performed best, with a MAE of 1.5 ± 0.8 h within 24 h and 14.5 ± 4.4 h within 15 days after death for PMI prediction. These findings suggested the combination of multiple AI methods might improve the accuracy of PMI estimation.

Although many exciting results have been achieved to date to prove that microbial communities combined with AI are potentially effective tools for predicting PMI, there are still many problems with using AI analysis of microbiological data to study PMI (Figure 1). First, there is lack of unified standardization in experimental models, collected samples, and data analysis, which means the predicted results of PMI are not credible for the courtroom (Diez Lopez et al., 2022). Many complex environmental and artificial factors can potentially affect the succession of microorganisms. Second, NGS has the limitation of short reads and low accuracy of species identification (Yakun et al., 2019); consequently, most studies have predominantly targeted amplification of the V3 and V4 regions of the 16S rRNA gene, and these fragments only provide an approximate picture of the bacterial phyla (Verma et al., 2018). Accurate detailed taxonomy annotation of microorganisms requires full-length amplification of DNA. In addition, more advanced methods to disclose all microbial community species are needed. A recent study started to use third-generation sequencing technology for microbial research (Wang et al., 2021). Third, the main microbial datasets (Silva, Greengenes) for forensic PMI studies were mainly established based on clinical or environmental studies (Quast et al., 2013; Balvočiūtė and Huson, 2017). These datasets contain different numbers and types of microbial species, which could result in differences in annotation even when using the same sequencing data. Finally, the black box and uncertainty are central challenges in designing AI tools (Saffiotti, 1987). Although AI techniques are widely used for PMI estimation, the different predicted models for PMI present difference in estimated effectiveness, especially using detailed taxonomic levels, such as species and genera. Consequently, it is necessary to explore a well-recognized AI method for its application in forensic medicine (Metcalf, 2019).

Future outlook

The widespread use of AI provides new insights into forensic PMI estimation. However, current advances in AI techniques using the microbiome for PMI prediction highlight three key points to improve the accuracy of PMI studies in the future.

1. The establishment and development of microbiome biobanks for forensic purposes are necessary. Considering the complex influences of models, samples, locations, environmental factors, and postmortem intervals, forensic researchers should collaborate to pool raw data and construct a microbiome biobank for forensic purposes.

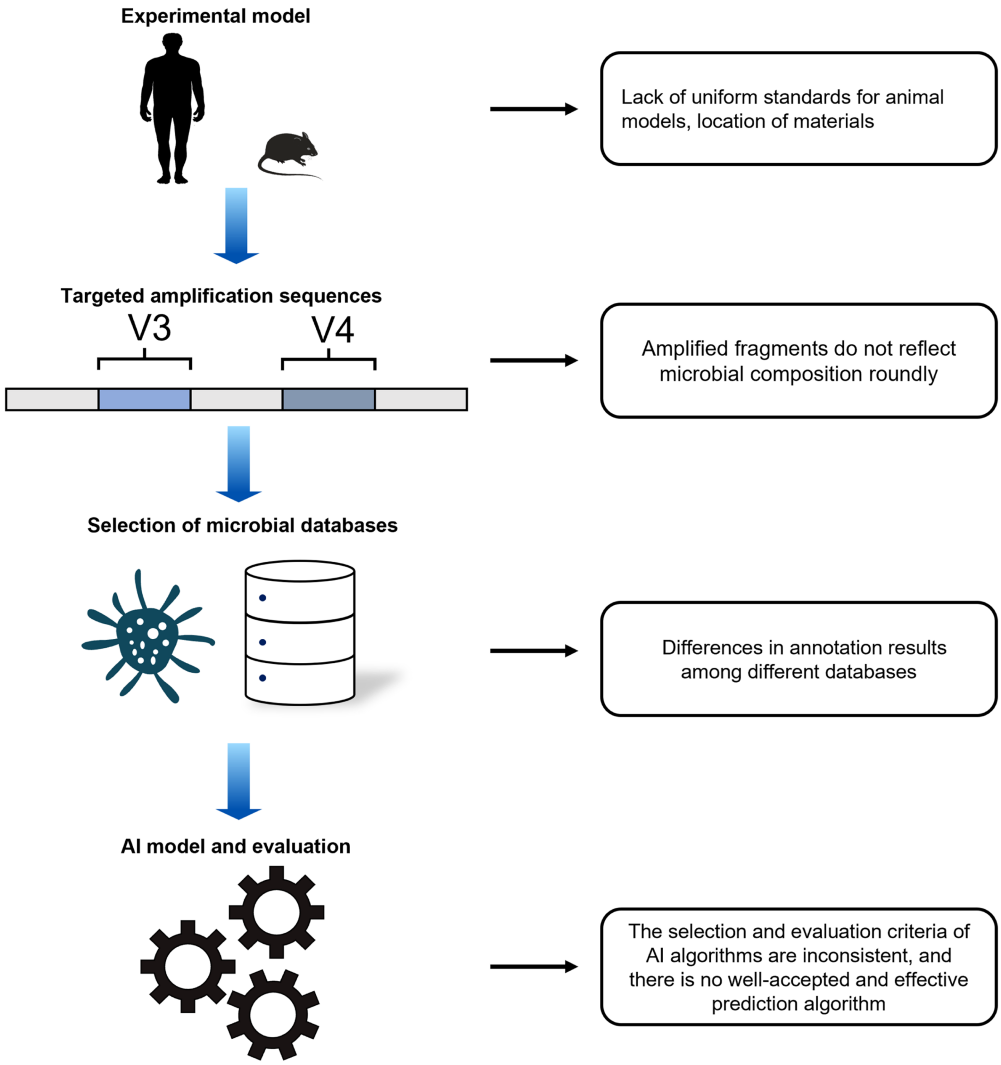


FIGURE 1
Problems for AI prediction of PMI.

2. Deep learning (DL) may shed new light on accurate predicted models for PMI. DL are considered one of the cutting-edge areas of development and study in almost all scientific and technological fields and has allowed for resolving many challenges faced by standard ML algorithms. The basis of DL is often implicated in neural network systems, where they are used to create systems that have the capability to complete complex data recognition, interpretation, and generation (Rahaman et al., 2020). AI techniques for analyzing microbiota data are still in their infancy because the amount of data used in most studies is still too low to meet the demands of DL. Deep learning—which allows neural networks to learn how to capture features by themselves (Cheng et al., 2018)—will enhance the accuracy of AI models for PMI prediction.
3. AI technologies for multi-omics provide a future direction for PMI estimation. Although microbiome analysis with AI has been shown to be effective for predicting PMI, integrated omics—including microbiomes, metabonomics, transcriptomics, and proteomics—will further improve the accuracy of PMI inference with the development of AI techniques.

Author contributions

RZ and DG designed the manuscript and edited the manuscript. RZ and ZW wrote the manuscript. FZ, LW, and HY searched, edited, and reviewed the literature. All authors have read and commented on the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (grant numbers: 81971793 and 81772023),

Shenyang Science and Technology Innovation Support Plan for Young and Middle-aged Talent (grant number: RC200412), the Natural Science Foundation of Liaoning Province (2022-YGJC-74), and National Key Research and Development Program of China (grant number: 2022YFC3302002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adserias-Garriga, J., Hernandez, M., Quijada, N. M., Rodriguez Lazaro, D., Steadman, D., and Garcia-Gil, J. (2017a). Daily thanatomicrobiome changes in soil as an approach of postmortem interval estimation: an ecological perspective. *Forensic Sci. Int.* 278, 388–395. doi: 10.1016/j.forsciint.2017.07.017

Adserias-Garriga, J., Quijada, N. M., Hernandez, M., Rodriguez Lazaro, D., Steadman, D., and Garcia-Gil, L. J. (2017b). Dynamics of the oral microbiota as a tool to estimate time since death. *Mol. Oral. Microbiol.* 32, 511–516. doi: 10.1111/omi.12191

Ao, Y., Li, H., Zhu, L., Ali, S., and Yang, Z. (2019). The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling. *J. Pet. Sci. Eng.* 174, 776–789. doi: 10.1016/j.petrol.2018.11.067

Balvočiūtė, M., and Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics* 18:114. doi: 10.1186/s12864-017-3501-4

Bell, C. R., Wilkinson, J. E., Robertson, B. K., and Javan, G. T. (2018). Sex-related differences in the thanatomicrobiome in postmortem heart samples using bacterial gene regions V1-2 and V4. *Lett. Appl. Microbiol.* 67, 144–153. doi: 10.1111/lam.13005

Benbow, M. E., Pechal, J. L., Lang, J. M., Erb, R., and Wallace, J. R. (2015). The potential of high-throughput metagenomic sequencing of aquatic bacterial communities to estimate the postmortem submersion interval. *J. Forensic Sci.* 60, 1500–1510. doi: 10.1111/1556-4029.12859

Can, I., Javan, G. T., Pozhitkov, A. E., and Noble, P. A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *J. Microbiol. Methods* 106, 1–7. doi: 10.1016/j.mimet.2014.07.026

Cao, J., Li, W. J., Wang, Y. F., An, G. S., Lu, X. J., Du, Q. X., et al. (2021). Estimating postmortem interval using intestinal microbiota diversity based on 16S rRNA high-throughput sequencing technology. *Fa Yi Xue Za Zhi* 37, 621–626. doi: 10.12116/j.issn.1004-5619.2020.400708

Carter, D. O., Yellowlees, D., and Tibbett, M. (2007). Cadaver decomposition in terrestrial ecosystems. *Naturwissenschaften* 94, 12–24. doi: 10.1007/s00114-006-0159-1

Cartozzo, C., Simmons, T., Swall, J., and Singh, B. (2021). Postmortem submersion interval (PMSI) estimation from the microbiome of Sus scrofa bone in a freshwater river. *Forensic Sci. Int.* 318:110480. doi: 10.1016/j.forsciint.2020.110480

Cecchini, F., Iacumin, L., Fontanot, M., Comi, G., and Manzano, M. (2012). Identification of the unculturable bacteria *Candidatus arthromitus* in the intestinal content of trouts using dot blot and southern blot techniques. *Vet. Microbiol.* 156, 389–394. doi: 10.1016/j.vetmic.2011.11.020

Cheng, C., Dang, W., Bai, S., Pan, L., and Liu, C. (2018). Brain functional hyper network construction and classification research based-on group lasso method. *TUT* 49, 861–866. doi: 10.16355/j.cnki.issn1007-9432tyut.2018.06.009

Chernov, T. I., and Zhelezova, A. D. (2020). The dynamics of soil microbial communities on different timescales: a review. *Eurasian Soil Sci.* 53, 643–652. doi: 10.1134/s106422932005004x

Cobaugh, K. L., Schaeffer, S. M., and DeBruyn, J. M. (2015). Functional and structural succession of soil microbial communities below decomposing human cadavers. *PLoS One* 10:e0130201. doi: 10.1371/journal.pone.0130201

Dash, H. R., and Das, S. (2022). Microbial community signatures for estimation of postmortem time intervals. *Adv. Appl. Microbiol.* 118, 91–113. doi: 10.1016/bs.aambs.2022.02.002

DeBruyn, J. M., and Hauther, K. A. (2017). Postmortem succession of gut microbial communities in deceased human subjects. *PeerJ* 5:e3437. doi: 10.7717/peerj.3437

Diez Lopez, C., Vidaki, A., and Kayser, M. (2022). Integrating the human microbiome in the forensic toolkit: current bottlenecks and future solutions. *Forensic Sci. Int. Genet.* 56:102627. doi: 10.1016/j.fsigen.2021.102627

Finley, S. J., Pechal, J. L., Benbow, M. E., Robertson, B. K., and Javan, G. T. (2016). Microbial signatures of cadaver Gravesoil during decomposition. *Microb. Ecol.* 71, 524–529. doi: 10.1007/s00248-015-0725-1

Geradts, Z. (2018). Digital, big data and computational forensics. *Forensic. Sci. Res.* 3, 179–182. doi: 10.1080/20961790.2018.1500078

Habtom, H., Pasternak, Z., Matan, O., Azulay, C., Gafny, R., and Jurkevitch, E. (2019). Applying microbial biogeography in soil forensics. *Forensic Sci. Int. Genet.* 38, 195–203. doi: 10.1016/j.fsigen.2018.11.010

Hauther, K. A., Cobaugh, K. L., Jantz, L. M., Sparer, T. E., and DeBruyn, J. M. (2015). Estimating time since death from postmortem human gut microbial communities. *J. Forensic Sci.* 60, 1234–1240. doi: 10.1111/1556-4029.12828

He, J., Guo, J., Fu, X., and Cai, J. (2019). Potential use of high-throughput sequencing of bacterial communities for postmortem submersion interval estimation. *Braz. J. Microbiol.* 50, 999–1010. doi: 10.1007/s42770-019-00119-w

Iancu, L., Junkins, E. N., Necula-Petrareanu, G., and Purcarea, C. (2018). Characterizing forensically important insect and microbial community colonization patterns in buried remains. *Sci. Rep.* 8:15513. doi: 10.1038/s41598-018-33794-0

Javan, G. T., Finley, S. J., Can, I., Wilkinson, J. E., Hanson, J. D., and Tarone, A. M. (2016). Human Thanatomicrobiome succession and time since death. *Sci. Rep.* 6:29598. doi: 10.1038/srep29598

Javan, G. T., Finley, S. J., Smith, T., Miller, J., and Wilkinson, J. E. (2017). Cadaver Thanatomicrobiome signatures: the ubiquitous nature of clostridium species in human decomposition. *Front. Microbiol.* 8:2096. doi: 10.3389/fmicb.2017.02096

Javan, G. T., Finley, S. J., Tuomisto, S., Hall, A., Benbow, M. E., and Mills, D. (2019). An interdisciplinary review of the thanatomicrobiome in human decomposition. *Forensic Sci. Med. Pathol.* 15, 75–83. doi: 10.1007/s12024-018-0061-0

Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., et al. (2016). A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLoS One* 11:e0167370. doi: 10.1371/journal.pone.0167370

Kaszubinski, S. F., Receveur, J. P., Nestle, E. D., Pechal, J. L., and Benbow, M. E. (2022). Microbial community succession of submerged bones in an aquatic habitat. *J. Forensic Sci.* 67, 1565–1578. doi: 10.1111/1556-4029.15036

Kuiper, I. (2016). Microbial forensics: next-generation sequencing as catalyst: the use of new sequencing technologies to analyze whole microbial communities could become a powerful tool for forensic and criminal investigations. *EMBO Rep.* 17, 1085–1087. doi: 10.15252/embr.201642794

Lauber, C. L., Metcalf, J. L., Keepers, K., Ackermann, G., Carter, D. O., and Knight, R. (2014). Vertebrate decomposition is accelerated by soil microbes. *Appl. Environ. Microbiol.* 80, 4920–4929. doi: 10.1128/AEM.00957-14

Li, H., Yang, E., Zhang, S., Zhang, J., Yuan, L., Liu, R., et al. (2020). Molecular characterization of gut microbial shift in SD rats after death for 30 days. *Arch. Microbiol.* 202, 1763–1773. doi: 10.1007/s00203-020-01889-w

Liu, R., Gu, Y., Shen, M., Li, H., Zhang, K., Wang, Q., et al. (2020). Predicting postmortem interval based on microbial community sequences and machine learning algorithms. *Environ. Microbiol.* 22, 2273–2291. doi: 10.1111/1462-2920.15000

Liu, R., Wang, Q., Zhang, K., Wu, H., Wang, G., Cai, W., et al. (2021). Analysis of postmortem intestinal microbiota successional patterns with application in postmortem interval estimation. *Microb. Ecol.* doi: 10.1007/s00248-021-01923-4 [Epub ahead of print].

Lutz, H., Vangelatos, A., Gottel, N., Osculati, A., Visona, S., Finley, S. J., et al. (2020). Effects of extended postmortem interval on microbial communities in organs of the human cadaver. *Front. Microbiol.* 11:569630. doi: 10.3389/fmicb.2020.569630

Metcalf, J. L. (2019). Estimating the postmortem interval using microbes: knowledge gaps and a path to technology adoption. *Forensic Sci. Int. Genet.* 38, 211–218. doi: 10.1016/j.fsigen.2018.11.004

Metcalf, J. L., Wegener Parfrey, L., Gonzalez, A., Lauber, C. L., Knights, D., Ackermann, G., et al. (2013). A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *eLife* 2:e01104. doi: 10.7554/eLife.01104

Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646

Pechal, J. L., Crippen, T. L., Benbow, M. E., Tarone, A. M., Dowd, S., and Tomberlin, J. K. (2014). The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. *Int. J. Legal Med.* 128, 193–205. doi: 10.1007/s00414-013-0872-1

Procopio, N., Ghignone, S., Williams, A., Chamberlain, A., Mello, A., and Buckley, M. (2019). Metabarcoding to investigate changes in soil microbial communities within forensic burial contexts. *Forensic Sci. Int. Genet.* 39, 73–85. doi: 10.1016/j.fsigen.2018.12.002

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., et al. (2020). Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. *J. Xray Sci. Technol.* 28, 821–839. doi: 10.3233/xst-200715

Randall, S., Cartozzo, C., Simmons, T., Swall, J. L., and Singh, B. (2021). Prediction of minimum postmortem submersion interval (PMSImin) based on eukaryotic community succession on skeletal remains recovered from a lentic environment. *Forensic Sci. Int.* 323:110784. doi: 10.1016/j.forsciint.2021.110784

Saffiotti, A. J. T. K. E. R. (1987). An AI view of the treatment of uncertainty. *Knowl. Eng. Rev.* 2, 75–97. doi: 10.1017/S0269888900000795

Speruda, M., Piecuch, A., Borzecka, J., Kadej, M., and Ogorek, R. (2021). Microbial traces and their role in forensic science. *J. Appl. Microbiol.* 132, 2547–2557. doi: 10.1111/jam.15426

Tuomisto, S., Karhunen, P. J., Vuento, R., Aittoniemi, J., and Pessi, T. (2013). Evaluation of postmortem bacterial migration using culturing and real-time quantitative PCR. *J. Forensic Sci.* 58, 910–916. doi: 10.1111/1556-4029.12124

Vass, A. A., Bass, W. M., Wolt, J. D., Foss, J. E., and Ammons, J. T. (1992). Time since death determinations of human cadavers using soil solution. *J. Forensic Sci.* 37, 1236–1253. doi: 10.1520/JFS13311J

Verma, D., Garg, P. K., and Dubey, A. K. (2018). Insights into the human oral microbiome. *Arch. Microbiol.* 200, 525–540. doi: 10.1007/s00203-018-1505-3

Wang, Y., Wang, M., Xu, W., Wang, Y., Zhang, Y., and Wang, J. (2021). Estimating the postmortem interval of carcasses in the water using the carrion insect, brain tissue RNA, bacterial biofilm, and algae. *Front. Microbiol.* 12:774276. doi: 10.3389/fmicb.2021.774276

Wang, L. L., Zhang, F. Y., Dong, W. W., Wang, C. L., Liang, X. Y., Suo, L. L., et al. (2020). A novel approach for the forensic diagnosis of drowning by microbiological analysis with next-generation sequencing and unweighted Uni Frac-based PCoA. *Int. J. Legal Med.* 134, 2149–2159. doi: 10.1007/s00414-020-02358-1

Yakun, X., Yue, M., Xiaoxi, H., and Jun, W. (2019). Analysis of prospective microbiology research using third-generation sequencing technology. *Biodivers. Sci.* 27, 534–542. doi: 10.17520/biods.2018201

Yang, T., Shi, Y., Zhu, J., Zhao, C., Wang, J., Liu, Z., et al. (2021). The spatial variation of soil bacterial community assembly processes affects the accuracy of source tracking in ten major Chinese cities. *Sci. China Life Sci.* 64, 1546–1559. doi: 10.1007/s11427-020-1843-6

Yuan, H. Y., Zhao, R., Gao, L. N., Xu, E. Y., Wang, L. L., Guan, D. W., et al. (2020). Research Progress on estimation of postmortem submersion interval. *Fa Yi Xue Za Zhi* 36, 801–806. doi: 10.12116/j.issn.1004-5619.2020.06.010

Zaharchuk, G., Gong, E., Wintermark, M., Rubin, D., and Langlotz, C. P. (2018). Deep learning in neuroradiology. *AJNR Am. J. Neuroradiol.* 39, 1776–1784. doi: 10.3174/ajnr.A5543

Zhang, J., Wang, M., Qi, X., Shi, L., Zhang, J., Zhang, X., et al. (2021). Predicting the postmortem interval of burial cadavers based on microbial community succession. *Forensic Sci. Int. Genet.* 52:102488. doi: 10.1016/j.fsigen.2021.102488

Zhang, F., Wang, P., Zeng, K., Yuan, H., Wang, Z., Li, X., et al. (2022). Postmortem submersion interval estimation of cadavers recovered from freshwater based on gut microbial community succession. *Front. Microbiol.* doi: 10.3389/fmicb.2022.988297 [Epub ahead of print].

Zhou, W., and Bian, Y. (2018). Thanatomicrobiome composition profiling as a tool for forensic investigation. *Sci. Res.* 3, 105–110. doi: 10.1080/20961790.2018.1466430

Zhao, X., Zhong, Z., and Hua, Z. (2022). Estimation of the post-mortem interval by modelling the changes in oral bacterial diversity during decomposition. *J. Appl. Microbiol.* doi: 10.1111/jam.15771 [Epub ahead of print].

Zou, Y., Zhuang, C., Fang, Q., and Li, F. (2020). Big data and artificial intelligence new insight into the estimation of postmortem interval. *Fa Yi Xue Za Zhi* 36, 86–90. doi: 10.12116/j.issn.1004-5619.2020.01.017

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 15 November 2022
DOI 10.3389/fmicb.2022.1052808

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Di Wen,
Hebei Medical University,
China
Libing Yun,
Sichuan University,
China
Yadong Guo,
Central South University,
China
Olga Lavrukova,
Petrozavodsk State University,
Russia

*CORRESPONDENCE
Rui Zhao
rzhao@cmu.edu.cn
Dawei Guan
dwguan@cmu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 24 September 2022
ACCEPTED 31 October 2022
PUBLISHED 15 November 2022

CITATION
Wang L, Zhang F, Zeng K, Dong W, Yuan H, Wang Z, Liu J, Pan J, Zhao R and Guan D (2022) Microbial communities in the liver and brain are informative for postmortem submersion interval estimation in the late phase of decomposition: A study in mouse cadavers recovered from freshwater. *Front. Microbiol.* 13:1052808. doi: 10.3389/fmicb.2022.1052808

Microbial communities in the liver and brain are informative for postmortem submersion interval estimation in the late phase of decomposition: A study in mouse cadavers recovered from freshwater

Linlin Wang[1,2†], Fuyuan Zhang[1†], Kuo Zeng[3], Wenwen Dong[1,2], Huiya Yuan[1,2], Ziwei Wang[1], Jin Liu[1], Jiaqing Pan[1], Rui Zhao[1,2*] and Dawei Guan[1,2*]

[1]Department of Forensic Pathology, China Medical University School of Forensic Medicine, Shenyang, China, [2]Liaoning Province Key Laboratory of Forensic Bio-evidence Science, Shenyang, China, [3]Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Beijing, China

Introduction: Bodies recovered from water, especially in the late phase of decomposition, pose difficulties to the investigating authorities. Various methods have been proposed for postmortem submersion interval (PMSI) estimation and drowning identification, but some limitations remain. Many recent studies have proved the value of microbiota succession in viscera for postmortem interval estimation. Nevertheless, the visceral microbiota succession and its application for PMSI estimation and drowning identification require further investigation.

Methods: In the current study, mouse drowning and CO_2 asphyxia models were developed, and cadavers were immersed in freshwater for 0 to 14 days. Microbial communities in the liver and brain were characterized *via* 16S rDNA high-throughput sequencing.

Results: Only livers and brains collected from 5 to 14 days postmortem were qualified for sequencing. There was significant variation between microbiota from liver and brain. Differences in microbiota between the cadavers of mice that had drowned and those only subjected to postmortem submersion decreased over the PMSI. Significant successions in microbial communities were observed among the different subgroups within the late phase of the PMSI in livers and brains. Eighteen taxa in the liver which were mainly related to *Clostridium_sensu_stricto* and *Aeromonas*, and 26 taxa in the brain which were mainly belonged to *Clostridium_sensu_stricto*, *Acetobacteroides*, and *Limnochorda*, were selected as potential biomarkers for PMSI estimation based on a random forest algorithm. The PMSI estimation models established yielded accurate prediction results with mean absolute errors ± the standard error of 1.282±0.189 d for the liver and 0.989±0.237 d for the brain.

Conclusions: The present study provides novel information on visceral postmortem microbiota succession in corpses submerged in freshwater which sheds new light on PMSI estimation based on the liver and brain in forensic practice.

KEYWORDS

aquatic habitat, decomposition, internal organ, microbial community, postmortem submersion interval

Introduction

Human cadavers are often discovered in a range of natural aquatic habitats such as lakes, rivers, and oceans due to drowning, disasters, and accidents (Cartozzo et al., 2021b). Bodies retrieved from water pose difficulties to the investigating authorities, particularly corpses at an advanced stage of decay. A forensic pathologist is generally required to determine the cause of death and the postmortem submersion interval (PMSI; Humphreys et al., 2013). Many studies have been conducted to address these questions. Though accumulated degree-days based on the morphological state of decomposition has been suggested to determine the PMSI (Heaton et al., 2010), it is unsuitable for corpses that have been submerged in very cold water (Dickson et al., 2011; Palazzo et al., 2020). In addition, typical macroscopic signs including the classic plume of white froth from the nose or mouth, overinflated, crepitant lungs, pulmonary edema, and water in the stomach have frequently been used to identify drowning as a cause of death, but such indicators gradually become less reliable with the progression of decomposition (Schneppe et al., 2021). Given this, novel methods for PMSI estimation and the identification of drowning are required for use in forensic practice.

Aquatic bacteria have recently attracted widespread interest from forensic experts (Uchiyama et al., 2012; Lang et al., 2016). Bacteria are ubiquitous in natural bodies of water, and they are small (0.2–2.0 μm), which facilitates their entrance into blood circulation and their deposition in the viscera during drowning (Oliveira and Amorim, 2018). A previous study indicates that various bacteria spread around the entire corpse after death (Wójcik et al., 2021). Specific genera, including *Aeromonas* in freshwater and *Vibrio* and *Photobacterium* in seawater, are indicators of drowning when they are detected in the blood and viscera of victims *via* culture-dependent and/or PCR-based methods (Kakizaki et al., 2008; Aoyagi et al., 2009). Ubiquitous microbes including the internal microbiota of the carcass as well as those of the surrounding environment play an important role in the natural decomposition of carcasses in aquatic systems (Metcalf et al., 2016). However, these methods (i.e., culture-dependent and/or PCR-based methods) could provide only limited information. With the advancement of sequencing technologies, especially next-generation sequencing, it is possible to obtain a more comprehensive understanding of microbial community succession during the decay process. Many studies in human and animal corpses indicate the potential value of microbial succession for PMI or PMSI estimation (Li et al., 2021; Randall et al., 2021). To date, microbial studies investigating aquatic ecosystems in this context have mainly focused on microbes that have colonized the surface of remains or specific body parts (e.g., bones; Wallace et al., 2021; Cartozzo et al., 2021a), which are vulnerable to environmental changes (Kaszubinski et al., 2022). However, there is a lack of studies assessing the succession pattern of microbial communities colonized in the internal organs, which are relatively resistant to environmental abiotic factors (i.e., pH and temperature) and biotic factors (i.e., insects and scavenger activities; Tomberlin et al., 2011). The liver and brain are believed to be sterile in living hosts (Javan et al., 2016). The microbes discovered in these organs of cadavers could represent those directly associated with decomposition, making them ideal subjects for postmortem microbiota investigation.

Using 16S rDNA sequencing, our previous study demonstrated that microbial communities in the viscera differed in drowning and postmortem submersion groups at 3 days postmortem (Wang et al., 2020). In another study microbiota succession in the gut was helpful for estimating the PMSI (Zhang et al., 2022b). Whether microbial succession in other visceral organs could be used for PMSI estimation and the determination of cause of death requires investigation. In the present study, to verify this hypothesis, mouse drowning and postmortem submersion models were developed, and corpses were maintained in freshwater for 0 to 14 days. Microbial communities in liver and brain were characterized by 16S rDNA high-throughput sequencing, and data were analyzed with machine-learning algorithms.

Materials and methods

Sample collection and experimental setup

All animal experiments were approved by the Animal Experiment Committee of China Medical University (approval number CMU2021202). All experiments were performed in October in a natural freshwater river (Shenyang, China; N41°57′, E123°27′). Adult male C57BL/6J mice (20–25 g, aged 8–10 weeks, $n = 180$) were purchased from the Experimental Animal Center of

China Medical University, then housed in micro-isolator cages under standard lighting (light/dark periods of 12 h) with free access to drinking water and food. Five water samples (1 l each) were taken from the experimental sites before the animal experiments and filtered through sterile 0.2-μm filters (Fisher Scientific, Hampton, NH). A total of 144 mice were randomly distributed into drowning ($n = 72$) and postmortem submersion ($n = 72$) groups. The drowning model was established as previously reported (Zhang et al., 2022a). Briefly, mice were deposited in sterile string bags and immersed in 30-cm-deep water for 1 min before being retrieved from the water for 30 s. The above steps were repeated until the animals died, then the corpses were submerged underwater. Mice in the postmortem submersion group were killed by CO_2 inhalation then submersed underwater. Nine timepoints were investigated; immediately after death, 6 h and 12 h after death, and 1, 3, 5, 7, 10, and 14 days after death. At each timepoint liver (the right lobe) and brain (the right hemispheres) specimens were harvested from 16 mice (8 per group). To assess the percentages of intestine-derived bacteria in liver and brain microbial communities during decomposition, 16 cecal content samples were collected from the corpses immediately after death. All samples were immediately frozen in liquid nitrogen and stored at −80°C for subsequent sequencing. The remaining 36 mice were processed in accordance with the above-described procedures (drowning group 18 mice, postmortem submersion subgroup 18 mice; 2 mice at each indicated timepoint) as an independent validation experiment. The specific grouping is presented in Supplementary Table 1.

16S rDNA extraction and amplification

Bacterial genomic DNA from all samples, including cecal content ($n = 16$), liver ($n = 180$), brain ($n = 180$), and water ($n = 5$) was extracted using the CTAB method. DNA concentration and purity were then determined *via* 1% agarose gels. The V3-V4 region of 16S rDNA was amplified by PCR (98°C for 1 min, followed by 30 cycles of 98°C for 10 s, 50°C for 30 s, and 72°C for 30 s, then final extension at 72°C for 10 min) using the primers 341F (CCTACGGGNGGCWGCAG) and 806R (GGACTACHVGGGTATCTAAT), which were synthesized by Sangon Biotech (Sangon, Shanghai, China). DNA was then sequenced on the Illumina NovaSeq platform (Illumina, United States), and 250-bp paired-end reads were generated.

Sequence analysis

The 16S rRNA gene sequences were processed using QIIME 1.9.1 (Caporaso et al., 2010), USEARCH 10.0 (Edgar, 2010), and in-house scripts. Paired-end Illumina reads were checked by FastQC (de Sena Brandine and Smith, 2019), and further processed by USEARCH (including joining of paired-end reads, relabeling of sequencing names, removal of barcodes and primers, filtering of low-quality reads, and finding non-redundancy reads). Based on high-confidence 16S representative sequences, an amplicon sequence variants (ASVs) table was generated. The taxonomy of the representative sequences was classified with the "RDP trainset 16" database (Cole et al., 2014) on the basis of the sintax algorithm in USEARCH (−sintax command). ASVs assigned to chloroplasts and mitochondria were removed. An ASV table was generated within USEARCH (−otutab command). For alpha and beta diversity, samples were first rarefied at minimal sequences by USEARCH (−otutab_norm command).

Analysis of microbial communities

Data analyses were conducted using R (v.4.1.1).[1] Alpha diversity was measured by the Chao1 and Shannon indexes with the "vegan" package[2] in R. The Chao1 index was used to estimate alpha diversity richness and the Simpson index to evaluate evenness in addition to richness. Analysis of the difference in alpha diversity between drowning and postmortem submersion groups was performed using Wilcoxon rank-sum tests, and corresponding p values were corrected for multiple tests using a false discovery rate set at 0.05. Differences in beta diversity metrics (unweighted UniFrac and Bray–Curtis) were assessed visually using principal coordinates analysis (PCoA) and statistically using permutational multivariate analysis of variance tests (PERMANOVAs), with a total of 999 permutations ("vegan" package). Unweighted UniFrac considers phylogeny and taxa, while Bray–Curtis takes taxa and relative abundances into account. Multiple PERMANOVAs were performed, and the groups tested included cause of death (drowning and postmortem submersion), sample type (liver and brain), and PMSI (5, 7, 10, and 14 days). Fast expectation–maximization microbial source tracking (FEAST) was used to calculate the contributions of water and intestine bacterial communities as described previously (Shenhav et al., 2019), with the "FEAST" package of R. FEAST can identify the origins of complex microbial communities based on a statistical model that assumes each sink is a complex combination of known and unknown sources. In this study, water and intestine samples were defined as "sources," and liver and brain samples were defined as "sinks."

Random forest models

Datasets derived from microbiomics have the characteristics of high dimensionality and large amounts of noise and redundancy (Li et al., 2020; Vidanaarachchi et al., 2020), and they are not amenable to analysis with traditional analytical methods (Zhang et al., 2019; Park et al., 2021). Random forest (RF) has become a popular tool for the analysis of microbial data, given that it is relatively robust with respect to outliers and noise, and is not

1 http://www.r-project.org/

2 https://cran.r-project.org/package=vegan

prone to over-fitting (Knights et al., 2011). RF reportedly exhibits satisfactory performance when used to analyze microbial data to address unanswered forensic questions such as cause of death and postmortem interval (Metcalf et al., 2013; Zhang et al., 2019).

The present study investigated the use of different organs for PMSI estimation and drowning determination. RF regression and classification models were established based on microbiota profiles (the abundance data of each ASV) using default parameters of the R implementation of the algorithm (R package "randomForest"; ntree = 1,000, square root of the number of variables for the classification model and one-third of the variables for the regression model). To visualize the similarity of samples from different groups, a multidimensional scaling (MDS) ordination plot was generated using the MDSplot function of the "randomForest" package. Final performance was assessed *via* the mean absolute error (MAE) for the regression model, and the area under the receiver operating characteristic (ROC) curve for classification. Bacterial ASVs were ranked in order of their feature importance (the percentage increase in the mean-squared error; %IncMSE) in the regression model. Biomarker sets were generated by selecting the minimum error using 10-fold cross-validation.

Results

Overview of liver and brain microbial communities during decomposition

A total of 360 viscera samples including 180 from the liver and 180 from the brain were collected and analyzed in the exploratory and validation experiments at nine PMSIs spanning 14 days (Supplementary Table 1). Macroscopically, no significant signs of decomposition were observed within 5-day postmortem. Liver and brain showed minimal autolysis at 7 days. Mild liquefaction was observed in liver and brain at 10 days. At 14 days, there were scattered putrefactive blisters on the surface and parenchyma of the liver. The brain presented apparent liquefaction. After PCR amplification and agarose gel electrophoresis detection, almost all samples collected before 5-day postmortem were not qualified enough for use in subsequent experiments (The target region of 16S rDNA could not be amplified efficiently after multiple PCR, implying the low abundance of bacteria; Supplementary Table 1). Accordingly, only samples with PMSIs ranging from 5 to 14 days were further analyzed. The V3-V4 hypervariable region of the 16S rDNA gene was sequenced to characterize the microbial community. Following quality filtering and rarefaction, a total of 7,089,046 high-quality sequences were generated from 181 sample libraries, which were clustered into 3,071 ASVs. Rarefaction curves indicated that as sequence depth increased, species richness rose considerably and then reached asymptotes (Supplementary Figure 1), demonstrating that the tissues were sufficiently sequenced to observe all taxa.

After taxonomy classification, composition analysis of microbial communities in the liver and brain was performed at different levels. At the phylum level, *Firmicutes* and *Proteobacteria* were dominant in all samples (Figure 1A). The relative abundance of *Proteobacteria* was higher in liver samples than in brain samples. The opposite was true for *Firmicutes*. In brain samples, the abundance of *Firmicutes* was increased and reached a plateau at 10 days, in conjunction with a decrease in *Proteobacteria*. As the taxonomy level increased, the difference between liver and brain became greater. At the family level higher abundance of *Clostridiaceae 1*, *Morganellaceae*, and *Enterobacteriaceae* was observed in liver samples compared to brain samples during the decomposition process, whereas the abundance of *Peptostreptococcaceae* was lower (Figure 1B). The relative abundance of *Aeromonadaceae* declined from 5 days and became relatively stable after 10 days both in liver samples and in brain samples. At the genus level, *Clostridium_sensu_stricto* and *Proteus* were more prevalent in liver samples, whereas *Proteocatella* and *Desnuesiella* were more common in brain samples (Figure 1C). The abundances of *Aeromonas* at 10 days and 14 days were lower than those at 5 days and 7 days in both organs.

To assess the possible source of microbes acquired from liver and brain samples during decomposition, an additional 16 cecal content samples and 5 water samples were obtained. The compositions of microbial communities in the liver and brain were compared with those of the water and intestine. The microorganism compositions in different samples were distinctly different (Figure 1; Supplementary Figure 2). None of the top 10 genera in the water, gut, and viscera were the same, indicating that great care must be taken to avoid microbial contamination from water or other organs during sampling in forensic practice. Given that water-derived bacteria may penetrate the viscera *via* the circulation during drowning, and the bacteria in the intestinal tract could disseminate to different parts of the body during decomposition, FEAST analysis was performed to assess the effects of water-derived and intestine-derived bacteria on viscera microbiota succession. For the liver samples, the contribution of water-derived bacteria decreased in the drowning and postmortem submersion groups as PMSI increased (Figure 1D). Similar results were observed in brain samples (Figure 1E). The contribution of intestine-derived bacteria to the brain microbial community was close to zero throughout, which was lower than that in liver (mean 7.8% ± 1.2%).

Microbial diversity in liver and brain samples

Alpha diversity was estimated using the Chao1 and Shannon indices (Figures 2A,B; Table 1). For the liver samples, there was no significant difference in the Chao1 and Shannon indexes between drowning and postmortem submersion groups at each timepoint. For the brain samples, there were only marked differences at 7-day postmortem. To visualize similarities and dissimilarities in postmortem bacterial compositions in different samples, PCoA was performed and represented in two-dimensional space. In an

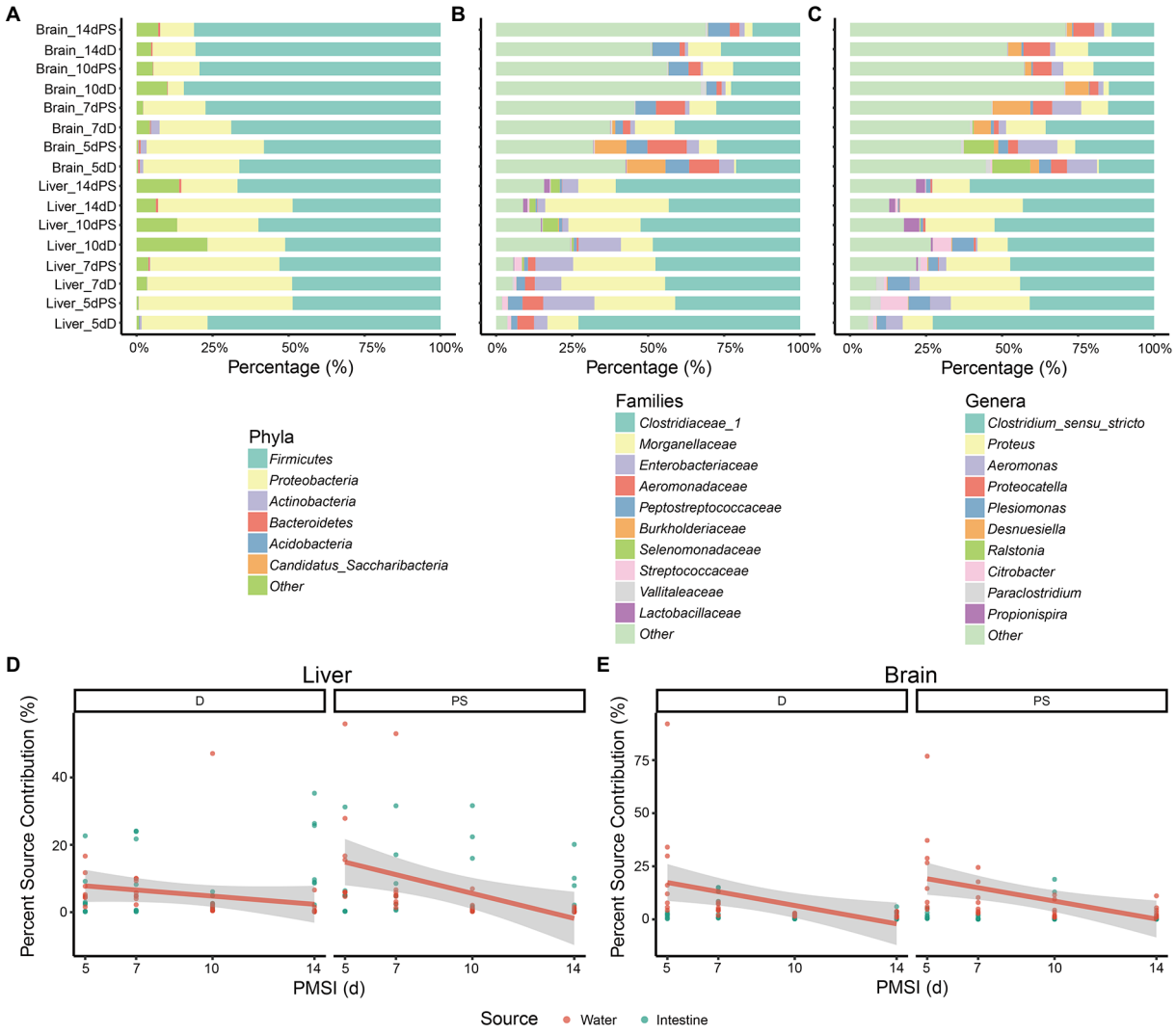


FIGURE 1
Composition of microbial communities in internal organs and microbial source tracking. **(A–C)** Relative abundance of bacterial taxa at different taxonomic levels. Stacked bar charts of the top 6 bacterial phyla **(A)**, top 10 bacterial families **(B)**, and top 10 bacterial genera **(C)** with the largest mean relative abundance in the liver and brain. Percentage source contributions of water-derived and intestine-derived bacteria to microbial communities in liver **(D)**, and brain **(E)**, over time were determined using FEAST. D, drowning group; PS, postmortem submersion group.

unweighted UniFrac distance-based PCoA plot (Figure 2C), principal coordinate 1 (PCo1) and PCo2 (42.5 and 12.1% of variance explained, respectively) axes showed that the microbial communities in the liver and brain were clearly separated during 14 days of decomposition (PERMANOVA, $R^2 = 0.247$, $p = 0.001$). PCoA2 separated the communities mainly by PMSI (5–7 days and 10–14 days). Separation between sample types and among PMSIs was more notable in a Bray–Curtis-based PCoA plot (Figure 2D). However, no difference between drowning and postmortem submersion groups was observed based on unweighted UniFrac ($p = 0.135$) or Bray–Curtis distance ($p = 0.275$) analysis (Figures 2C,D). The PERMANOVA test indicated that both sample type and PMSI could significantly affect the microbial community ($p < 0.05$). Sample type explained more variance ($R^2 = 0.247$ or 0.270) in microbial community compared to PMSI ($R^2 = 0.145$ or 0.068) and cause of death ($R^2 = 0.013$ or 0.009; Figures 2C,D). These results indicated that there were significant differences in microbial communities between the two types of viscera and among PMSIs.

Applicability of liver and brain microbial communities for drowning determination

To further assess the applicability of microbiota in different organs for drowning determination, cause-of-death classification models were established based on the relative abundance of microbiota at the level of ASV using the RF machine-learning algorithm. In MDS plots, drowning and postmortem submersion groups were indistinguishable both in liver and brain models

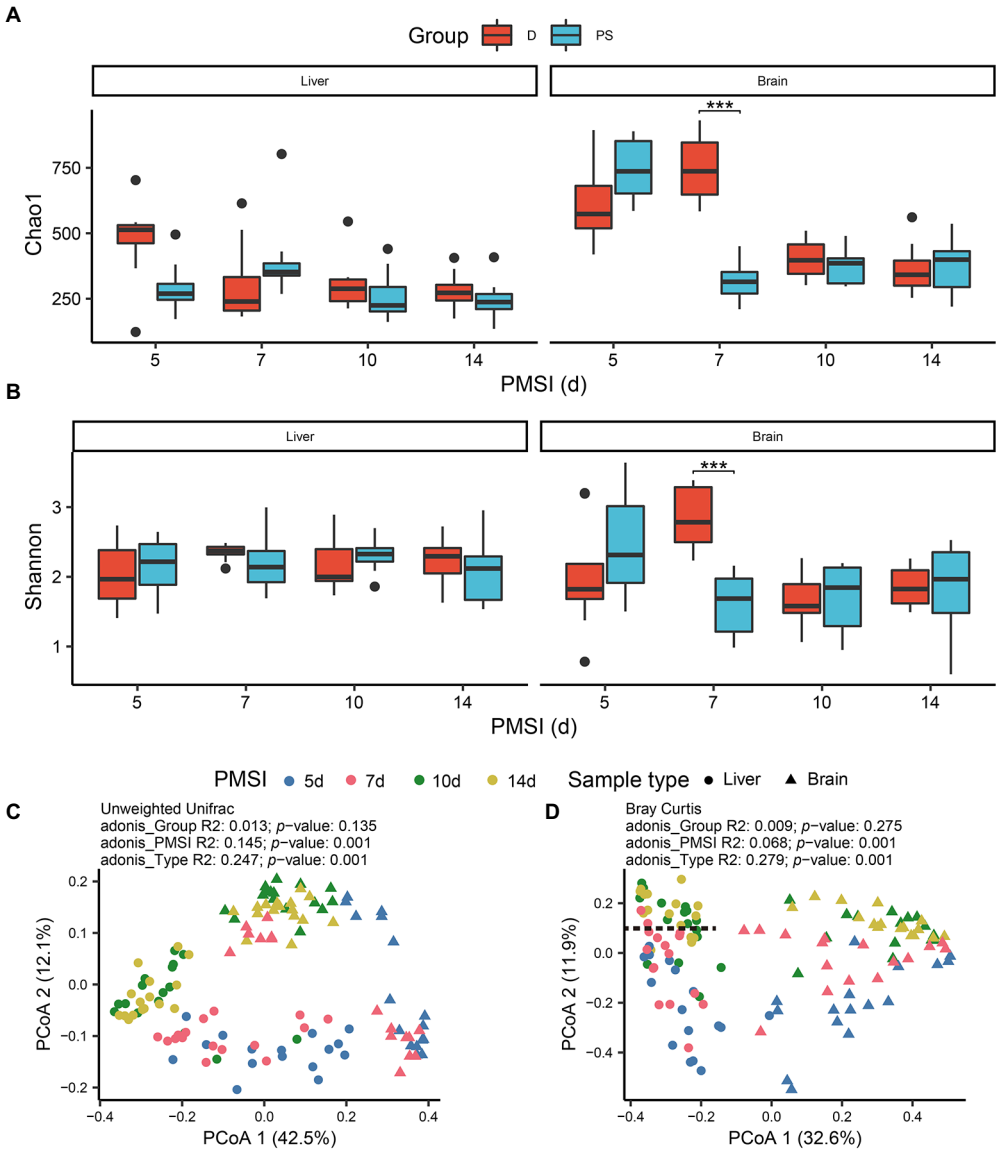


FIGURE 2
Alpha and beta diversities of microbiota in the liver and brain. Comparisons of the Chao1 **(A)**, and Shannon **(B)**, indices between drowning and postmortem groups at each PMSI. Ordination plot for the first two PCoA axes based on unweighted Unifrac **(C)**, and Bray–Curtis **(D)**, distances. Different colors indicate different PMSIs. Different sample types (liver or brain) are represented by different shapes. D, drowning group; PS, postmortem submersion group.

(Figures 3A,C). Similar results were observed in ROC curves. The areas under the curve (AUCs) were low (liver AUC exploratory 0.62, AUC validation 0.62; brain AUC exploratory 0.62, AUC validation 0.66; Figures 3B,D), indicating the performance of the classification models was poor. Thus, there was no significant difference in the microbial communities of liver and brain between the drowning and postmortem submersion groups when the individuals at different PMSIs were taken as a whole. Considering that the effect of PMSI on the bacterial community (R^2 = 0.145 or 0.068) was stronger than that of cause of death (R^2 = 0.013 or 0.009), we further analyzed the difference in microbiota between the groups at each timepoint. In PCoA analysis microbiomes from drowned corpses were clearly separated from those of postmortem submersion corpses at 5 days and 7 days, both in liver and brain (Supplementary Figures 3, 4). The same patterns were reflected in the MDS plots and ROC curves from RF classification models at each PMSI (Supplementary Figures 5, 6). Classification models were then generated based on the microbial communities in the brain and liver, which were collected at 5 days and 7 days. Overall performance is shown in Supplementary Figures 7 and 8 (liver AUC exploratory 0.94, AUC validation 0.56; brain AUC exploratory 0.79, AUC validation 0.94). These results demonstrated that the difference in microbial communities

TABLE 1 Comparisons of alpha diversity indexes (Chao1 and Shannon) between drowning and postmortem submersion by Wilcoxon rank-sum test.

PMSI	Sample type	*p*_Chao1	*p*.adjust_Chao1*	*p*_Shannon	*p*.adjust_Shannon*
5d	Liver	0.028	0.112	0.574	0.574
7d	Liver	0.083	0.166	0.195	0.574
10d	Liver	0.234	0.312	0.328	0.574
14d	Liver	0.442	0.442	0.442	0.574
5d	Brain	0.083	0.166	0.328	0.656
7d	Brain	0	0.001	0	0.001
10d	Brain	0.505	0.673	0.878	0.878
14d	Brain	0.721	0.721	0.721	0.878

**p*-values were adjusted using Benjamini-Hochberg (BH) correction and the adjusted p value cut-off was 0.05.

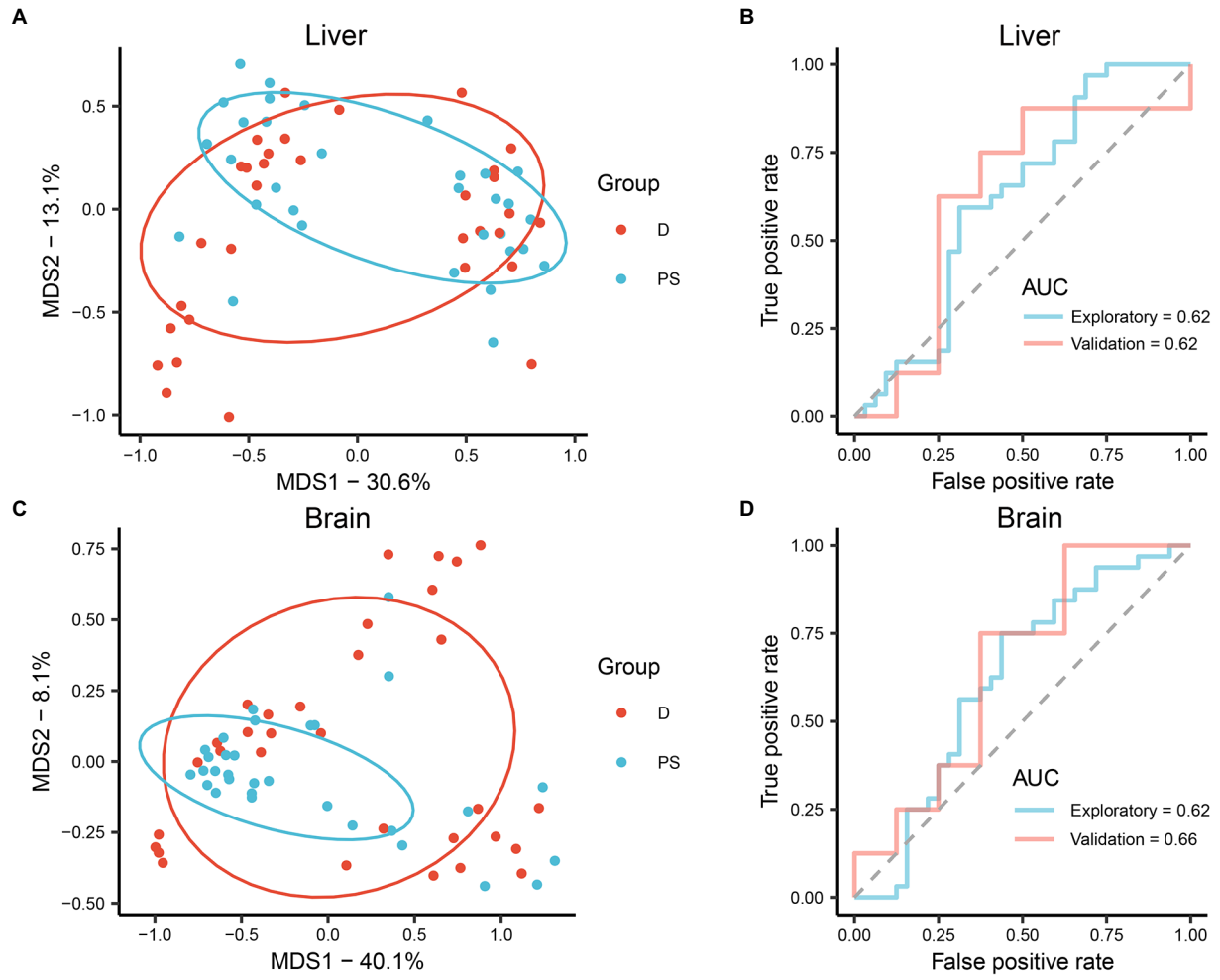


FIGURE 3
(A,B) Performance of the RF classification model built on microbiota in the liver. **(C,D)** Performance of the RF classification model built on microbiota in the brain. **(A,C)** MDS plot generated by the learning algorithm RF comparing the microbial community between drowning and postmortem submersion groups. **(B,D)** ROC curves of the RF classification model on data from exploratory and validation experiments. D, drowning group; PS, postmortem submersion group.

between drowned corpses and postmortem submersion corpses reduced gradually over the PMSI, and may only be helpful for drowning diagnosis for corpse retrieved at 5-day and 7-day postmortem. Overall, bacterial communities in liver and brain from corpses retrieved later could not be utilized for drowning determination. Thereafter, data from the two groups were assessed together to trace the common community succession for PMSI estimation.

PMSI estimation based on microbial community succession in liver and brain

The postmortem successions of microbial communities in liver and brain were assessed for PMSI estimation. PCoA based on unweighted UniFrac distance (Figures 4A,B) revealed obvious chronological ordination along PCoA1 both in liver and brain. Samples were briefly clustered into two categories; 5–7 days and 10–14 days. Subsequently, the relative abundance of microbiota at the level of ASV was analyzed with the RF algorithm to establish PMSI estimation models (the initial models). The variance explained in the liver model was 81.46%, and the variance explained in the brain was 82.45%. The regression models obtained satisfactory performances from the experimental data (liver MAE 1.137 d ± 0.115 d; brain MAE 1.114 d ± 0.111 d). The validation data were used to verify the efficiency of the models, and the MAE ± SE values were 1.229 d ± 0.146 d for the liver and 1.077 d ± 0.231 d for the brain (Figures 4C,D; Table 2). These results suggested that microbiota in liver and brain could be used for estimating PMSI in late-phase submerged corpses.

Though these models demonstrated satisfactory accuracy for PMSI prediction, there may be some ASVs contributing less to the models. Hence, cross-validation was performed to select the top informative indicator set (Figures 5A, 6A). Eighteen ASVs with high PMSI-discriminatory importance were selected as potential biomarkers in the liver, and 26 were selected in the brain (Figures 5B,C, 6B,C). Some microbes decreased in relative abundance over the PMSI, while others increased. According to species annotation, *Firmicutes* (13 ASVs in liver and 19 in brain) were dominant. Significant taxa in the liver were related to *Paraclostridium*, *Clostridium_sensu_stricto*, *Propionispira*, *Desnuesiella*, *Duncaniella*, and *Aeromonas* at the genus level. The microbes in brain mainly belonged to *Clostridium_sensu_stricto*, *Acetobacteroides*, and *Limnochorda*. Though the compositions of these indicator sets differed in liver and brain, *Clostridium_sensu_stricto* accounted for the dominant microbiota (10 ASVs in the liver and 9 in the brain). Three biomarkers (ASV 37, ASV 103, and ASV 1758) were shared between the two organs, which were assigned to *Clostridium_sensu_stricto*. A similar pattern in relative abundance for ASV 37 and ASV 1758 was observed in liver and brain. Lastly, the relative abundance of microbiota from the most

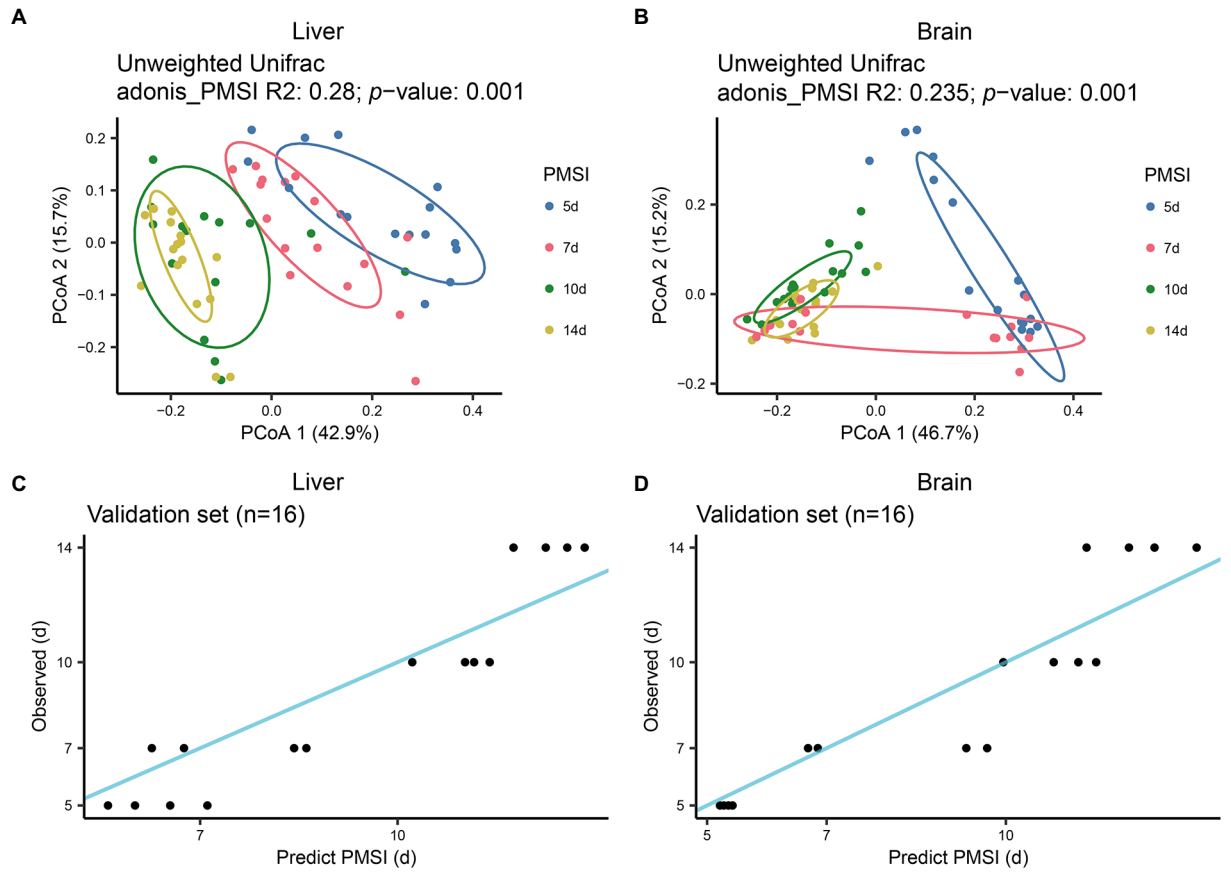


FIGURE 4
Successional dynamics of microbial communities in the liver and brain, and performances of regression models for PMSI estimation. PCoA of bacterial communities in liver **(A)**, and brain **(B)**. Different colors indicate different PMSIs. Predicted PMSI versus actual PMSI for liver **(C)**, and brain **(D)**, samples were plotted with a superimposed one-to-one reference line. Dots represent samples from the validation experiment (*n*=4 per PMSI).

TABLE 2 Prediction results of validation samples derived from the initial and refined regression models.

Sample	Group	Observed	Predic_initial_liver*	Error_initial_liver*	Predict_refined_liver*	Error_refined_liver*	Predict_initial_brain*	Error_initial_brain*	Predict_refined_brain*	Error_refined_brain*
V1	D	5	5.6	0.6	5.995	0.995	5.215	0.215	5.037	0.037
V2	D	5	6.01	1.01	6.246	1.246	5.357	0.357	5.216	0.216
V3	D	7	6.266	−0.734	6.257	−0.743	6.689	−0.311	6.737	−0.263
V4	D	7	6.755	−0.246	7.051	0.051	6.853	−0.147	6.997	−0.003
V5	D	10	10.223	0.223	10.09	0.09	10.8	0.8	10.873	0.873
V6	D	10	11.025	1.025	11.024	1.024	9.957	−0.043	9.876	−0.124
V7	D	14	12.251	−1.749	12.846	−1.154	12.488	−1.512	12.955	−1.045
V8	D	14	11.762	−2.238	11.433	−2.567	12.06	−1.94	12.545	−1.455
V9	PS	5	7.11	2.11	7.345	2.345	5.426	0.426	5.307	0.307
V10	PS	5	6.545	1.545	6.677	1.677	5.278	0.278	5.009	0.009
V11	PS	7	8.428	1.428	8.486	1.486	9.687	2.687	9.461	2.461
V12	PS	7	8.614	1.614	9.171	2.171	9.338	2.338	9.404	2.404
V13	PS	10	11.163	1.163	11.574	1.574	11.213	1.213	11.364	1.364
V14	PS	10	11.4	1.4	12.039	2.039	11.511	1.511	11.92	1.92
V15	PS	14	12.578	−1.422	13.463	−0.537	11.351	−2.649	11.327	−2.673
V16	PS	14	12.843	−1.157	13.187	−0.813	13.193	−0.807	13.336	−0.664

*Predict_initial and Error_initial: Prediction results of validation samples derived from the regression model based on all ASVs. Predict refined and Error refined: Prediction results of validation samples derived from the regression model established by selected taxa (liver 18 ASVs; brain 26 ASVs). PMSI was measured in units of days.

informative indicator sets was further regressed against PMSI (the refined models). Compared with the initial models, the explained variances in these refined models (liver 85.71%, brain 85.78%) were slightly increased. The MAEs of the refined models were similar to those of the initial models with respect to data from exploratory experiments (liver 0.906 d ± 0.114 d, brain 0.911 d ± 0.113 d) and validation experiments (liver 1.282 d ± 0.189 d, brain 0.989 d ± 0.237 d; Figures 5D, 6D; Table 2). These results indicated that these refined models based on selected microbial communities in liver and brain had powerful potential for late-phase PMSI estimation.

Discussion

Many studies have investigated carrion microbial succession using high-throughput sequencing in terrestrial habitats (Metcalf et al., 2016, 2017; Liu et al., 2021), but there are comparatively few studies on microbial communities in the internal organs in aquatic ecosystems. Thus, in the current study, we characterized shifts in bacterial communities in the livers and brains of mouse cadavers in natural freshwater, and compared differences between drowned corpses and those only subjected to postmortem submersion. The temporal succession of microbiota colonization in some liver and brain samples from these corpses was informative for PMSI estimation. Bacterial communities in the liver and brain were of little use for drowning identification.

Many internal organs including liver and brain are believed to be sterile in healthy living hosts (Javan et al., 2016). After death, various bacteria begin to spread throughout the entire corpse. The microbes discovered in these organs are mainly associated with decomposition. In the present study, most of the samples collected on days 0 to 3 postmortem were not qualified for subsequent detection, implying that there was little to no bacteria proliferating in the viscera at these timepoints. In our previous study, liver samples collected at 3 days postmortem could meet the requirements of next-generation sequencing (Wang et al., 2020). Because temperature is one of the most important environmental factors affecting the succession of microbes (Zhou et al., 2021), a reasonable explanation for the discrepancy between the previous study and the current study is that the ambient temperature in the present study (5 to 10°C) was substantially lower than that in the previous study (15 to 25°C), resulting in limited growth of microorganisms. This result suggests that in follow-up translational studies, the concept of accumulated degree-days, which integrates postmortem interval and ambient temperature (Heaton et al., 2010), should be used to minimize error caused by temperature fluctuation. Additionally, a previous study reported that liver remains sterile up to 5 days after death (Tuomisto et al., 2013). Our experiment further supported this finding. So, it is a wise choice to focus on the microbial communities in other organs when it comes to fresh corpses.

In the present study bacterial communities in the liver and brain from corpses at advanced stages of decay could not be used for drowning determination. There are likely several reasons for this. First, bacteria from the external environment and gut would influence the microbial community in viscera. Sterile organs could be colonized by infiltrating bacteria, and tissues where there is a

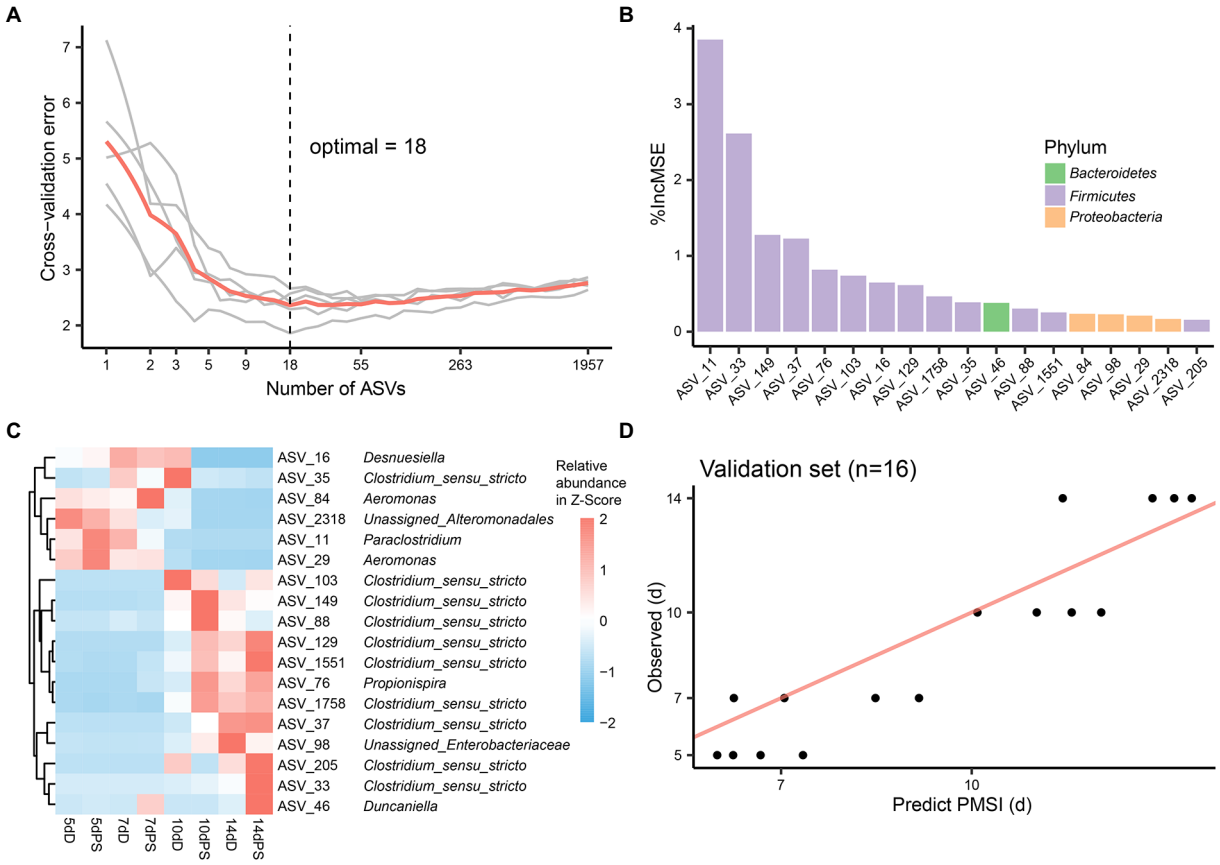


FIGURE 5
Liver biomarker identification and validation for PMSI estimation. **(A)** Cross-validation results of the initial model established using liver microbial communities. **(B)** The top 18 ASVs were identified by the RF algorithm. Biomarker taxa were ranked in decreasing order of importance (i.e., %IncMSE). **(C)** Heatmap demonstrating dynamic changes in abundance of the top 18 PMSI-predictive biomarkers. **(D)** Predicted PMSI versus actual PMSI for liver samples obtained by the refined regression model plotted with a superimposed one-to-one reference line. The dots represent liver samples from the validation experiment (*n*=4 per PMSI).

specific microbiota can be contaminated (Wójcik et al., 2021). The genus *Aeromonas* is ubiquitous in freshwater but absent in the healthy human body (Goncalves Pessoa et al., 2019), and it has been documented as a potential bacterial marker of freshwater drowning (Kakizaki et al., 2011; Huys et al., 2012; Uchiyama et al., 2012). In the current study, however, there was no significant difference in the relative abundance of *Aeromonas* in liver between drowning and postmortem submersion from 5-day postmortem, indicating the between-group differences caused by the exogenous species were relatively subtle at 5 days. Due to drastic changes in the environment, water-derived microorganisms that entered the internal organs during drowning might die gradually, resulting in decreased contribution of water-derived bacteria. Cartozzo et al. (Cartozzo et al., 2021b) reported that microorganisms inherent to the surrounding water environment contributed little to the dominating bone microbial communities with respect to relative abundance. That report is concordant with the present study. These results indicated that some changes in the microenvironment or in the microbial community during drowning became less pronounced as the PMSI extended, making a drowning diagnosis more difficult at advanced stages of decomposition. Additionally, in the current study classification models based on liver and brain microbiota collected at 5 days and 7 days performed well in exploratory experiments, implying their potential usefulness for drowning identification. The liver model exhibited poor performance in validation experiments, however (AUC 0.56), probably due to the large individual differences. This finding should be viewed with caution due to the small sample size in the present study, and should be validated in future larger studies.

Despite its unsatisfactory performance in drowning identification, microbial information derived from liver and brain exhibited good chronological regularity conducive to use for PMSI estimation. Both of the models yielded satisfactory accuracy in independent validation samples, suggesting the usefulness of liver and brain microbiota for late-PMSI estimation. In the present study, the minimum cross-validation errors were obtained when using 18 ASVs for the liver and 26 ASVs for the brain. A list of candidate taxa, identified *via* analysis of cross-validation, changed in abundance over time. Some of the microbial taxa have also been reported in postmortem microbiomes in terrestrial carcass

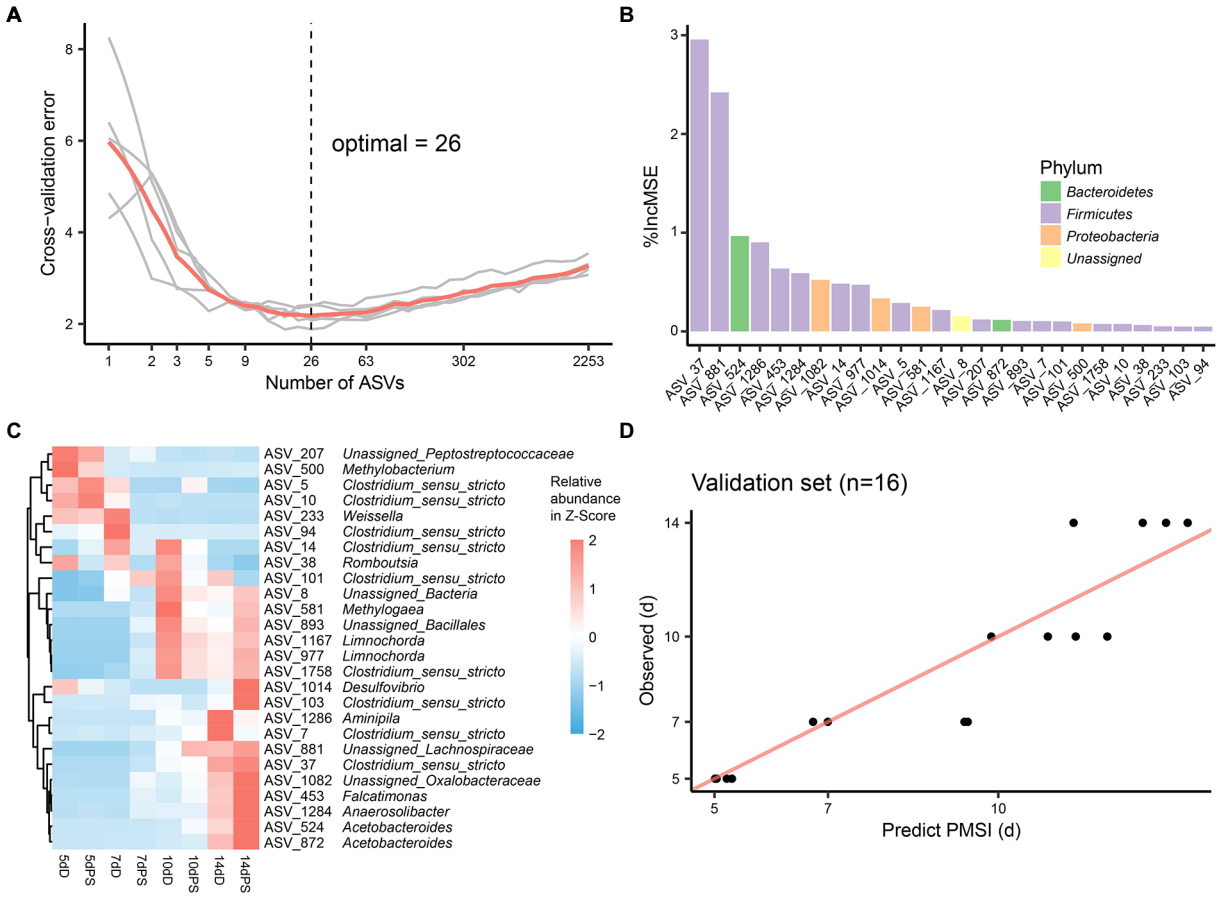


FIGURE 6
Brain biomarker identification and validation for PMSI estimation. **(A)** Cross-validation result of the initial model established using brain microbial communities. **(B)** The top 26 ASVs were identified by the RF algorithm. Biomarker taxa were ranked in decreasing order of importance (i.e., %IncMSE). **(C)** Heatmap demonstrating dynamic changes in abundance of the top 26 PMSI-predictive biomarkers. **(D)** Predicted PMSI versus actual PMSI for brain samples obtained by the refined regression model plotted with a superimposed one-to-one reference line. Dots represent brain samples from the validation experiment (*n*=4 per PMSI).

decomposition studies. For example, many selected indicators were assigned to *Clostridium_sensu_stricto*, which was also the dominant bacteria in terrestrial carcasses at 5–14-day postmortem. *Clostridium* is a widely variable oxygen-tolerant anaerobic genus found in diverse environments such as soil, freshwater, and marine sediments (Bergogne-Berezin and Towner, 1996), and it has been regarded as a key contributor to the general decomposition process on land (Hyde et al., 2015). The translocation and proliferation of *Clostridium* in postmortem human internal organs has been reported in several studies (Tuomisto et al., 2013; Javan et al., 2016). Recent research has even defined a new scientific concept, the "postmortem *Clostridium* effect"—which refers to the ubiquitous *Clostridium* spp. present during human decomposition (Javan et al., 2017). The current study demonstrates that this phenomenon can also be observed in submerged corpses. Further, some taxa unique to aquatic systems were proposed as indicators of PMSI. For instance, taxa defined as *Aeromonas* which gradually decreased in the liver could be used for PMSI estimation. *Desulfovibrio* is commonly found in aquatic environments (Amrani et al., 2014). *Methylogaea* is isolated from the soil-water interface of rice paddy fields (Tarlera, 2016). A species of *Limnochorda* has been isolated from sediment from a brackish meromictic lake (Watanabe et al., 2015). *Acetobacteroides* is found in reed swamps. The abundances of ASVs related to *Methylogaea*, *Limnochorda*, and *Acetobacteroides* were higher at the advanced decomposition stage, implying that sediment-dwelling bacteria may play an important role during degradation. This warrants further research. Lastly, these bioindicators were useful for PMSI estimation with high predictive accuracy (liver MAE 1.282 d ± 0.189 d; brain MAE 0.989 d ± 0.237 d).

Although some of the findings in the present study are novel, it should be viewed as an initial investigation into microbial succession associated with decomposition in a natural freshwater environment. The study had some limitations. Bacterial is reportedly more resistant to harsh environmental conditions (i.e., chemical and physical agents) due to the wall of peptidoglycan matrix, which renders bacteria applicable to studies in corpses at advanced stages of decomposition (Tozzo et al., 2020). Further

research should monitor the entire decomposition process (e.g., fresh to skeletonization). The environmental conditions surrounding the cadaver influence the bacterial communities present and the stages of decomposition, but the present study was conducted at a single location during a single season. For broader application of the findings in forensic science, it would be helpful to develop reliable and robust databases of microbiomes obtained in multiple aquatic environments and seasons.

The succession of postmortem microbiota that colonize internal organs (including the gut, brain, liver, spleen, and heart) has proven useful for PMI estimation in terrestrial environments (Can et al., 2014; Javan et al., 2016). The present study provides novel and informative context for better understanding the decomposition processes that submerged corpses undergo, which have important implications for forensic practice. It also sheds new light on PMSI estimation based on the succession of microbial populations in liver and brain specimens from corpses in water.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA883605.

Ethics statement

The animal study was reviewed and approved by Animal Experiment Committee of China Medical University.

Author contributions

DG and RZ conceived and designed the research. LW and FZ performed the lab experiments and wrote the main manuscript text. HY, ZW, JL, and JP performed the animal experiments. FZ, KZ, and WD performed the bioinformatic analysis. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grant from the National Key Research and Development Program of China (grant number 2022YFC3302002), National Natural Science Foundation of China (grant numbers 82271926, 81971793, 81772023, 81801874), and Shenyang Science and Technology innovation support plan for young and middle-age talent (grant number RC200412).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1052808/full#supplementary-material

References

Amrani, A., Bergon, A., Holota, H., Tamburini, C., Garel, M., Ollivier, B., et al. (2014). Transcriptomics reveal several gene expression patterns in the piezophile Desulfovibrio hydrothermalis in response to hydrostatic pressure. *PLoS One* 9:e106831. doi: 10.1371/journal.pone.0106831

Aoyagi, M., Iwadate, K., Fukui, K., Abe, S., Sakai, K., Maebashi, K., et al. (2009). A novel method for the diagnosis of drowning by detection of Aeromonas sobria with PCR method. *Leg. Med. (Tokyo)* 11, 257–259. doi: 10.1016/j.legalmed.2009.07.003

Bergogne-Berezin, E., and Towner, K. J. (1996). Acinetobacter spp. as nosocomial pathogens: microbiological, clinical, and epidemiological features. *Clin. Microbiol. Rev.* 9, 148–165. doi: 10.1128/CMR.9.2.148

Can, I., Javan, G. T., Pozhitkov, A. E., and Noble, P. A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *J. Microbiol. Methods* 106, 1–7. doi: 10.1016/j.mimet.2014.07.026

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Cartozzo, C., Simmons, T., Swall, J., and Singh, B. (2021a). Postmortem submersion interval (PMSI) estimation from the microbiome of Sus scrofa bone in a freshwater river. *Forensic Sci. Int.* 318:110480. doi: 10.1016/j.forsciint.2020.110480

Cartozzo, C., Singh, B., Swall, J., and Simmons, T. (2021b). Postmortem submersion interval (PMSI) estimation from the microbiome of sus scrofa bone in a freshwater lake. *J. Forensic Sci.* 66, 1334–1347. doi: 10.1111/1556-4029.14692

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., et al. (2014). Ribosomal database project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642. doi: 10.1093/nar/gkt1244

de Sena Brandine, G., and Smith, A. D. (2019). Falco: high-speed FastQC emulation for quality control of sequencing data. *F1000Res* 8:1874. doi: 10.12688/f1000research.21142.2

Dickson, G. C., Poulter, R. T., Maas, E. W., Probert, P. K., and Kieser, J. A. (2011). Marine bacterial succession as a potential indicator of postmortem submersion interval. *Forensic Sci. Int.* 209, 1–10. doi: 10.1016/j.forsciint.2010.10.016

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Goncalves Pessoa, R. B., De Oliveira, W. F., Marques, D. S. C., Dos Santos Correia, M. T., de Carvalho, E., and Coelho, L. (2019). The genus Aeromonas: a general approach. *Microb. Pathog.* 130, 81–94. doi: 10.1016/j.micpath.2019.02.036

Heaton, V., Lagden, A., Moffatt, C., and Simmons, T. (2010). Predicting the postmortem submersion interval for human remains recovered from U.K. waterways. *J. Forensic Sci.* 55, 302–307. doi: 10.1111/j.1556-4029.2009.01291.x

Humphreys, M. K., Panacek, E., Green, W., and Albers, E. (2013). Comparison of protocols for measuring and calculating postmortem submersion intervals for human analogs in fresh water. *J. Forensic Sci.* 58, 513–517. doi: 10.1111/1556-4029.12033

Huys, G., Coopman, V., Van Varenbergh, D., and Cordonnier, J. (2012). Selective culturing and genus-specific PCR detection for identification of Aeromonas in tissue samples to assist the medico-legal diagnosis of death by drowning. *Forensic Sci. Int.* 221, 11–15. doi: 10.1016/j.forsciint.2012.03.017

Hyde, E. R., Haarmann, D. P., Petrosino, J. F., Lynne, A. M., and Bucheli, S. R. (2015). Initial insights into bacterial succession during human decomposition. *Int. J. Legal Med.* 129, 661–671. doi: 10.1007/s00414-014-1128-4

Javan, G. T., Finley, S. J., Can, I., Wilkinson, J. E., Hanson, J. D., and Tarone, A. M. (2016). Human Thanatomicrobiome succession and time since death. *Sci. Rep.* 6:29598. doi: 10.1038/srep29598

Javan, G. T., Finley, S. J., Smith, T., Miller, J., and Wilkinson, J. E. (2017). Cadaver Thanatomicrobiome signatures: the ubiquitous nature of clostridium species in human decomposition. *Front. Microbiol.* 8:2096. doi: 10.3389/fmicb.2017.02096

Kakizaki, E., Kozawa, S., Imamura, N., Uchiyama, T., Nishida, S., Sakai, M., et al. (2011). Detection of marine and freshwater bacterioplankton in immersed victims: post-mortem bacterial invasion does not readily occur. *Forensic Sci. Int.* 211, 9–18. doi: 10.1016/j.forsciint.2011.03.036

Kakizaki, E., Takahama, K., Seo, Y., Kozawa, S., Sakai, M., and Yukawa, N. (2008). Marine bacteria comprise a possible indicator of drowning in seawater. *Forensic Sci. Int.* 176, 236–247. doi: 10.1016/j.forsciint.2007.09.018

Kaszubinski, S. F., Receveur, J. P., Nestle, E. D., Pechal, J. L., and Benbow, M. E. (2022). Microbial community succession of submerged bones in an aquatic habitat. *J. Forensic Sci.* 67, 1565–1578. doi: 10.1111/1556-4029.15036

Knights, D., Costello, E. K., and Knight, R. (2011). Supervised classification of human microbiota. *FEMS Microbiol. Rev.* 35, 343–359. doi: 10.1111/j.1574-6976.2010.00251.x

Lang, J. M., Erb, R., Pechal, J. L., Wallace, J. R., McEwan, R. W., and Benbow, M. E. (2016). Microbial biofilm community variation in flowing habitats: potential utility as bioindicators of postmortem submersion intervals. *Microorganisms* 4:0001. doi: 10.3390/microorganisms4010001

Li, J., Lu, Q., and Wen, Y. (2020). Multi-kernel linear mixed model with adaptive lasso for prediction analysis on high-dimensional multi-omics data. *Bioinformatics* 36, 1785–1794. doi: 10.1093/bioinformatics/btz822

Li, H., Zhang, S., Liu, R., Yuan, L., Wu, D., Yang, E., et al. (2021). Potential use of molecular and structural characterization of the gut bacterial community for postmortem interval estimation in Sprague Dawley rats. *Sci. Rep.* 11:225. doi: 10.1038/s41598-020-80633-2

Liu, R., Wang, Q., Zhang, K., Wu, H., Wang, G., Cai, W., et al. (2021). Analysis of postmortem intestinal microbiota successional patterns with application in postmortem interval estimation. *Microb. Ecol.* doi: 10.1007/s00248-021-01923-4

Metcalf, J. L., Wegener Parfrey, L., Gonzalez, A., Lauber, C. L., Knights, D., Ackermann, G., et al. (2013). A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *elife* 2:e01104. doi: 10.7554/eLife.01104

Metcalf, J. L., Xu, Z. Z., Bouslimani, A., Dorrestein, P., Carter, D. O., and Knight, R. (2017). Microbiome tools for forensic science. *Trends Biotechnol.* 35, 814–823. doi: 10.1016/j.tibtech.2017.03.006

Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., Van Treuren, W., Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646

Oliveira, M., and Amorim, A. (2018). Microbial forensics: new breakthroughs and future prospects. *Appl. Microbiol. Biotechnol.* 102, 10377–10391. doi: 10.1007/s00253-018-9414-6

Palazzo, C., Pelletti, G., Fais, P., Boscolo-Berto, R., Fersini, F., Gaudio, R. M., et al. (2020). Postmortem submersion interval in human bodies recovered from fresh water in an area of Mediterranean climate. Application and comparison of preexisting models. *Forensic Sci. Int.* 306:110051. doi: 10.1016/j.forsciint.2019.110051

Park, M. B., Wang, J. M., and Bulwer, B. E. (2021). Global dieting trends and seasonality: social big-data analysis may be a useful tool. *Nutrients* 13:069. doi: 10.3390/nu13041069

Randall, S., Cartozzo, C., Simmons, T., Swall, J. L., and Singh, B. (2021). Prediction of minimum postmortem submersion interval (PMSImin) based on eukaryotic community succession on skeletal remains recovered from a lentic environment. *Forensic Sci. Int.* 323:110784. doi: 10.1016/j.forsciint.2021.110784

Schneppe, S., Dokter, M., and Bockholdt, B. (2021). Macromorphological findings in cases of death in water: a critical view on "drowning signs". *Int. J. Legal Med.* 135, 281–291. doi: 10.1007/s00414-020-02469-9

Shenhav, L., Thompson, M., Joseph, T. A., Briscoe, L., Furman, O., Bogumil, D., et al. (2019). FEAST: fast expectation-maximization for microbial source tracking. *Nat. Methods* 16, 627–632. doi: 10.1038/s41592-019-0431-x

Tarlera, S. (2016). "Methylogaea" in *Bergey's manual of systematics of Archaea and bacteria*. eds. W. B. Whitman, P. DeVos and S. Dedysh (Wiley), 1–5.

Tomberlin, J. K., Mohr, R., Benbow, M. E., Tarone, A. M., and VanLaerhoven, S. (2011). A roadmap for bridging basic and applied research in forensic entomology. *Annu. Rev. Entomol.* 56, 401–421. doi: 10.1146/annurev-ento-051710-103143

Tozzo, P., D'Angiolella, G., Brun, P., Castagliuolo, I., Gino, S., and Caenazzo, L. (2020). Skin microbiome analysis for forensic human identification: what do we know so far? *Microorganisms* 8:873. doi: 10.3390/microorganisms8060873

Tuomisto, S., Karhunen, P. J., Vuento, R., Aittoniemi, J., and Pessi, T. (2013). Evaluation of postmortem bacterial migration using culturing and real-time quantitative PCR. *J. Forensic Sci.* 58, 910–916. doi: 10.1111/1556-4029.12124

Uchiyama, T., Kakizaki, E., Kozawa, S., Nishida, S., Imamura, N., and Yukawa, N. (2012). A new molecular approach to help conclude drowning as a cause of death: simultaneous detection of eight bacterioplankton species using real-time PCR assays with TaqMan probes. *Forensic Sci. Int.* 222, 11–26. doi: 10.1016/j.forsciint.2012.04.029

Vidanaarachchi, R., Shaw, M., Tang, S. L., and Halgamuge, S. (2020). IMPARO: inferring microbial interactions through parameter optimisation. *BMC Mol Cell Biol* 21:34. doi: 10.1186/s12860-020-00269-y

Wallace, J. R., Receveur, J. P., Hutchinson, P. H., Kaszubinski, S. F., Wallace, H. E., and Benbow, M. E. (2021). Microbial community succession on submerged vertebrate carcasses in a tidal river habitat: implications for aquatic forensic investigations. *J. Forensic Sci.* 66, 2307–2318. doi: 10.1111/1556-4029.14869

Wang, L. L., Zhang, F. Y., Dong, W. W., Wang, C. L., Liang, X. Y., Suo, L. L., et al. (2020). A novel approach for the forensic diagnosis of drowning by microbiological analysis with next-generation sequencing and unweighted UniFrac-based PCoA. *Int. J. Legal Med.* 134, 2149–2159. doi: 10.1007/s00414-020-02358-1

Watanabe, M., Kojima, H., and Fukui, M. (2015). Limnochorda pilosa gen. Nov., sp. nov., a moderately thermophilic, facultatively anaerobic, pleomorphic bacterium and proposal of Limnochordaceae fam. Nov., Limnochordales Ord. Nov. and Limnochordia classis nov. in the phylum Firmicutes. *Int. J. Syst. Evol. Microbiol.* 65, 2378–2384. doi: 10.1099/ijs.0.000267

Wójcik, J., Tomsia, M., Drzewiecki, A., and Skowronek, R. (2021). Thanatomicrobiome – state of the art and future directions. *Postępy Mikrobiologii - Advancements of Microbiology* 60, 21–29. doi: 10.21307/PM-2021.60.1.03

Zhang, Y., Pechal, J. L., Schmidt, C. J., Jordan, H. R., Wang, W. W., Benbow, M. E., et al. (2019). Machine learning performance in a microbial molecular autopsy context: a cross-sectional postmortem human population study. *PLoS One* 14:e0213829. doi: 10.1371/journal.pone.0213829

Zhang, F. Y., Wang, L. L., Dong, W. W., Zhang, M., Tash, D., Li, X. J., et al. (2022a). A preliminary study on early postmortem submersion interval (PMSI) estimation and cause-of-death discrimination based on nontargeted metabolomics and machine learning algorithms. *Int. J. Legal Med.* 136, 941–954. doi: 10.1007/s00414-022-02783-4

Zhang, F. Y., Wang, P. F., Zeng, K., Yuan, H. Y., Wang, Z. W., Li, X. J., et al. (2022b). Postmortem submersion interval estimation of cadavers recovered from freshwater based on gut microbial community succession. *Front. Microbiol.* doi: 10.3389/fmicb.2022.988297

Zhou, Z., Tang, H., Wang, W., Zhang, L., Su, F., Wu, Y., et al. (2021). A cold shock protein promotes high-temperature microbial growth through binding to diverse RNA species. *Cell Discov* 7:15. doi: 10.1038/s41421-021-00246-5

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 15 November 2022
DOI 10.3389/fmicb.2022.1046733

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Hu ShanQing,
Beijing Institute of Technology, China
Fenglin Zhuo,
Capital Medical University,
China

*CORRESPONDENCE
Rui-Qun Qi
xiaoqiliumin@163.com
Min Liu
liuminxiaoqi@163.com

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 17 September 2022
ACCEPTED 31 October 2022
PUBLISHED 15 November 2022

CITATION
He Q, Niu X, Qi R-Q and Liu M (2022) Advances in microbial metagenomics and artificial intelligence analysis in forensic identification.
Front. Microbiol. 13:1046733.
doi: 10.3389/fmicb.2022.1046733

Advances in microbial metagenomics and artificial intelligence analysis in forensic identification

Qing He[1,2], Xueli Niu[1,2], Rui-Qun Qi[1,2]* and Min Liu[1,3]*

[1]Department of Dermatology, The First Hospital of China Medical University, Shenyang, China, [2]Key Laboratory of Immunodermatology, Ministry of Education and NHC, National Joint Engineering Research Center for Theranostics of Immunological Skin Diseases, Shenyang, China, [3]Institute of Respiratory Disease, China Medical University, Shenyang, China

Microorganisms, which are widely distributed in nature and human body, show unique application value in forensic identification. Recent advances in high-throughput sequencing technology and significant reductions in analysis costs have markedly promoted the development of forensic microbiology and metagenomics. The rapid progression of artificial intelligence (AI) methods and computational approaches has shown their unique application value in forensics and their potential to address relevant forensic questions. Here, we summarize the current status of microbial metagenomics and AI analysis in forensic microbiology, including postmortem interval inference, individual identification, geolocation, and tissue/fluid identification.

KEYWORDS

artificial intelligence, microbiome, machine learning, forensic microbiology, forensic science, microbial forensics

Introduction

"Microorganism" is a general term for tiny organisms that exist in nature, mainly including bacteria, viruses, and fungi, which are invisible to the naked eye or cannot be observed clearly. Microorganisms are small, simple in structure, and widely present in nature and the human body. The microbiome and metagenomics are rapidly emerging due to the progress of genome sequencing technology, improved microbial sampling methods and the rise of bioinformatics. In the era of big data, artificial intelligence and its related technologies continue to be developed and innovated, and corresponding results have been widely used in many disciplines, including forensics (Rahaman et al., 2020; Zhang et al., 2021a; Chen et al., 2022).

Postmortem interval estimation

Inference of the time of death, or the postmortem interval (PMI), is an important task during forensic examination. Host- and environment-related microbial community

succession during postmortem decay, which occurs in a regular, clock-like manner after human death, provides novel ideas for PMI inference (Metcalf et al., 2016). Johnson et al. (2016) sampled the skin microbiota in the nasal and ear canals of decomposing human cadavers to establish an algorithm for predicting the PMI, and thereby successfully demonstrated that the skin microbiota is a promising tool in forensic death investigations. The application of microbial community changes for PMI estimation has gradually become a topic of major interest in forensic research.

The oral cavity is one of the key research fields of human microbial communities, and its microbial community richness is one of the most abundant areas and the second largest human complex after gastrointestinal tract. Adserias-Garriga et al. (2017) monitored the oral microbiota of donated human bodies within 12 days after death. They found that Firmicutes and Actinobacteria are the predominant phyla in the fresh stage, Tenericutes is the predominant phyla in bloat stage, and Firmicutes is the predominant phyla in advanced decay. Dong et al. (2019) found that when the PMI was 0 h, the dominant phyla in the oral cavity of mice were Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes. Within 240 h after the death of mice, the Proteobacteria and Firmicutes always occupied the dominant position. The oral microbiota changes in mice are different from those in human decaying bodies. By constructing linear regression models between relative abundance and postmortem intervals, Gamma-proteobacteria and Proteus species were the best candidates for use to infer the PMI, especially the late PMI. The R^2 value of both constructed linear models was 0.99.

Microorganisms play a vital role in the decomposition process. However, relatively few studies are available on the postmortem migration behavior of microbial communities inside cadavers. Liu et al. (2020) assessed the microbial community structure in the brain, heart, and cecum of mice at 15 d postmortem and found that an artificial neural network (ANN) combined with the postmortem microbial dataset from the cecum was the optimal model; mean absolute error of 1.5 ± 0.8 h within 24-h decomposition and 14.5 ± 4.4 h within 15-day decomposition. This model is potential to serve as an advantageous technique in PMI inference, however, further verification is needed.

The above studies exposed cadavers to the air during decomposition. However, the microbial community in buried decomposing cadavers may be different from that in cadavers exposed to air due to different conditions such as oxygen content, humidity, light, and soil composition. Zhang et al. (2021b) analyzed postmortem microorganisms in the gravesoil, rectum, and skin of buried rats using the random forest algorithm to predict the PMI. The results showed that the predicted MAEs of the microorganisms in the rectum, cadaver skin, and gravesoil were 2.06, 2.13, and 1.82 days, respectively, within 60 days after death. This study developed the first model to predict the PMI based on microbial community succession and machine learning algorithms for buried bodies, which can provide information on the timing of buried body cases for forensic investigations.

Deel et al. (2021) placed six human donor subjects remains outdoors to decompose on the soil surface, with three samples each placed in spring and summer. Microorganisms in the skin and soil can naturally decompose the corpse to expose the ribs. The investigators developed a PMI prediction model using colonies on ribs in combination with the random forest algorithm. The accuracy of PMI prediction within 9 months was approximately (±34) days. This study represents a preliminary attempt to study the continuity of microbial communities in postmortem corpse remains, which may provide a tool for forensic investigators to estimate the time since death of skeletal remains. However, limitations remain, such as small sample sizes, differences between seasons, including differences in soil moisture, inorganic salts, and microbial contents, and variations in the organic composition of bones and other skeletal degradation indicators.

Individual identification

Individual identification is one of the most important tasks in forensic science. A number of studies have shown that microecosystems such as the skin, oral cavity, and intestine have obvious polymorphisms and individual differences. The differences in microbial community composition and abundance in human microecosystems constitute the basis of microbial use for individual identification. In theory, each individual carries a unique set of microorganisms that differs from those of other individuals, which can be identified through microbiome analysis, and this particular microbial community can persist over long periods. Therefore, microbiome characterization is potentially applicable to forensic human identification.

Franzosa et al. found that the microbiome of an individual can specifically identify its source host in a population of more than 100 people, the performance of the gut microbiome is very stable, and more than 80% of individuals can still be accurately located after 1 year (Franzosa et al., 2015). Another study found that the genotypic composition of the 16S rRNA of Cutibacterium acnes is individual-specific. The random forest machine learning method was used to combine the 16S rRNA genotype of C. acnes with the skin microbiome profile data, and the accuracy of individual identification was ~90% (Yang et al., 2019). Over time, the 16S rRNA genotype of C. acnes was more stable than that of the skin microbiome profile.

The Budowle team conducted a series of studies on the application of forensic individual identification using skin microorganisms (Schmedes et al., 2017, 2018). The core microbiota of the skin was determined, and clade-specific markers were identified. A novel targeted sequencing panel, the hidSkinPlex, was developed, which contains 286 markers covering a range of taxonomies of specific microorganisms that are in high abundance on the human skin. Schmedes et al. (2018) achieved accuracy rates between 54.20% and 100.00% when classifying eight individuals with samples from three body sites (i.e., foot,

hand and manubrium) by using regularized multinomial logistic regression and 1-nearest-neighbor classification. Woerner et al. (2019) used the same panel to classify 51 individuals across three body sites with nearest neighbor machine learning approaches. The accuracy rates of using phylogenetic distance or nucleotide diversity were 78.00% and 83.70%, respectively. As the number of individuals increased, the classification accuracy decreased.

Sherier et al. (2021, 2022) proposed that single nucleotide polymorphism genetic markers are more individualized than taxonomic markers. They designed an improved "hidSkinPlex+" system, which comprises 365 SNPs residing in 135 markers, fewer markers than the original hidSkinPlex. Eliminating the markers that do not contribute to classification accuracy can improve the enrichment process and increase the efficiency of machine learning. They reanalyzed the same sequencing data as those in Woerner et al. (2019), and found that the highest Wright's fixation index (F_{ST}) combined with support vector machine (SVM) could achieve higher accuracy in individual identification ($p = 0.03$, chi-squared test).

Tissue/fluid identification

During forensic reconstruction of crime scene activities, identification of biological traces and their bodily origin provides valuable evidence that can be presented in court. However, traces and stains at the crime scene are often exposed to the environment outside the human body for a period before being processed in the laboratory. Dobay et al. (2019) detected some characteristic microorganisms with high abundance in semen, saliva, vaginal secretions, menstrual blood, peripheral blood, and skin. The study found that samples with 30 days of indoor exposure still harbor a microbial signature that can be used to identify bodily origins. The dominant microbial signature in skin, saliva, semen are Propionibacterium, Prevotella, and Bacteroides, respectively. Vaginal fluid and menstrual blood share their microbial signatures, as Lactobacillus makes up on average 75% and 86% of the bacterial reads. Hanssen et al. (2017) used standard pattern recognition based on principal component analysis in combination with linear discriminant analysis and found that the microbial community was well differentiated between saliva and skin, and the saliva microorganisms of different individuals have specificity. The accuracy of cross-validation was 94%. Based on massively parallel sequencing of the microbiome, Díez López et al. achieved accurate tissue-type classification of skin, saliva, and vaginal secretions by using taxonomy-independent deep learning networks (Díez López et al., 2019). Body-site classification accuracy of these test samples was very high as indicated by AUC values of 0.99 for skin, 0.99 for oral, and 1 for vaginal secretion. It can also provide forensically relevant blood samples (e.g., menstrual blood, nasal blood, fingertip blood, and venous blood) with accurate information about the source of blood in the body (Díez López et al., 2020). By analyzing the sequencing data of different body parts and soil mixture samples, Tackmann et al. (2018) identified a core set of ecologically informed microbial biomarkers for human body sites. Using Generalized Local Learning, 635 operational taxonomic units (OTUs) were reported as biomarkers, between 92 (nostril) and 326 (skin). Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria were dominant in all investigated body sites. They found high fractions of positive Firmicutes and Bacteroidetes biomarkers in feces and Proteobacteria biomarkers in skin.

Geolocation

The International Metagenomics and Metadesign of Subways and Urban Biomes (MetaSUB), which was launched in 2015, is a global network of scientists and clinicians developing knowledge of urban microbiomes by studying mass transit systems, the built environment, and hospitals. In forensic casework, a link is evident between a crime scene investigation, the suspect, and an object, location, or victim. The study of environmental metagenomics also introduces potential for new forensic applications such as geographical identification.

Researchers and volunteers of MetaSUB Consortium collected ~5,000 samples from the mass transit systems of 60 cities around the world. Analysis was performed using next-generation sequencing and genome sequencing technology, and the largest set of global urban microbial metagenomics research results to date was reported (Danko et al., 2021). Public data from MetaSUB Consortium were used by multiple research teams to perform geographic origin inference by using various bioinformatics and artificial intelligence algorithms. Huang et al. (2020) extracted features from metagenomic abundance profiles. By using logistic regression with L2 normalization, the prediction accuracy of the model reached 86% to infer city affiliation. Walker and Datta (2019) analyzed whole-genome sequenced microbiota sampled from 12 cities in seven different countries. The authors applied machine learning techniques to identify the geographical provenance of the microbiome samples. Up to 90% of the samples were correctly classified, demonstrating the potential of machine learning applications in biogeography, although further evidence is necessary to extend these applications to an evidentiary context. Ryan (2019) constructed a random forest classifier based on a dataset of 311 urban microbiome samples and correctly classified 83.3% of the samples.

Conclusion

Microorganisms are widely distributed in nature and the human body. Microbial traces from the human body or crime scenes can be effectively used in forensic medicine to solve crime problems, showing huge potential and unique application value in forensic medicine. The rapid progress of artificial intelligence and its related technologies has markedly promoted the development of forensic microbiology, which has introduced novel ideas and

tools for solving the problems in forensic practice. Research is still at the preliminary stage, and many challenges need further addressed, for example, limited sample sizes, model accuracies, unrealistic environmental settings, etc. As artificial intelligence analysis in forensic identification is novel innovation, there are only limited relevant research reports. Many researchers conducted a single study from a single perspective, and there was insufficient data to cross-verify the accuracy of these results. Although we have summarized these reports, it is not known how accurate the studies are. Nonetheless, it is foreseeable that microbiome-based evidence could contribute to forensic investigations in the future.

Author contributions

QH: searched and analyzed the published literature, drafted the manuscript. XN: searched and analyzed the published literature. R-QQ and ML: reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded and supported by National Natural Science Foundation of China (82173401 and U1908206).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adserias-Garriga, J., Quijada, N. M., Hernandez, M., Rodríguez Lázaro, D., Steadman, D., and Garcia-Gil, L. J. (2017). Dynamics of the oral microbiota as a tool to estimate time since death. *Mol Oral Microbiol* 32, 511–516. doi: 10.1111/omi.12191

Chen, H., Li, C., Wang, G., Li, X., Mamunur Rahaman, M., Sun, H., et al. (2022). GasHis-transformer: a multi-scale visual transformer approach for gastric histopathological image detection. *Pattern Recogn.* 130:108827. doi: 10.1016/j.patcog.2022.108827

Danko, D., Bezdan, D., Afshin, E. E., Ahsanuddin, S., Bhattacharya, C., Butler, D. J., et al. (2021). A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cells* 184, 3376–3393.e17. doi: 10.1016/j.cell.2021.05.002

Deel, H., Emmons, A. L., Kiely, J., Damann, F. E., Carter, D. O., Lynne, A., et al. (2021). A pilot study of microbial succession in human rib skeletal remains during terrestrial decomposition. *mSphere* 6:e0045521. doi: 10.1128/mSphere.00455-21

Díez López, C., Montiel González, D., Haas, C., Vidaki, A., and Kayser, M. (2020). Microbiome-based body site of origin classification of forensically relevant blood traces. *Forensic Sci. Int. Genet.* 47:102280. doi: 10.1016/j.fsigen.2020.102280

Díez López, C., Vidaki, A., Ralf, A., Montiel González, D., Radjabzadeh, D., Kraaij, R., et al. (2019). Novel taxonomy-independent deep learning microbiome approach allows for accurate classification of different forensically relevant human epithelial materials. *Forensic Sci. Int. Genet.* 41, 72–82. doi: 10.1016/j.fsigen.2019.03.015

Dobay, A., Haas, C., Fucile, G., Downey, N., Morrison, H. G., Kratzer, A., et al. (2019). Microbiome-based body fluid identification of samples exposed to indoor conditions. *Forensic Sci. Int. Genet.* 40, 105–113. doi: 10.1016/j.fsigen.2019.02.010

Dong, K., Xin, Y., Cao, F., Huang, Z., Sun, J., Peng, M., et al. (2019). Succession of oral microbiota community as a tool to estimate postmortem interval. *Sci. Rep.* 9:13063. doi: 10.1038/s41598-019-49338-z

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J., et al. (2015). Identifying personal microbiomes using metagenomic codes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E2930–E2938. doi: 10.1073/pnas.1423854112

Hanssen, E. N., Avershina, E., Rudi, K., Gill, P., and Snipen, L. (2017). Body fluid prediction from microbial patterns for forensic application. *Forensic Sci. Int. Genet.* 30, 10–17. doi: 10.1016/j.fsigen.2017.05.009

Huang, L., Xu, C., Yang, W., and Yu, R. (2020). A machine learning framework to determine geolocations from metagenomic profiling. *Biol. Direct* 15:27. doi: 10.1186/s13062-020-00278-z

Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., et al. (2016). A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLoS One* 11:e0167370. doi: 10.1371/journal.pone.0167370

Liu, R., Gu, Y., Shen, M., Li, H., Zhang, K., Wang, Q., et al. (2020). Predicting postmortem interval based on microbial community sequences and machine learning algorithms. *Environ. Microbiol.* 22, 2273–2291. doi: 10.1111/1462-2920.15000

Metcalf, J. L., Xu, Z. Z., Weiss, S., Lax, S., van Treuren, W., Hyde, E. R., et al. (2016). Microbial community assembly and metabolic function during mammalian corpse decomposition. *Science* 351, 158–162. doi: 10.1126/science.aad2646

Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., et al. (2020). Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches. *J. Xray Sci. Technol.* 28, 821–839. doi: 10.3233/XST-200715

Ryan, F. J. (2019). Application of machine learning techniques for creating urban microbial fingerprints. *Biol. Direct* 14:13. doi: 10.1186/s13062-019-0245-x

Schmedes, S. E., Woerner, A. E., and Budowle, B. (2017). Forensic human identification using skin microbiomes. *Appl. Environ. Microbiol.* 83:e01672-17. doi: 10.1128/AEM.01672-17

Schmedes, S. E., Woerner, A. E., Novroski, N. M. M., Wendt, F. R., King, J. L., Stephens, K. M., et al. (2018). Targeted sequencing of clade-specific markers from skin microbiomes for forensic human identification. *Forensic Sci. Int. Genet.* 32, 50–61. doi: 10.1016/j.fsigen.2017.10.004

Sherier, A. J., Woerner, A. E., and Budowle, B. (2021). Population informative markers selected using Wright's fixation index and machine learning improves human identification using the skin microbiome. *Appl. Environ. Microbiol.* 87:e0120821. doi: 10.1128/AEM.01208-21

Sherier, A. J., Woerner, A. E., and Budowle, B. (2022). Determining informative microbial single nucleotide polymorphisms for human identification. *Appl. Environ. Microbiol.* 88:e0005222. doi: 10.1128/aem.00052-22

Tackmann, J., Arora, N., Schmidt, T. S. B., Rodrigues, J. F. M., and von Mering, C. (2018). Ecologically informed microbial biomarkers and accurate classification of mixed and unmixed samples in an extensive cross-study of human body sites. *Microbiome* 6:192. doi: 10.1186/s40168-018-0565-6

Walker, A. R., and Datta, S. (2019). Identification of city specific important bacterial signature for the meta SUB CAMDA challenge microbiome data. *Biol. Direct* 14:11. doi: 10.1186/s13062-019-0243-z

Woerner, A. E., Novroski, N. M. M., Wendt, F. R., Ambers, A., Wiley, R., Schmedes, S. E., et al. (2019). Forensic human identification with targeted microbiome markers using nearest neighbor classification. *Forensic Sci. Int. Genet.* 38, 130–139. doi: 10.1016/j.fsigen.2018.10.003

Yang, J., Tsukimi, T., Yoshikawa, M., Suzuki, K., Takeda, T., Tomita, M., et al. (2019). Cutibacterium acnes (*Propionibacterium acnes*) 16S rRNA genotyping of microbial samples from possessions contributes to owner identification. *mSystems* 4:e00594-19. doi: 10.1128/mSystems.00594-19

Zhang, J., Li, C., Kosov, S., Grzegorzek, M., Shirahama, K., Jiang, T., et al. (2021a). LCU-net: a novel low-cost U-net for environmental microorganism image segmentation. *Pattern Recogn.* 115:107885. doi: 10.1016/j.patcog.2021.107885

Zhang, J., Wang, M., Qi, X., Shi, L., Zhang, J., Zhang, X., et al. (2021b). Predicting the postmortem interval of burial cadavers based on microbial community succession. *Forensic Sci. Int. Genet.* 52:102488. doi: 10.1016/j.fsigen.2021.102488

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 01 February 2023
DOI 10.3389/fmicb.2023.1112010

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Ying Tu,
The First Affiliated Hospital of Kunming Medical University, China
Ke Hui,
Vietnamese – German Center of Excellence in Medical Research, Vietnam
Yifeng Guo,
Shanghai Jiao Tong University,
China

*CORRESPONDENCE
Fujun Chen
✉ 30348107@qq.com
Rui-Qun Qi
✉ xiaoqiliumin@163.com

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 30 November 2022
ACCEPTED 05 January 2023
PUBLISHED 01 February 2023

CITATION
Sun T, Niu X, He Q, Chen F and Qi R-Q (2023) Artificial Intelligence in microbiomes analysis: A review of applications in dermatology. *Front. Microbiol.* 14:1112010. doi: 10.3389/fmicb.2023.1112010

Artificial Intelligence in microbiomes analysis: A review of applications in dermatology

Te Sun[1,2], Xueli Niu[1,2], Qing He[1,2], Fujun Chen[3]* and Rui-Qun Qi[1,2]*

[1]Department of Dermatology, The First Hospital of China Medical University, Shenyang, China, [2]Key Laboratory of Immunodermatology, Ministry of Education and NHC, National Joint Engineering Research Center for Theranostics of Immunological Skin Diseases, Shenyang, China, [3]Liaoning Center for Drug Evaluation and Inspection, Shenyang, China

Microorganisms are closely related to skin diseases, and microbiological imbalances or invasions of exogenous pathogens can be a source of various skin diseases. The development and prognosis of such skin diseases are also closely related to the type and composition ratio of microorganisms present. Therefore, through detection of the characteristics and changes in microorganisms, the possibility for diagnosis and prediction of skin diseases can be markedly improved. The abundance of microorganisms and an understanding of the vast amount of biological information associated with these microorganisms has been a formidable task. However, with advances in large-scale sequencing, artificial intelligence (AI)-related machine learning can serve as a means to analyze large-scales of data related to microorganisms along with determinations regarding the type and status of diseases. In this review, we describe some uses of this exciting, new emerging field. In specific, we described the recognition of fungi with convolutional neural networks (CNN), the combined application of microbial genome sequencing and machine learning and applications of AI in the diagnosis of skin diseases as related to the gut-skin axis.

KEYWORDS

Artificial Intelligence, microbiome, machine learning, convolutional neural networks, microbial sequencing, gut-skin axis

Introduction

A tremendous array of microorganisms widely exist in nature and the human body. In healthy humans various types of microorganisms and dominant microbiota are present within different parts of the body and a stable number and ratio of microorganisms are maintained through competition or synergy (Dominguez-Bello et al., 2019). With changes in the ratio between the dominant microbiota and individual microbiota, a microecological imbalance occurs, which can lead to specific skin diseases and, the characteristics of these changes are related to the progression of the disease. Due to the large number and variety of microorganisms, analyzes using previous techniques have been incapable of handling these data (Goodswen et al., 2021). However, with the development of AI and machine learning related technologies, information based on microbial image recognition or genomics data can be applied in many fields, including the identification of specific conditions for application in forensic science and clinical disease diagnosis. In this review, we provide a detailed introduction to the application of AI as based on microbial information for use in diagnosing skin diseases and predicting disease progression of these conditions.

Recognition of fungi with convolutional neural networks

Convolutional Neural Network (CNN) is a special type of machine learning that can assimilate both isolated topographies as well as entire images and classify these images according to their unique features (Dildar et al., 2021). Dermatologists can also apply this method for disease diagnosis. For example, an image is annotated according to the corresponding medical records and pathological results and, after generating standardized data, these data can then be analyzed using CNN to distinguish and thus diagnose skin lesion images from that of normal skin images (Haenssle et al., 2018; Liopyris et al., 2022). This method is often used in the diagnosis of fungal infections, such as onychomycosis, as this condition represents the most common nail disease infected with fungi. The traditional clinical diagnosis for this condition is based on direct microscopy with potassium hydroxide (KOH), a periodic acid schiff stain (PAS) and/or a fungal culture. However, as colony formation requires an extended period of time and is susceptible to antifungal drugs, this diagnostic approach can be problematic (Gupta et al., 2020). The diagnosis of this condition, which involves the observation of a specific type of fungal morphology, is particularly suitable for that of CNN (Hogarty et al., 2020; Zhang et al., 2021).

The differences in diagnosing onychomycosis using CNN vs. manual microscopic examinations have been compared and analyzed. Results from one report found that the accuracy of CNN diagnosis was 10% greater than that of traditional diagnostic methods (Yilmaz et al., 2022). In that study, 60 nail samples from patients with onychomycosis and 297 nail samples from healthy controls were treated with KOH. Two different CNN diagnostic performance models (VGG16 and InceptionV3) were developed. These two models have different algorithms, but both accomplish the purpose of diagnosis by extracting fungal-specific structures. As compared with that of the traditional clinical method, these two CNN models not only demonstrated a higher degree of accuracy, but also showed a better sensitivity (75.04% and 74.93% vs. 74.81%) and specificity (92.67% and 93.78% vs. 74.25%). Similar findings were reported in another study with the specificity of diagnosis using VGG16 being 72.7% vs. 49.3% with the traditional clinical method, however, the sensitivity of CNN diagnosis was slightly lower at 70.2% vs. 73%, as determined in a group of 90 patients (Kim et al., 2020). In addition, CNN showed a greater degree of specificity in the diagnosis of onychomycosis as compared with conventional diagnostic methods using the periodic acid-Schiff reaction (PAS stain). As reported in a study with 199 cases, CNN showed an increased level of specificity (98% vs. 90.35%) and area under the receiver operating characteristic curve (AUC – 0.960 vs. 0.932) as compared to that obtained with three dermatopathologists (Decroos et al., 2021). Moreover, CNN can also be used as a primary screening tool to assist manual microscopic examinations to greatly improve diagnostic accuracy. As the specificity and sensitivity of machine learning can be adjusted by changing the intersection over union (IOU) parameter, the specificity of machine learning can be increased by increasing IOU to ensure a higher true positive rate. Subsequently, the clinician can re-screen samples diagnosed as negative in hyphae with use of CNN to reduce the false negative rates. Such an approach not only improves detection efficiency and reduces expenses, but also increases diagnostic accuracy (Koo et al., 2021).

In addition to nails, fungal infections within other regions, such as the skin and hair, can also be diagnosed using CNN. However, due to the expansive areas involving skin and hair, the lesions are not concentrated. As a result, the low fungal content in an individual lesion hinders the recognition of mycelial characteristics by CNN, which can then decrease the diagnostic accuracy in these regions (Gao et al., 2021). One approach to alleviate the deficiency of CNN to extract effective information from small-scale data sets, is to combine the CNN model with the attention mechanism (AM) to build an IL-MCAM framework. IL-MCAM is based on attention mechanisms and interactive learning and can be applied to add misclassified images to the training sets using an interactive approach after the images have been classified with CNN to improve the classification ability of the CNN model. Although, to our knowledge, no reports are available using IL-MCAM to diagnose fungal skin diseases, a 99.77% correct diagnosis rate for colorectal cancer has been reported with this model (Chen et al., 2022a).

Combined application of microbial genome sequencing and machine learning

Machine learning can provide the means for identifying patterns in the sequencing data of a pathogen to generate a system for subdividing that pathogen. In this way, it can be used to determine which branch of the pathogen is infected to provide a basis for administration of the most appropriate medication. This method can be applied for the diagnosis of syphilis, as the internal structure of treponema is like that of bacteria with five genera, among which treponema is the pathogenic bacteria resulting in syphilis. In contrast, the clinical diagnosis of syphilis requires serological tests including TPPA and TPHA, which lack the ability to determine which branch of the pathogen is infecting the patient, thus precluding decisions regarding the most effective medication (Forrestel et al., 2020). One example of this approach has been applied to enable an advanced determination as to whether a patient was infected with drug-resistant spirochetes. Investigators collected and sequenced treponema pallidum from syphilis-infected patients in 8 countries and 6 continents and classified the spirochete branches using machine learning – maximum likelihood phylogeny method (ss14 and Nichols). It was found that the clades recognized by this model as Nichols C and Nichols B were consistent with resistance to azithromycin (Lieberman et al., 2021). This observation not only helped in clinically diagnosing the type of syphilis, as achieved using PC, but also provided a guide with regard to the initial administrations of medications to avoid use of ineffective antibiotics for patients infected with drug-resistant strains. While promising, these findings were based on a small sample size which lacked South Asian and South American populations, which raises an issue regarding the reliability and validity of these experimental results.

Machine learning can process 16S sequencing results to obtain information on differences in microbial species and composition ratios between patients and healthy individuals. Compared with traditional diagnostic methods, machine learning has the capacity to obtain additional information regarding body skin status and pathogen type. The 16S sequencing approach is a commonly used sequencing method to reveal species composition and evolution, mainly *via* its ability to detect partial fragments of microbial ribosomal DNA. This fragment includes 9 variable and 10 conserved regions, with the former determining relationships between species and the latter providing an understanding of differences between species. Diversity information of microorganisms in a sample can be obtained by amplifying and

sequencing partial regions of the rRNA DNA sequence in the extracted sample (Abellan-Schneyder et al., 2021). As an example, males with an HPV infection may present with insidious symptoms, but their penile microbiota will change, with this change placing their sexual partners at risk for HPV infection (Onywera et al., 2020b). The V3-V4 hypervariable regions of the 16S rRNA gene from the penile skin microbiota of 238 South African males were analyzed, and 6 distinct community state types (CSTs) were identified. With use of the machine learning – linear discriminant analysis effect size algorithm, differences in the abundance of microbial populations were observed as a function of different HPV infection subtypes. High-risk (HR)-HPV males had a significantly greater relative abundance of prevotella, dialister, peptoniphilus, and unclassified clostridiales and CST types 2–6, as compared with those not infected with HR-HPV. Males with a CST type dominated by corynebacterium were less likely to be infected with HR-HPV, but the opposite was true for women (Onywera et al., 2020a). While all HPVs are contagious, different types of HPV infection can lead to different diseases including genital warts, flat warts and even genital cancer. Due to its mode of transmission (contact transmission) and the potential for latent clinical symptoms following infection, HPV infections can readily affect the health of sexual partners, if no treatments and/or protection are undertaken. Therefore, 16S sequencing and machine learning represent important tools which can be used to predict, not only the HPV type, but even the type in their partner, which enables the possibility for an early detection and treatment.

In addition to being used in the diagnosis of HPV infection, 16S sequencing has also been used in evaluating skin status and generating probabilities for the prediction of skin diseases. In one study, 1,200 microbial samples were obtained from the legs of Canadian women aged 21–65 and subjected to 16S rRNA sequencing. Combined with skin hydration status (including PH value and conductance capacitance), three machine learning methods – random forest (RF), XGBoost, and LightGBM were used to analyze these samples. In addition, samples from the legs of 278 British women were also obtained for analysis using machine learning. The results from this study revealed that skin moisture levels were higher and a better skin condition was observed as a function of increased levels of lactobacilli. With an abundance of Bergeyella, the skin was dehydrated and the probability for dermatitis was relatively high (Carrieri et al., 2021).

Metagenomics, which differs from that of 16S sequencing, directly extracts DNA from all microorganisms of environmental samples, with the detection object including all microbial genomes, due to its more prolific genome database (Gu et al., 2019). The diagnosis of acne can be performed by analyzing the metagenomic sequencing data of acne using machine learning methods. Acne, which is associated with adipogenic fibroblasts, genetic factors and skin and intestinal microbiota, is one of the most common skin diseases worldwide (Mitchell et al., 2022; Sánchez-Pellicer et al., 2022). As it remains unclear whether changes in skin microbiota play an indicative role in acne, diseased skin (DS) and healthy skin (HS) samples from 35 acne patients and 35 normal control (NC) skin samples were collected for analysis. Through metagenomics analysis, 2,520 sequence data points from each volunteer were selected. Using machine learning – principal component analysis (PCA) and kernel principal component analysis (KPCA) methods, the corresponding lipids that largely contributed to the status of each type of skin were identified. Using a multiset canonical correlation analysis (MCCA) method, lipids which can effectively differentiate among the three different skin states were revealed, with the results that lipid No. 1240 can distinguish a DS sample set, lipids No. 608 and 2334 can distinguish a HS sample set and a decrease in lipids No. 95, 1069, and 1108 indicates an improvement in the disease. Accordingly, the results of this study have significant implications with regard to the diagnosis of acne (Wang et al., 2021).

AI and skin diseases as related to the gut-skin axis

Gut microbiota play an important role in maintaining human health. The host and microbiota maintain a state of homeostasis within the body through subtle interactions, with disruptions in this balance affecting the entire organism, even within organs far removed from the gut, such as the integumentary system (De Pessemier et al., 2021). In fact, increasing evidence has accrued which indicates that many skin diseases are accompanied by alterations in the gut microbiome (e.g., atopic dermatitis, psoriasis, vitiligo, and acne vulgaris; Szántó et al., 2019). Such findings have led to development of the gut-skin axis concept. That is, when the relationship between gut microbes and the immune system is compromised, subsequent effects on the skin can be triggered and even develop into skin diseases. Therefore, skin diseases may be diagnosed through the detection of gut microbes (Mahmud et al., 2022).

Results from previous studies have shown that microbes on the skin surface are highly related to the occurrence and development of vitiligo, and the progression of this disease can be estimated by observing changes in skin microbes. For example, increased levels of streptomycin and streptococci are observed in active vs. stable vitiligo as detected by the Novaseq sequencer; and differences in Beta diversity (Non-Metric Multi-Dimensional Scaling) are present between patients with active vs. stable vitiligo (Lu et al., 2021). While the composition of gut microbiomes remains stable from infancy, skin surface microbes are susceptible to environmental influences. Only in the presence of vitiligo does the proportion of microorganisms in gut microbes change as a function of disease progression. When gut microbes of 30 patients with vitiligo were compared with that of 30 matched healthy controls, results from the 16S rRNA sequencing assay revealed that the Shannon and Simpson index was higher and the ratio of Bacteroides/Firmicutes decreased in vitiligo patients (Ni et al., 2020). The alpha-diversity (a measure of richness and uniformity) in patients experiencing vitiligo for >5 years was greater than that of patients experiencing a shorter interval of illness. Finally, when combining machine learning with assay results indicating the presence of Corynebacterium 1 and Psychrobacter, a diagnosis of vitiligo with an accuracy rate of 0.929 was obtained. These data suggest that gut microbiota can not only be used to distinguish vitiligo in patients vs. healthy individuals, but can also provide a determination for the duration of this disease.

Atopic dermatitis (AD) is a chronic inflammatory disease that may result from a complex interaction among genetic predisposition, immune dysfunction, environmental allergens and skin barrier abnormalities. Interestingly, results from previous studies have suggested that AD patients show abnormal gut microbiomes prior to the onset of this disease. Infants (2 months of age) whose fecal calprotectin was greater than normal showed an increased risk of developing AD at 6 years of age (Lunjani et al., 2018). In addition, these children had increased *E. coli*, fewer bifidobacteria, bacteroides and lower levels of alpha-diversity (Arboleya et al., 2016). Such findings indicate the importance of intestinal microbial changes in the diagnosis of AD as can be determined using machine learning (Jiang et al., 2022). In that study,

data from intestinal epithelial cell transcriptomes and flora were collected from 88 AD patients and 73 healthy controls (the average age of the healthy group was 3 months younger than that of the AD group) and the supervised machine learning pipeline—Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest Classifier (RFC) were constructed as based on 44,608 gene expression probes and 366 species of microorganisms in transcriptome and microbial databases. Fifty microbial characteristic maps as related to AD were screened, including akkermanisia, verrucomicrobia, propionibacterium, and those with the highest F1 scores (high precision 0.70 and recall 0.88), could then be used as AD predictors (Jiang et al., 2022). Finally, results from a literature review have verified that these microbial characteristics are highly correlated with AD, and therefore cannot only be used to predict AD but even distinguish among disease subtypes.

Discussion

The human microbiome is closely related to skin diseases. Accordingly, an understanding of the microbial community composition, structure, function and its changes within the skin can serve as critical indices for the diagnosis of skin diseases. With this method, different machine learning models can be used to analyze changes in the abundance, type, and composition ratio of microorganisms in different aspects of the disease state versus that of healthy people, which can then enhance the accuracy of diagnosing skin diseases. Moreover, as compared with that of traditional methods, this procedure can also serve to predict the occurrence and progression of diseases.

As this new technology currently resides in developmental stages, many limitations remain. For example, sequencing data have low rates of representation and insignificant features as well as easy under-segmentation caused by the image characteristic which can result in a reduction in the accuracy of conclusions derived with this technique (Zhang et al., 2022). In addition, it is difficult for CNN to identify spores and hyphae with varying degrees of linear curvature and the relatively low resolution of pathological images can make it difficult for AI to distinguish between serum particles and fungal elements in the images. However, this reduction in accuracy resulting from low-resolution images can be resolved by altering the method used for image generation, such as using whole slide imaging (WSI) or through application of visual transformer models to obtain more stable results (Chen et al., 2022b; Li et al., 2022). The development of 269 AI has accelerated the progress of research in the study of microorganisms. In addition to the capacity for disease diagnosis and prediction as described in this report, AI's automatic identification of microbial genomes can also be applied for forensic identification of deceased individuals and thus offers the potential to assist forensic investigators in resolving medical disputes. While combining AI with microbes remains a challenge, this technology holds great promise for applications in the fields of medicine and forensics.

Author contributions

TS: searched and analyzed the published literature and drafted the manuscript. XN and QH: searched and analyzed the published literature. R-QQ and FC: reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was funded and supported by National Natural Science Foundation of China (82173401 and U1908206).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abellan-Schneyder, I., Schusser, A. J., and Neuhaus, K. (2021). ddPCR allows 16S rRNA gene amplicon sequencing of very small DNA amounts from low-biomass samples. *BMC Microbiol.* 21:349. doi: 10.1186/s12866-021-02391-z

Arboleya, S., Watkins, C., Stanton, C., and Ross, R. P. (2016). Gut Bifidobacteria populations in human health and aging. *Front. Microbiol.* 7:1204. doi: 10.3389/fmicb.2016.01204

Carrieri, A. P., Haiminen, N., Maudsley-Barton, S., Gardiner, L. J., Murphy, B., Mayes, A. E., et al. (2021). Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences. *Sci. Rep.* 11:4565. doi: 10.1038/s41598-021-83922-6

Chen, H., Li, C., Li, X., Rahaman, M. M., Hu, W., Li, Y., et al. (2022a). IL-MCAM: an interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach. *Comput. Biol. Med.* 143:105265. doi: 10.1016/j.compbiomed.2022.105265

Chen, H., Li, C., Wang, G., Li, X., Mamunur Rahaman, M., Sun, H., et al. (2022b). GasHis-Transformer: A multi-scale visual transformer approach for gastric histopathological image detection. *Pattern Recognition*, 130. doi: 10.1016/j.patcog.2022.108827

De Pessemier, B., Grine, L., Debaere, M., Maes, A., Paetzold, B., and Callewaert, C. (2021). Gut-skin Axis: current knowledge of the interrelationship between microbial Dysbiosis and skin conditions. *Microorganisms* 9:353. doi: 10.3390/microorganisms9020353

Decroos, F., Springenberg, S., Lang, T., Päpper, M., Zapf, A., Metze, D., et al. (2021). A deep learning approach for histopathological diagnosis of onychomycosis: not inferior to analogue diagnosis by Histopathologists. *Acta Derm. Venereol.* 101:adv00532. doi: 10.2340/00015555-3893

Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., et al. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. *International Journal of Environmental Research and Public Health*, 18, 5479. doi: 10.3390/ijerph18105479

Dominguez-Bello, M. G., Godoy-Vitorino, F., Knight, R., and Blaser, M. J. (2019). Role of the microbiome in human development. *Gut* 68, 1108–1114. doi: 10.1136/gutjnl-2018-317503

Forrestel, A. K., Kovarik, C. L., and Katz, K. A. (2020). Sexually acquired syphilis: historical aspects, microbiology, epidemiology, and clinical manifestations. *J. Am. Acad. Dermatol.* 82, 1–14. doi: 10.1016/j.jaad.2019.02.073

Gao, W., Li, M., Wu, R., Du, W., Zhang, S., Yin, S., et al. (2021). The design and application of an automated microscope developed based on deep learning for fungal detection in dermatology. *Mycoses* 64, 245–251. doi: 10.1111/myc.13209

Goodswen, S. J., Barratt, J. L. N., Kennedy, P. J., Kaufer, A., Calarco, L., and Ellis, J. T. (2021). Machine learning and applications in microbiology. *FEMS Microbiol. Rev.* 45:fuab015. doi: 10.1093/femsre/fuab015

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. *Annu. Rev. Pathol.* 14, 319–338. doi: 10.1146/annurev-pathmechdis-012418-012751

Gupta, A. K., Stec, N., Summerbell, R. C., Shear, N. H., Piguet, V., Tosti, A., et al. (2020). Onychomycosis: a review. *J. Eur. Acad. Dermatol. Venereol.* 34, 1972–1990. doi: 10.1111/jdv.16394

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., et al. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann. Oncol.* 29, 1836–1842. doi: 10.1093/annonc/mdy166

Hogarty, D. T., Su, J. C., Phan, K., Attia, M., Hossny, M., Nahavandi, S., et al. (2020). Artificial intelligence in dermatology-where we are and the way to the future: a review. *Am. J. Clin. Dermatol.* 21, 41–47. doi: 10.1007/s40257-019-00462-6

Jiang, Z., Li, J., Kong, N., Kim, J. H., Kim, B. S., Lee, M. J., et al. (2022). Accurate diagnosis of atopic dermatitis by combining transcriptome and microbiota data with supervised machine learning. *Sci. Rep.* 12:290. doi: 10.1038/s41598-021-04373-7

Kim, Y. J., Han, S. S., Yang, H. J., and Chang, S. E. (2020). Prospective, comparative evaluation of a deep neural network and dermoscopy in the diagnosis of onychomycosis. *PLoS One* 15:e0234334. doi: 10.1371/journal.pone.0234334

Koo, T., Kim, M. H., and Jue, M. S. (2021). Automated detection of superficial fungal infections from microscopic images through a regional convolutional neural network. *PLoS One* 16:e0256290. doi: 10.1371/journal.pone.0256290

Li, X., Li, C., Rahaman, M. M., Sun, H., Li, X., Wu, J., et al. (2022). A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* 55, 4809–4878. doi: 10.1007/s10462-021-10121-0

Lieberman, N. A. P., Lin, M. J., Xie, H., Shrestha, L., Nguyen, T., Huang, M. L., et al. (2021). Treponema pallidum genome sequencing from six continents reveals variability in vaccine candidate genes and dominance of Nichols clade strains in Madagascar. *PLoS Negl. Trop. Dis.* 15:e0010063. doi: 10.1371/journal.pntd.0010063

Liopyris, K., Gregoriou, S., Dias, J., and Stratigos, A. J. (2022). Artificial intelligence in dermatology: challenges and perspectives. *Dermatol. Ther. (Heidelb).* 12, 2637–2651. doi: 10.1007/s13555-022-00833-8

Lu, H., Xu, J., Hu, Y., Luo, H., Chen, Y., Xie, B., et al. (2021). Differences in the skin microbial community between patients with active and stable vitiligo based on 16S rRNA gene sequencing. *Australas. J. Dermatol.* 62, e516–e523. doi: 10.1111/ajd.13721

Lunjani, N., Satitsuksanoa, P., Lukasik, Z., Sokolowska, M., Eiwegger, T., and O'mahony, L. (2018). Recent developments and highlights in mechanisms of allergic diseases: microbiome. *Allergy* 73, 2314–2327. doi: 10.1111/all.13634

Mahmud, M. R., Akter, S., Tamanna, S. K., Mazumder, L., Esti, I. Z., Banerjee, S., et al. (2022). Impact of gut microbiome on skin health: gut-skin axis observed through the lenses of therapeutics and skin diseases. *Gut Microbes* 14:2096995. doi: 10.1080/19490976.2022.2096995

Mitchell, B. L., Saklatvala, J. R., Dand, N., Hagenbeek, F. A., Li, X., Min, J. L., et al. (2022). Genome-wide association meta-analysis identifies 29 new acne susceptibility loci. *Nat. Commun.* 13, –702. doi: 10.1038/s41467-022-28252-5

Ni, Q., Ye, Z., Wang, Y., Chen, J., Zhang, W., Ma, C., et al. (2020). Gut microbial Dysbiosis and plasma metabolic profile in individuals with vitiligo. *Front. Microbiol.* 11:592248. doi: 10.3389/fmicb.2020.592248

Onywera, H., Williamson, A. L., Cozzuto, L., Bonnin, S., Mbulawa, Z. Z. A., Coetzee, D., et al. (2020a). The penile microbiota of black south African men: relationship with human papillomavirus and HIV infection. *BMC Microbiol.* 20:78. doi: 10.1186/s12866-020-01759-x

Onywera, H., Williamson, A. L., Ponomarenko, J., and Meiring, T. L. (2020b). The penile microbiota in uncircumcised and circumcised men: relationships with HIV and human papillomavirus infections and Cervicovaginal microbiota. *Front. Med. (Lausanne)* 7:383. doi: 10.3389/fmed.2020.00383

Sánchez-Pellicer, P., Navarro-Moratalla, L., Núñez-Delegido, E., Ruzafa-Costas, B., Agüera-Santos, J., and Navarro-López, V. (2022). Acne, microbiome, and probiotics: the gut-skin Axis. *Microorganisms* 10:1303. doi: 10.3390/microorganisms10071303

Szántó, M., Dózsa, A., Antal, D., Szabó, K., Kemény, L., and Bai, P. (2019). Targeting the gut-skin axis-probiotics as new tools for skin disorder management? *Exp. Dermatol.* 28, 1210–1218. doi: 10.1111/exd.14016

Wang, Y., Sun, M., and Duan, Y. (2021). Metagenomic sequencing analysis for acne using machine learning methods adapted to single or multiple data. *Comput. Math. Methods Med.* 2021, 8008731–8008711. doi: 10.1155/2021/8008731

Yilmaz, A., Göktay, F., Varol, R., Gencoglan, G., and Uvet, H. (2022). Deep convolutional neural networks for onychomycosis detection using microscopic images with KOH examination. *Mycoses* 65, 1119–1126. doi: 10.1111/myc.13498

Zhang, J., Li, C., Kosov, S., Grzegorzek, M., Shirahama, K., Jiang, T., et al. (2021). LCU-net: a novel low-cost U-net for environmental microorganism image segmentation. *Pattern Recogn.* 115:107885. doi: 10.1016/j.patcog.2021.107885

Zhang, J., Li, C., Yin, Y., Zhang, J., Grzegorzek, M.. (2022). Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer. *Artificial Intelligence Review*. doi: 10.1007/s10462-022-10192-7

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 23 March 2023
DOI 10.3389/fmicb.2023.1176339

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University,
China

REVIEWED BY
Yongchao Xu,
Wuhan University,
China
Dong Zhang,
Hong Kong University of Science and Technology,
Hong Kong SAR,
China

*CORRESPONDENCE
Qian Liu
✉ caixe_liu0222@tom.com
Xin Yang
✉ xinyang2014@hust.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 28 February 2023
ACCEPTED 09 March 2023
PUBLISHED 23 March 2023

CITATION
Chen Y, He H, Luo L, Liu K, Jiang M, Li S, Zhang X, Yang X and Liu Q (2023) Studying pulmonary fibrosis due to microbial infection *via* automated microscopic image analysis. *Front. Microbiol.* 14:1176339. doi: 10.3389/fmicb.2023.1176339

Studying pulmonary fibrosis due to microbial infection *via* automated microscopic image analysis

Yajie Chen[1†], Henghui He[2†], Licheng Luo[1†], Kangyi Liu[1], Min Jiang[2], Shiqi Li[2], Xianqi Zhang[2], Xin Yang[1*] and Qian Liu[2*]

[1]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Forensic Medicine, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Introduction: Pulmonary fibrosis is a consequential complication of microbial infections, which has notably been observed in SARS-CoV-2 infections in recent times. Macrophage polarization, specifically the M2-type, is a significant mechanism that induces pulmonary fibrosis, and its role in the development of Post- COVID-19 Pulmonary Fibrosis is worth investigating. While pathological examination is the gold standard for studying pulmonary fibrosis, manual review is subject to limitations. In light of this, we have constructed a novel method that utilizes artificial intelligence techniques to analyze fibro-pathological images. This method involves image registration, cropping, fibrosis degree classification, cell counting and calibration, and it has been utilized to analyze microscopic images of COVID-19 lung tissue.

Methods: Our approach combines the Transformer network with ResNet for fibrosis degree classification, leading to a significant improvement over the use of ResNet or Transformer individually. Furthermore, we employ semi-supervised learning which utilize both labeled and unlabeled data to enhance the ability of the classification network in analyzing complex samples. To facilitate cell counting, we applied the Trimap method to localize target cells. To further improve the accuracy of the counting results, we utilized an effective area calibration method that better reflects the positive density of target cells.

Results: The image analysis method developed in this paper allows for standardization, precision, and staging of pulmonary fibrosis. Analysis of microscopic images of COVID-19 lung tissue revealed a significant number of macrophage aggregates, among which the number of M2-type macrophages was proportional to the degree of fibrosis.

Discussion: The image analysis method provides a more standardized approach and more accurate data for correlation studies on the degree of pulmonary fibrosis. This advancement can assist in the treatment and prevention of pulmonary fibrosis. And M2-type macrophage polarization is a critical mechanism that affects pulmonary fibrosis, and its specific molecular mechanism warrants further exploration.

KEYWORDS

microbial infection, pulmonary fibrosis, microscopic image, artificial intelligence, image analysis, macrophage

1. Introduction

Microbial infections represent a major culprit behind the onset of pulmonary fibrosis. A variety of microorganisms have been identified as potential inducers of this debilitating disease, including the human T-cell leukemia virus, the human immunodeficiency virus, cytomegalovirus, Epstein–Barr virus, influenza virus, avian influenza virus, Middle East respiratory syndrome coronavirus, heavy acute respiratory syndrome coronavirus, SARS-CoV-2 (Huang and Tang, 2021), Mycobacterium tuberculosis, Chlamydia and Mycoplasma. The pathogenesis of pulmonary fibrosis due to microbial infections, especially viral infections, consists of two distinct mechanisms. The pathogenesis of viral-induced pulmonary fibrosis entails two distinct mechanisms. Firstly, direct viral damage during the infection leads to acute and heavy injury to the lungs. This acute insult results in persistent lung damage and/or abnormal wound healing, thus facilitating the progression of pulmonary fibrosis. Secondly, viral infections trigger an immune-mediated response that causes tissue damage. Upon infection, immune cells aggregate at the site of injury, releasing a vast array of pro-inflammatory and pro-fibrotic cytokines that mediate the progression of fibrosis. Consequently, the synergistic interplay between viruses and these factors culminates in sustained and substantial lung damage, ultimately leading to the development of pulmonary fibrosis.

In recent years, COVID-19 caused by SARS-CoV-2 infection has had a profound impact on the global population. In fact, over one-third of heavy COVID-19 pneumonia survivors discharged from hospitals have been found to develop pulmonary fibrosis (Han et al., 2021). Furthermore, forensic examination of deceased COVID-19 patients by the Liu Liang team has revealed copious amounts of viscous gray-white liquid and visible fibrous strands on lung sections (Liu et al., 2020). Grillo and colleagues conducted a systematic analysis of lung slice samples from eight COVID-19 patients who died in intensive care, and noted significant pulmonary fibrosis remodeling, characterized by fibroblast proliferation and alveolar occlusion (Grillo et al., 2021). This underscores the fact that pulmonary fibrosis is a significant complication that can lead to heavy illness and death in COVID-19 patients.

Tissue-resident macrophages are highly plastic cells that can polarize into classical activation phenotype (M1) macrophages or alternative activation phenotype (M2) macrophages (Sica and Mantovani, 2012; Vasse et al., 2021). M2 macrophages can be induced by various cytokines and are associated with fibrosis (Zhong et al., 2014). M2 macrophages can produce pro-fibrotic mediators such as transforming growth factor-β (TGF-β), which sustain the activation of fibroblasts and promote myofibroblast proliferation, leading to excessive deposition of extracellular matrix (ECM) and structural remodeling of lung tissue, ultimately resulting in pulmonary fibrosis and respiratory failure (Song et al., 2000). Therefore, after SARS-CoV-2 invasion, do pulmonary macrophages polarize into M2 macrophages in large numbers, participating in the initiation of pulmonary fibrosis? If the degree of lung tissue fibrosis can be correlated with the number of macrophages, it would provide solid evidence for this hypothesis.

For a long time, the gold standard for diagnosing pulmonary fibrosis has been pathological examination of the affected lung tissue, making histopathology the foundation for related research. Therefore, the analysis of microscopic images of pathological tissue sections is particularly important. However, traditional microscopic image analysis techniques have the following main limitations. Firstly, they rely on subjective judgments by humans and are experiential in nature, with results being significantly influenced by the observer's individual biases. Secondly, image analysis and cell counting only represent partial views and are difficult to analyze globally. Finally, because different staining for different markers often requires different sections of the same tissue block, even consecutive sections on different slides are difficult to accurately match for subsequent analysis, making it challenging to perform correlation analysis of different indicators in the same area. Therefore, using artificial intelligence algorithms is of high demand to effectively solve the aforementioned problems.

This paper presents an artificial intelligence-based image analysis method for registering, cropping, and classifying fibrosis degree in lung FFPE (Formalin-Fixed and Parrffin-Embedded) slice. To automate the classification of fibrosis degree, we utilize the ResNet network (He et al., 2016) to extract high-dimensional features of lung tissue from pathological microscopy images. Then, we use the self-attention mechanism of Transformer (Dosovitskiy et al., 2020) to select the most discriminative local features. We further employ semi-supervised learning to improve classification accuracy with a large number of unlabeled pathological images. Our results demonstrate that our classification model can accurately focus on the pathological tissue in the lung and classify images in a way that mimics human interpretation. Moreover, we apply the Trimap distance method to automatically count the number of macrophages in immunohistochemistry images of lung tissue. Given the characteristics of lung tissue, such as the presence of many empty areas like blood vessels and airway cavities, we employ an effective area calibration method in addition to cell counting to better reflect the positive density of the target cells. We applied the above-mentioned intelligent analysis method for lung FFPE slice images to quantitatively analyze SARS-CoV-2-induced lung fibrosis and investigate the role of macrophage polarization in the mechanism of SARS-CoV-2-induced lung fibrosis.

2. Materials and methods

2.1. Pathological tissue samples

The data used to train the algorithm model in this paper comes from 10 lung FFPE slices of COVID-19 death cases. The experimental data comes from lung tissue samples of 4 different COVID-19 death cases with pulmonary lesions, as well as 3 lung tissue samples without pulmonary lesions. HE staining, CD68 immunohistochemical staining, and CD163 immunohistochemical staining were used as staining methods, where CD68 represents total macrophages and CD163 represents M2 macrophages. HE staining was used for the classification task, while immunohistochemical staining was used for the counting task.

2.2. Preparing image analysis samples *via* image registration and cropping

Due to spatial position deviation between the original HE staining images and immunohistochemical images, we registered them using affine transformation in OpenCV to achieve optimal alignment.

Moreover, the original pathological image size was extremely large, ranging from 600 million to 1.2 billion pixels, which precluded direct input into the network for classification and counting. To overcome this challenge, we manually selected the region of interest and cropped it into patches with a resolution of 1,600 × 1,600 pixels, resulting in 1589 labeled patches. We separate all labeled patches into 1,082, 269, and 198 patches for training, validation, and testing, respectively. Furthermore, to enhance the classification accuracy, we obtained 1,420 unlabeled patches. We then used deep learning models to automatically classify registered HE staining patches and count macrophage on the corresponding immunohistochemical images. Finally, we spliced the target images to restore them to the original slice size and analyzed the corresponding classification and counting results in the corresponding regions.

2.3. Lung tissue fibrosis classification *via* deep learning

Various improvements have been made to the Visual Transformer (ViT) method (Dosovitskiy et al., 2020), including knowledge distillation (Touvron et al., 2021a), Re-attention (Zhou et al., 2022), and LayerScale (Touvron et al., 2021b). In this paper, we followed the automatic classification model Tokens-to-Token ViT (Yuan et al., 2021) which consists of a pre-trained ResNet backbone and a Transformer encoder-decoder. The ResNet network extracts high-dimensional features from image patches. We normalize these features and combine them with the positional encoding before they are input into the encoder-decoder with multi-head self-attention. The output of the encoder-decoder (i.e., the class token) is fed into a prediction head that is made up of fully connected layers. Then the output of the prediction head determines the specific classification (Figure 1). To further improve the classification accuracy, we employed the semi-supervised method which combines self-training and consistency learning (Tarvainen and Valpola, 2017), and designed a model from the perspective of pseudo-labeling (Arazo et al., 2020). We adopted a similar structure of Cross Pseudo Supervision (Chen et al., 2021) in our approach. The loss function is comprised of two parts: supervised and unsupervised. The supervised part calculates the cross-entropy loss between the predicted classes and the labels. For unlabeled data, one-hot pseudo-labels are generated using the output of another model, and the cross-entropy loss is calculated between the predicted values and the pseudo-labels generated by the other model. By merging the ResNet feature extractor and Transformer classifier, our approach is able to leverage both high-dimensional image features and long-term dependencies of sequence data for classification. This allows for training with advanced features, which reduces the need for extensive training data, making it ideal for medical datasets. Furthermore, through the use of semi-supervised learning, our model effectively integrates finely labeled and unlabeled data to improve classification accuracy.

2.4. Manual-labeled fibrosis rubric

After referencing to the refined Ashcroft score criteria (Hubner et al., 2008), our study evaluated the degree of pulmonary fibrosis in detail, assigning a score of 1–8 based on the specific morphological characteristics of the lung fibrosis. The score and detailed morphological descriptions are shown in Table 1. Based on the morphological characteristics, scores of 1–3 were classified as light, 4–5 as moderate, and 6–8 as heavy. The annotations were conducted by 5 pathologists with experience, who independently evaluated HE-stained patches in the algorithm training dataset and scored them according to the standard. The group results were then collected and compared for final calibration.

2.5. Cell counting and effective area calibration

The post-processing technique extracts the area of minimum distance in each local region of the predicted distance map, which represents the center of a cell. The number of centers detected corresponds to the number of cells in the image. Finally, the density of positive cells is calculated by adjusting for the effective tissue area. We adopt Trimap (Arteta et al., 2016) which involves three semantic segmentation classes to count the number of target cells. Specifically, the cell region is divided into two semantic classes, namely the segmentation region closer to the cell center and the cell region farther away from the center. The third semantic class is defined as the background class. To differentiate the two classes in the cell region based on proximity to the cell center, such as the orange region in Figure 2, the two regions closest to the center (black regions) can be used to separate two adherent cells. To implement Trimap, we leveraged the UNet segmentation network (Ronneberger et al., 2015) which employs a combination of upsampling and downsampling layers to extract high-level features and predict a distance map for each pixel in the input image. Since the center of a cell is typically located within regions closer to the cell centers, the post-processing algorithm extracts the local minimum region and calculates the centroid to obtain the cell center. To obtain the density of positive cells, we calculated the area of valid tissues in the patch to determine the effective tissue area. The invalid tissues are typically white in color, and threshold can be used to extract the blank connected regions in the image. Any small connected regions that are not relevant are then excluded based on their minimum area of the alveolar space in the patch. The effective tissue area is calculated by subtracting the sum of areas of all connected regions larger than the minimum area from the total area of the patch.

Finally, the density of cells in each patch is calculated as follow:

$$density = \frac{p_i}{R_e} \tag{1}$$

$$R_e = \frac{A_e}{A_{total}} \tag{2}$$

$$A_e = A_{total} - A_{white} \tag{3}$$

where R_e is the rate of effect area A_e to the total area of the patch A_{total}, A_{white}, is the area of the white regions in the patch and p_i is the predicted count of cells in the image.

Compared to existing counting algorithms (Wan and Chan, 2019; Wan et al., 2021), the Trimap distance algorithm is capable of handling

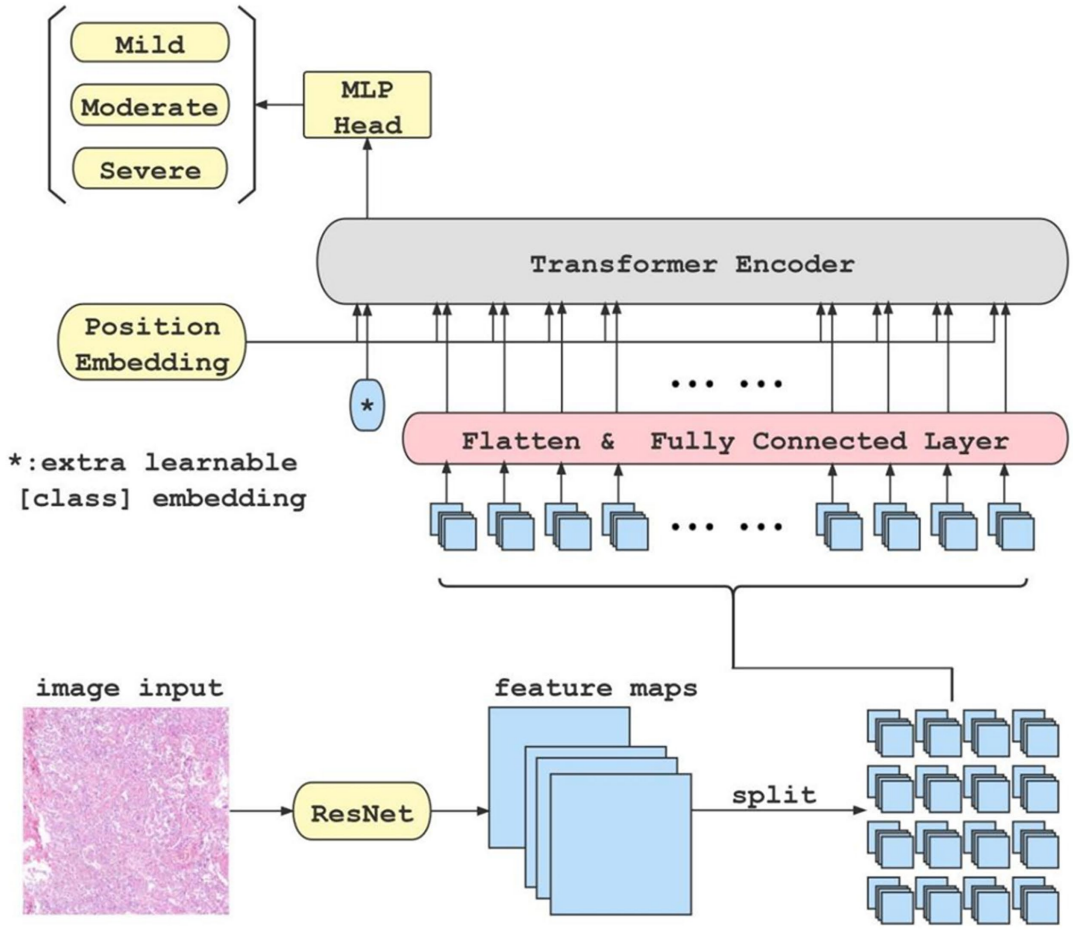


FIGURE 1
This is the flow chart of our classification model. When an original image is input to the ResNet feature extractor, the extractor outputs high-dimensional features maps of the image. Afterwards, the feature maps are split into blocks(in this case, there are 4×4 blocks), flattened, and fed into fully connected layer block by block. Meanwhile, a class token is initialized separately and fed into the transformer encoder together with each block of the output from the fully connected layer. Finally, the class token is extracted and fed into an MLP to get the classification result.

adhesive samples by utilizing the distance between pixels and the centroid and the use of effective area calibration allows the calculation of the effective positive density in combination with the cell count.

3. Results

3.1. Image registration and cropping

The original images contained approximately 900 million pixels. After image registration and manual delineation to maximize the target area, the samples were cropped into 1600×1600 patches for subsequent model training and analysis (as shown in Figure 3 for the registration and segmentation results).

3.2. Classification experiments

3.2.1. Evaluation metrics

We utilized the widely-used metrics: Precision, Recall and F1_score to evaluate the performance of fibrosis degree classification. Precision refers to the proportion of correctly predicted positive samples among all samples predicted as positive, as shown in Eq. (4). In Eqs. (4, 5) TP (True Positive) represents the number of positive samples correctly predicted as positive by the model, FP (False Positive) represents the number of negative samples incorrectly predicted as positive, and FN (False Negative) represents the number of positive samples incorrectly predicted as negative. Recall refers to the proportion of correctly predicted positive samples among all actual positive samples. F1_score is the harmonic mean of precision and recall as shown in Eq. (6), which comprehensively reflects the predictive ability and coverage of the model.

$$\text{Precision} = \frac{TP}{TP + FP} \times \%100 \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \times \%100 \tag{5}$$

$$\text{F1\_score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times \%100 \tag{6}$$

TABLE 1 Modified Ashcroft score(Hubner et al., 2008).

Rate	Description
1	Alveolar septa: Isolated gentle fibrotic changes (septum ≤3× thicker than normal)
	Lung structure: Alveoli partly enlarged and rarefied, but no fibrotic masses present
2	Alveolar septa: Clearly fibrotic changes (septum >3× thicker than normal) with knot-like formation but not connected to each other
	Lung structure: Alveoli partly enlarged and rarefied, but no fibrotic masses
3	Alveolar septa: Contiguous fibrotic walls (septum >3× thicker than normal) predominantly in whole microscopic field
	Lung structure: Alveoli partly enlarged and rarefied, but no fibrotic masses
4	Alveolar septa: Variable
	Lung structure: Single fibrotic masses (≤10% of microscopic field)
5	Alveolar septa: Variable
	Lung structure: Confluent fibrotic masses (>10% and ≤ 50% of microscopic field). Lung structure severely damaged but still preserved
6	Alveolar septa: Variable, mostly not existent
	Lung structure: Large contiguous fibrotic masses (>50% of microscopic field). Lung architecture mostly not preserved
7	Alveolar septa: Non-existent
	Lung structure: Alveoli nearly obliterated with fibrous masses but still up to several air bubbles
8	Alveolar septa: Non-existent
	Lung structure: Only fibrotic masses in microscopic field

3.2.2. Results of classification

The results of pulmonary fibrosis degree classification are shown in Table 2.

We compare our final method, i.e., ResNet+Transformer (w/SSL) with several variants, including using only ResNet18 for classification, using only Transformer for classification, using ResNet+Transformer without semi-supervised learning (SSL). Experimental results suggest that utilizing ResNet as a feature extractor in combination with a Transformer network leads to a significant improvement in performance compared to using either ResNet or Transformer individually. Additionally, using semi-supervised learning can further enhance the model's capability to focus on the target area and analyze complex samples. The normalized confusion matrix for the classification task is shown in Figure 4.

3.2.3. Visualization of lung tissue fibrosis classification

In Transformer, the key component is the multi-head attention module. In each head, the correlation between every two tokens is calculated to obtain an attention map. We focus on those values involving the class token, which represent the degree of the influence of features from different patches on the classification decision. The attention maps in all heads (4 heads) of the multi-head attention module of each layer (10 layers) are extracted, then averaged and the elements related to the class token(calculated between a patch' s token and the class token) are selected and adjusted to a two-dimensional matrix in order to match them with the corresponding locations in the original image for visualization. The shades of color are used to indicate the relative position of each vector to the class token in this two-dimensional space. Darker shades represent closer relative distances (i.e., higher correlation with the classification result), while lighter shades represent farther relative distances (i.e., lower correlation with the classification result; Figure 5). The B4 and D2 regions are the darkest in the image, indicating that the corresponding areas of pathological images have more prominent fibrosis features. As a comparison, the token map without SSL(as shown as Figure 5C), though it can also focus on some of the fibrotic areas, the areas of focus are incomplete and partially incorrect.

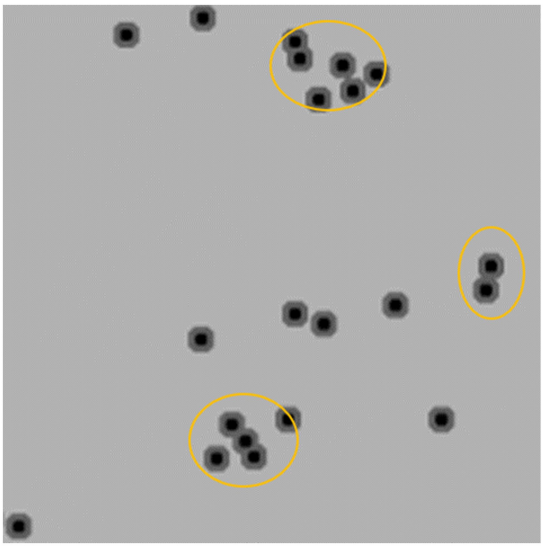


FIGURE 2
Distance map visualization.

In the global visualization map (Figure 6), the deep blue region represents heavy fibrosis classification, the lighter region represents moderate fibrosis classification, and the colorless region is not included in the study due to insufficient tissue.

3.3. Results of cell counting and calibration

The visualization results of the positive cell identified by our cell counting method in immunohistochemical staining images (Figure 7), in which Figure 7A contains some white regions, while 7B only has tissue. Direct counting and calibrated counting results of positive cells in Figure 7 are shown in Table 3.

3.4. Results of image analysis of Covid-19 lung tissue sections

3.4.1. Lung tissue image patch fibrosis classification

The HE-stained section images of the new coronal lung tissue were cut into 282 patches, and after the classification, the lung fibrosis light group (COV-L) 10 patches, the lung fibrosis moderate group (COV-M) 144 patches, the lung fibrosis heavy group (COV-H) 128

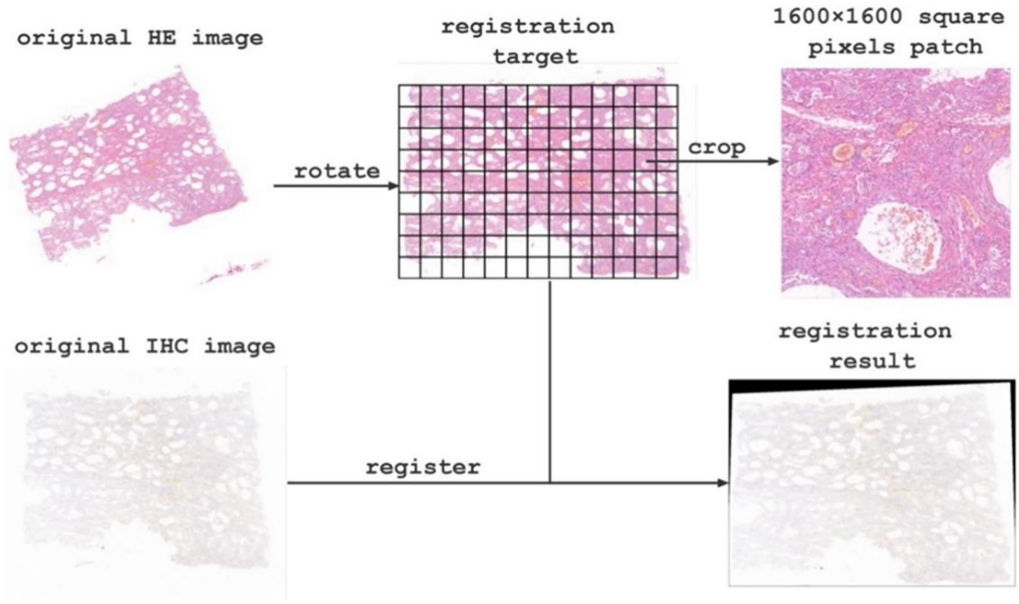


FIGURE 3
Image registration and cropping visualization.

TABLE 2 Results of pulmonary fibrosis degree classification.

Method	Precision	Recall	F1_Score
ResNet18	0.928	0.941	0.934
Transformer	0.936	0.925	0.929
ResNet+Transformer(w/ o SSL)	0.954	0.960	0.957
ResNet+Transformer(w/ SSL)	0.960	0.966	0.963

patches, and the normal lung tissue sections were segmented into 477 control groups (Control). The pathological images of lung tissues in each group are shown in Figure 8. Because the number of patches in the COVID-19 light group was too small, they were combined with COVID-19 moderate and analyzed as COVID-19 light-moderate group (COV-L/M) to reduce the error, reflecting the early stage of COVID-19 fibrosis.

3.4.2. Morphological manifestations of immunohistochemical staining of macrophages in each group of COVID-19 and control group

CD68 and CD163 immunohistochemically positive cells were yellow-brown darkly stained cells in the sections. The common results of these four immunohistochemistry groups were characterized by a higher number of CD68 positive staining fine than CD163 positive staining cells in the field of view (e.g., Figures 9–12).

3.4.3. Macrophage counts in COV-L/M Group, COV-H group and control group

The results of macrophage counts and M2/total macrophage ratios in the COV-L/M, COV-H and Control groups are shown in Table 4.

The total macrophage count and M2/total macrophage ratio were significantly higher in the COV-L/M and COV-H groups compared with the Control group, but there was no statistical difference between the COV-L/M and COV-H groups (e.g., Figures 13A,C). The M2-type macrophages were elevated in the COV-H and COV-L/M groups compared with the Control group, and the elevation in the COV-H group compared with the COV-L /M group were more significantly elevated (e.g., Figure 13B).

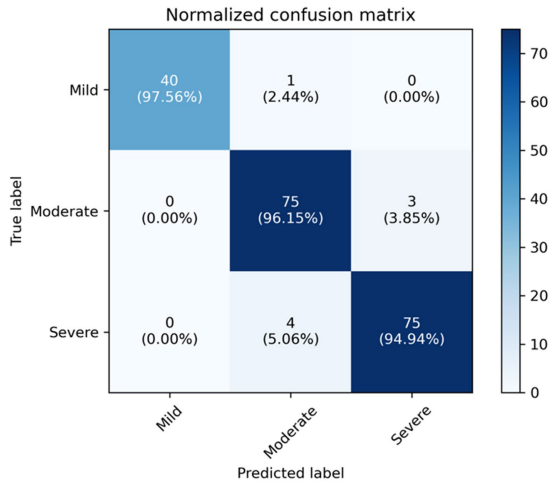


FIGURE 4
The normalized confusion matrix for the classification task.

4. Discussion

Image registration can be understood as the consistency of two or more images in spatial coordinates or visual perception. In

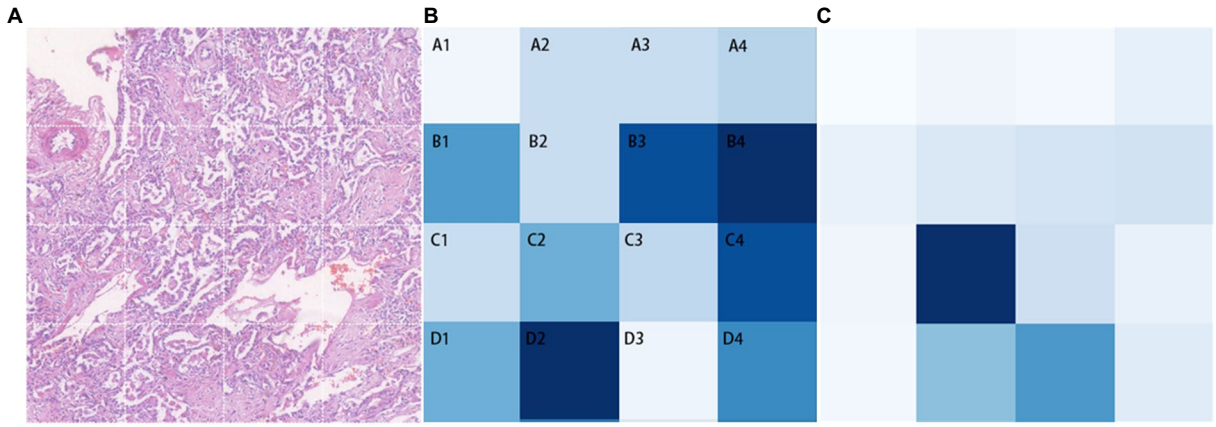


FIGURE 5
Classification task visualization. From left to right is the original input image **(A)**, the token map with SSL method **(B)** and the token map without SSL **(C)**.

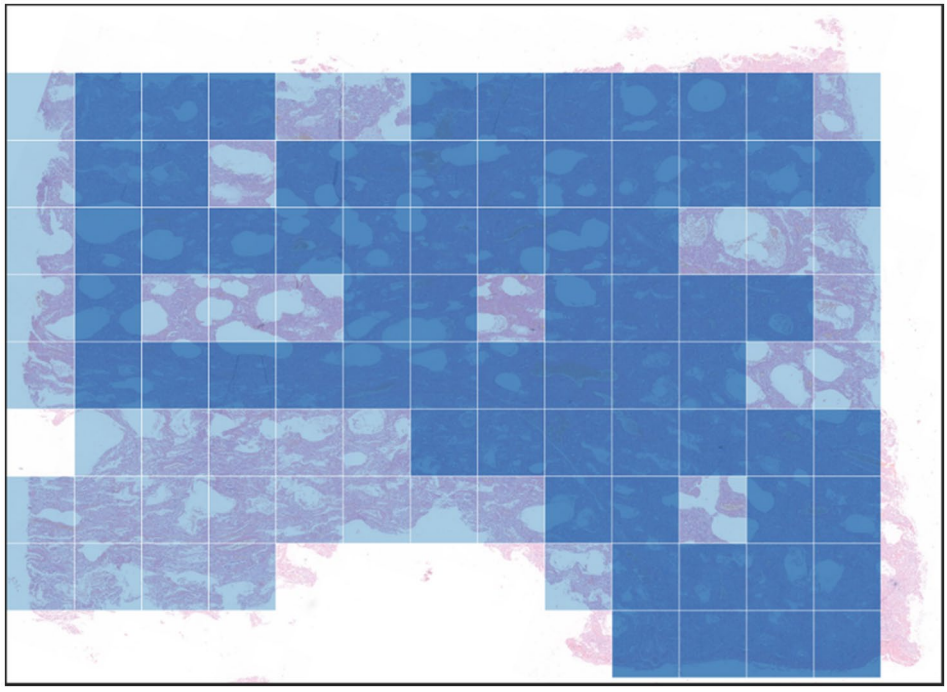


FIGURE 6
Global visualization map.

pathological research, different staining methods are used to display different structures or components of the same tissue. In this case, it is necessary to align the images of different staining methods to the same coordinate system. The AI-based image analysis method proposed in this paper solves the problem of image registration for histological sections. By registering the HE-stained images with the immunohistochemical CD68 and CD163-stained images, the differences caused by changes in orientation for each section are eliminated, allowing HE staining results to be combined seamlessly with immunohistochemical results. Therefore, the image registration method proposed in this paper can unify information from different staining methods and lay the foundation for subsequent correlation analysis of research indicators.

In the process of building the classification model, we compared several state-of-the-art models in the literature with ResNet, which served as the baseline. By utilizing various combinations of data augmentation, we were able to achieve the best performance, as shown in Table 2. In this task, since image features are highly abstract, ResNet is better equipped to capture the complex features in the images. On the other hand, the Transformer pays more attention to local features and spatial relationships. However, because only one convolutional layer is used for feature extraction, the Transformer is not as effective

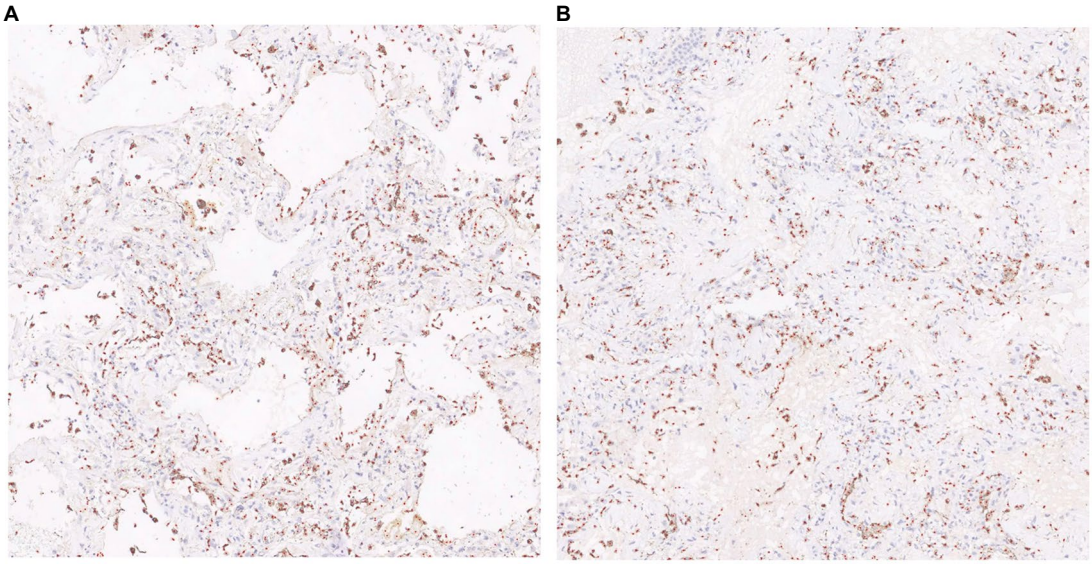


FIGURE 7
Visualization of cell prediction results (red dots are predicted cell locations). **A**:The patch contains some white regions; **B**:The patch only contains tissue.

TABLE 3 Cell count results of patch with white area (Figure 7A) and patch only contains tissues (Figure 7B).

Program	Figure 7A	Figure 7B
Before calibration	867.0	994.0
After calibration	851.0	859.0

The second and the third row correspond to the original counting results and the density results after the effective area calibration.

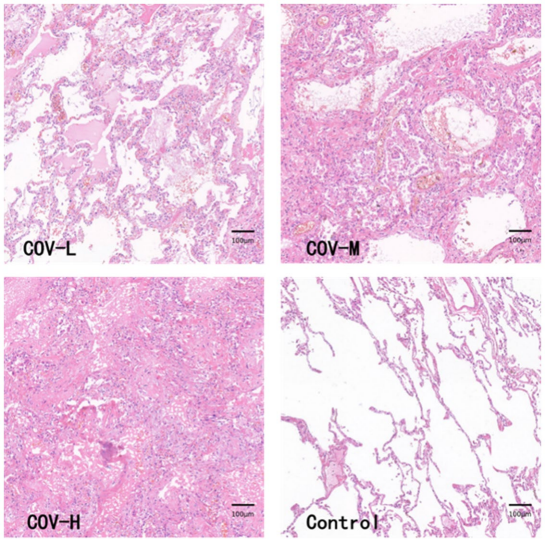


FIGURE 8
HE staining PATCH (1,600×1,600 pixels).

as ResNet in this regard. After using the pre-trained ResNet model as a feature extractor, the model was able to extract higher-dimensional features, leading to significant improvements in model accuracy, recall, and F1 score. Furthermore, after augmenting the unlabeled samples with semi-supervised learning, the model's performance was further enhanced. We observed that, with the use of SSL, our model provides more accurate and reasonable feature representations for different target regions, as evidenced by the visualized token map results.

From an application perspective, our proposed method is not only applicable to lung fibrosis classification, but also has broad potential applications in other medical imaging tasks or industrial image classification that require similar features. However, the current model also has certain limitations. For example, since the task is limited by the number of light cases, we supplemented some control cases into the light group, making it difficult for the current model to distinguish between the light and control groups. Patches with limited useful information were removed during annotation and were not included in the training set. In addition, the differences in semi-supervised learning algorithms may also affect the performance of the model. Currently, the two mainstream directions are to explore more suitable pseudo label strategies and to start with consistency learning to explore better data augmentation methods. In future, researcher could continue to optimize the model from points above.

By combining cell counting with effective area calibration, the counting algorithm presented in this paper can to some extent mitigate the influence of large blood vessel and bronchial lumens on lung parenchyma area, providing a more accurate reflection of the number of target cells in the lung parenchyma. Compared with fibrosis classification annotation, manually annotating cells in each image requires much more effort than classification, which is a major burden for researchers. Therefore, designing algorithms that can learn features from a small amount of annotation is a future research direction.

This paper is the first to apply artificial intelligence-based image analysis methods to systematically analyze lung tissue samples from COVID-19 patients. The samples analyzed in this paper showed predominantly moderate to heavy levels of lung fibrosis, with only a

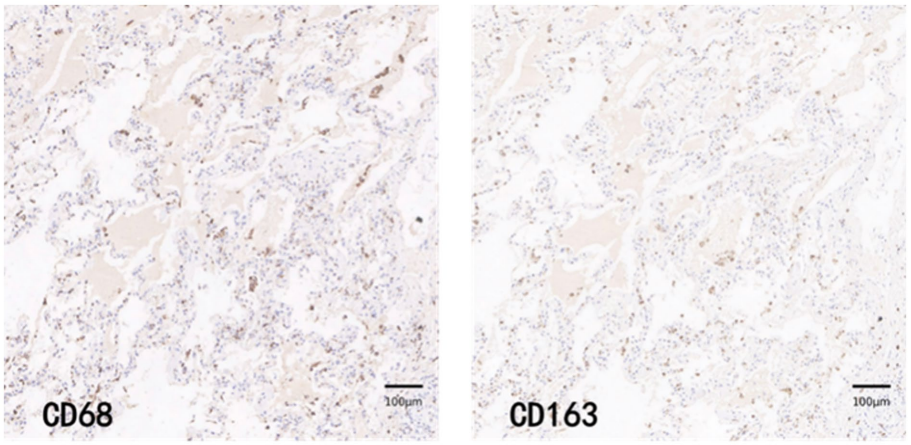


FIGURE 9
COV-L group immunohistochemical staining PATCH (1,600×1,600 pixels).

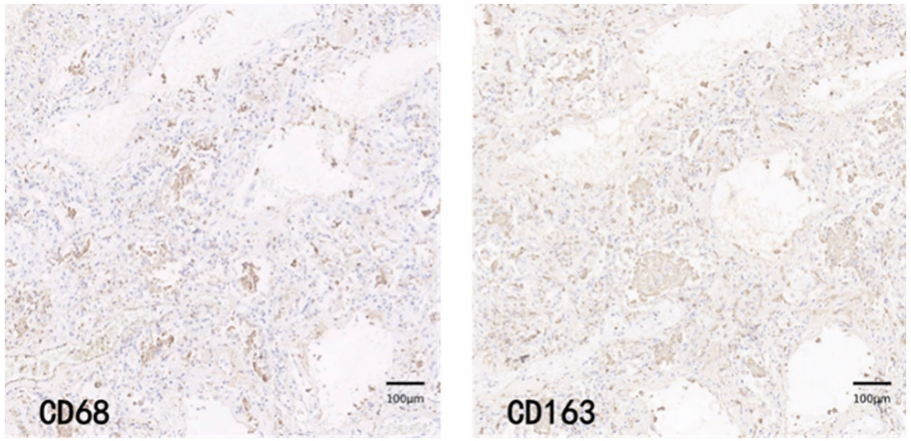


FIGURE 10
COV-M group immunohistochemical staining PATCH (1,600×1,600 pixels).

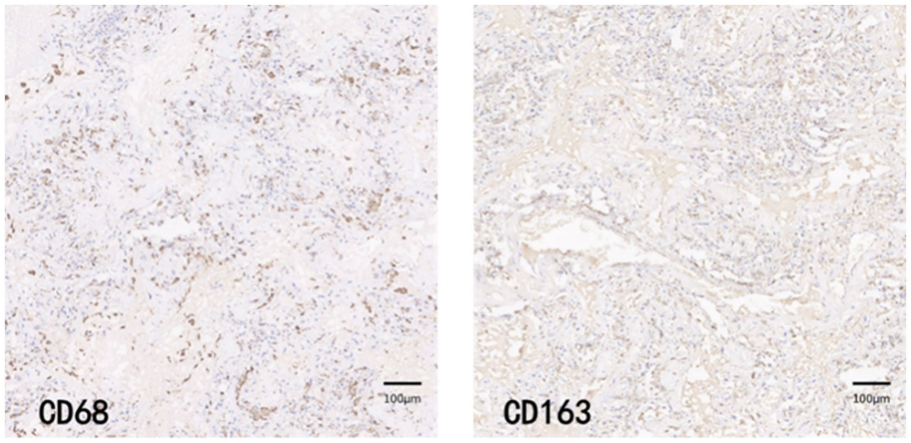


FIGURE 11
COV-H group immunohistochemical staining PATCH (1,600×1,600 pixels).

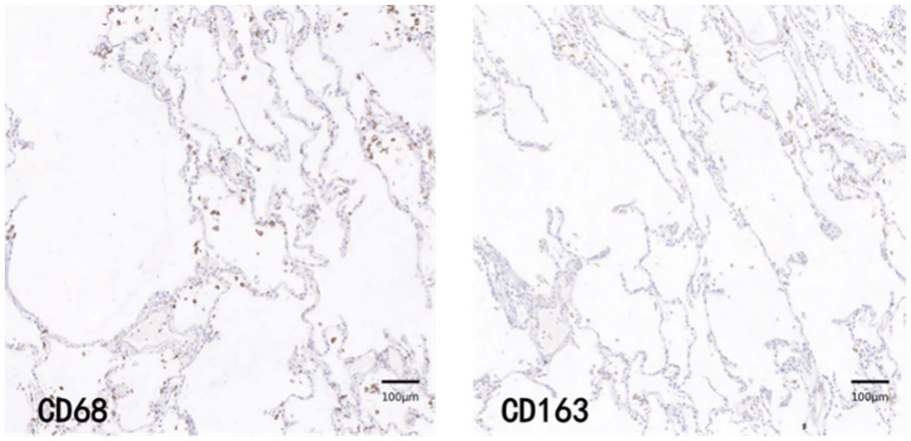


FIGURE 12
Control immunohistochemical staining PATCH (1,600×1,600 pixels).

TABLE 4 Results of macrophage counting in COVID-19 group, paraquat group and Control group.

	CD68 (Total macrophages)	CD163 (M2-type macrophages)	M2/total macrophage ratio
COVID-19 light to moderate group	940.2 ± 289.8	792.2 ± 272.2	1.015 ± 0.7863
COVID-19 heavy group	982.0 ± 355.1	1,020 ± 299.0	1.367 ± 1.070
Control group	440.7 ± 214.2	168.9 ± 135.9	0.3681 ± 0.2383

Values are mean ± standard deviation.

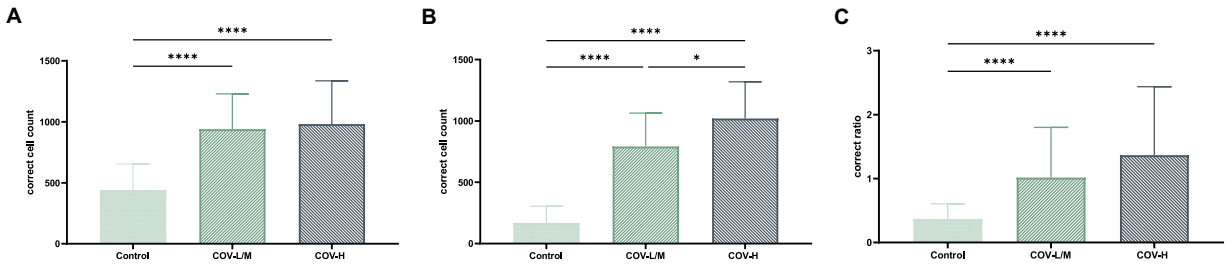


FIGURE 13
Comparative histogram of each macrophage count in each COVID-19 group and Control group **(A)** total macrophages, **(B)** M2-type macrophages, [**(C)** M2/total macrophage ratio; *p<0.05, ****p<0.0001].

small proportion showing light fibrosis, which may be attributed to the fact that the samples were obtained from COVID-19 patients with underlying conditions. HE staining revealed that the light fibrosis group had a higher incidence of exudative lesions in the pulmonary alveolar spaces, such as edema and inflammatory cells, and the pulmonary alveolar septa were slightly thickened without the formation of fibrotic nodules. In the moderate fibrosis group, fibrotic nodules appeared but did not exceed 50% of the field of view. Almost all cases in the heavy fibrosis group showed fibrotic masses with only a small amount of free space.

Through immunohistochemical staining and statistical analysis, it was found that compared to the Control group, the COVID-19 group had significantly elevated levels of total macrophages and M2-type macrophages. Furthermore, there was a positive correlation between the level of M2-type macrophages and the degree of fibrosis, with the level of M2-type macrophages being higher in the heavy fibrosis group than in the light to moderate fibrosis groups. These findings suggest that COVID-19 leads to the accumulation of macrophages and an increase in M2-type polarization, which is associated with the severity of pulmonary fibrosis. The M2/total macrophage ratio was significantly increased in all COVID-19 groups compared to the Control group, but there was no difference between the different fibrosis groups, which may be due to the increase in both total macrophages and M2-type macrophages during disease progression. These results indicate that both the total number of macrophages and the M2-type polarization phenomenon are significant factors in COVID-19-induced pulmonary fibrosis.

In summary, macrophages play an important role in the process of COVID-19-induced pulmonary fibrosis, with the

number of M2-type macrophages positively correlated with the degree of fibrosis. The specific molecular mechanism by which M2-type macrophage polarization leads to pulmonary fibrosis warrants further exploration.

5. Conclusion

Pulmonary fibrosis is not only a common complication of microbial infection of the lung but a serious threat to human health, and its pathogenesis and associated targets need to be further explored. In this paper, an artificially intelligent image analysis method is developed to align, slice, and discriminate the degree of fibrosis in microscopic images of lung tissue. The application of the artificial intelligence image analysis method constructed in this paper enables a standardized, precise, and staged study of pulmonary fibrosis, providing a more standardized method and more accurate data for the correlation of the degree of pulmonary fibrosis and aiding in the treatment and prevention of pulmonary fibrosis.

Furthermore, in this paper, we also applied this newly developed artificial intelligence image analysis method to explore the mechanism of Post-COVID-19 pulmonary fibrosis. We found that the accumulation of macrophages is a common pathological manifestation of Post-COVID-19-induced pulmonary fibrosis, in which M2-type macrophages play a major role. In the future, the related signaling molecules before and after the polarization of M2-type macrophages can be further explored to identify the key targets of Post-COVID-19-induced pulmonary fibrosis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Tongji Medical College, Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

Author contributions

QL and XY were responsible for the construction of the overall study design and the overall revision of the paper. YC was responsible for the construction of the cell counting algorithm model and calibration method. LL built the algorithm model for image alignment and automatic classification. HH was responsible for providing pathology image data and performing data analysis. KL assisted in the development of the algorithm model. MJ, XZ, and SL were involved in the annotation and analysis of the pathology image data. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Hubei Province (2021CFA053), the National Natural Science Foundation of China (62061160490) and the Applied Fundamental Research of Wuhan (2020010601012167).

Acknowledgments

The authors would like to thank the staff at HUST for their enthusiastic support in completing the groundwork.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Arteta, C., Lempitsky, V., and Zisserman, A.. (2016). *Counting in the Wild*. European Computer Vision Association, No. 9911.

Chen, X., Yuan, Y., Zeng, G., and Wang, J.. (2021). *Semi-supervised Semantic Segmentation with Cross Pseudo Supervision*. Conference on Computer Vision and Pattern Recognition (CVPR).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. Conference on Computer Vision and Pattern Recognition (CVPR).

Grillo, F., Barisione, E., Ball, L., Mastracci, L., and Fiocca, R. (2021). Lung fibrosis: an undervalued finding in covid-19 pathological series. *Lancet Infect. Dis.* 21:e72. doi: 10.1016/S1473-3099(20)30582-X

Han, X., Fan, Y., Alwalid, O., Li, N., Jia, X., Yuan, M., et al. (2021). Six-month follow-up chest ct findings after severe covid-19 pneumonia. *Radiology* 299, E177–E186. doi: 10.1148/radiol.2021203153

He, K., Zhang, X., Ren, S., and Sun, J. (2016). *Deep Residual Learning for Image Recognition*. Conference on Computer Vision and Pattern Recognition (CVPR).

Huang, W. J., and Tang, X. X. (2021). Virus infection induced pulmonary fibrosis. *J. Transl. Med.* 19:496. doi: 10.1186/s12967-021-03159-9

Hubner, R. H., Gitter, W., El Mokhtari, N. E., Mathiak, M., Both, M., Bolte, H., et al. (2008). Standardized quantification of pulmonary fibrosis in histological samples. *BioTechniques* 44, 507–517. doi: 10.2144/000112729

Liu, Q., Wang, R. S., Qu, G. Q., Wang, Y. Y., and Liu, L. (2020). Gross examination report of a covid-19 death autopsy. *Fa Yi Xue Za Zhi* 36, 21–23. doi: 10.12116/j.issn.1004-5619.2020.01.005

Ronneberger, O., Fischer, P., and Brox, T.. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Medical Image Computing and Computer-Assisted Intervention(MICCAI).

Sica, A., and Mantovani, A. (2012). Macrophage plasticity and polarization: *in vivo* veritas. *J. Clin. Invest.* 122, 787–795. doi: 10.1172/JCI59643

Song, E., Ouyang, N., Horbelt, M., Antus, B., Wang, M., and Exton, M. S. (2000). Influence of alternatively and classically activated macrophages on fibrogenic activities of human fibroblasts. *Cell. Immunol.* 204, 19–28. doi: 10.1006/cimm.2000.1687

Tarvainen, A., and Valpola, H.. (2017). *Mean Teachers are Better Role Models: Weight-averaged Consistency Targets Improve Semi-Supervised Deep Learning Results*. Neural Information Processing Systems(NIPS).

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jegou, H.. (2021a). *Training Data-efficient Image Transformers and Distillation through Attention*. The 38th International Conference on Machine Learning.

Touvron, H., Cord, M., Sablayrolles, A., Synnaeve, G., and Jégou, H. (2021b). *Going Deeper with Image Transformers*. Conference on Computer Vision (ICCV).

Vasse, G. F., Nizamoglu, M., Heijink, I. H., Schleputz, M., van Rijn, P., Thomas, M. J., et al. (2021). Macrophage-stroma interactions in fibrosis: biochemical, biophysical, and cellular perspectives. *J. Pathol.* 254, 344–357. doi: 10.1002/path.5632

Wan, J., and Chan, A. (2019). *Adaptive Density Map Generation for Crowd Counting*. International Conference on Computer Vision(ICCV).

Wan, J., Liu, Z., and Chan, A. B. (2021). *A Generalized Loss Function for Crowd Counting and Localization*. Conference on Computer Vision and Pattern Recognition(CVPR).

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z.-H., et al. (2021). *Tokens-to-token ViT: Training Vision Transformers from Scratch on ImageNet*. International Conference on Computer Vision (ICCV).

Zhong, B., Yang, X., Sun, Q., Liu, L., Lan, X., Tian, J., et al. (2014). Pdcd4 modulates markers of macrophage alternative activation and airway remodeling in antigen-induced pulmonary inflammation. *J. Leukoc. Biol.* 96, 1065–1075. doi: 10.1189/jlb.3A0313-136RRR

Zhou, D., Kang, B., Jin, X., Yang, L., Lian, X., Jiang, Z., et al. (2022). *DeepViT: Towards Deeper Vision Transformer*. Conference on Computer Vision and Pattern Recognition(CVPR).

frontiers | Frontiers in Microbiology

TYPE Mini Review
PUBLISHED 30 March 2023
DOI 10.3389/fmicb.2023.1163741

OPEN ACCESS

EDITED BY
Chen Li,
Northeastern University, China

REVIEWED BY
Dong Zhao,
China University of Political Science and Law, China
Libing Yun,
Sichuan University, China

*CORRESPONDENCE
Dawei Guan
✉ dwguan@cmu.edu.cn
Rui Zhao
✉ rzhao@cmu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

RECEIVED 11 February 2023
ACCEPTED 08 March 2023
PUBLISHED 30 March 2023

CITATION
Yuan H, Wang Z, Wang Z, Zhang F, Guan D and Zhao R (2023) Trends in forensic microbiology: From classical methods to deep learning. *Front. Microbiol.* 14:1163741. doi: 10.3389/fmicb.2023.1163741

Trends in forensic microbiology: From classical methods to deep learning

Huiya Yuan[1,2†], Ziwei Wang[3†], Zhi Wang[3], Fuyuan Zhang[3], Dawei Guan[2,3*] and Rui Zhao[2,3*]

[1]Department of Forensic Analytical Toxicology, China Medical University School of Forensic Medicine, Shenyang, China, [2]Liaoning Province Key Laboratory of Forensic Bio-Evidence Science, Shenyang, China, [3]Department of Forensic Pathology, China Medical University School of Forensic Medicine, Shenyang, China

Forensic microbiology has been widely used in the diagnosis of causes and manner of death, identification of individuals, detection of crime locations, and estimation of postmortem interval. However, the traditional method, microbial culture, has low efficiency, high consumption, and a low degree of quantitative analysis. With the development of high-throughput sequencing technology, advanced bioinformatics, and fast-evolving artificial intelligence, numerous machine learning models, such as RF, SVM, ANN, DNN, regression, PLS, ANOSIM, and ANOVA, have been established with the advancement of the microbiome and metagenomic studies. Recently, deep learning models, including the convolutional neural network (CNN) model and CNN-derived models, improve the accuracy of forensic prognosis using object detection techniques in microorganism image analysis. This review summarizes the application and development of forensic microbiology, as well as the research progress of machine learning (ML) and deep learning (DL) based on microbial genome sequencing and microbial images, and provided a future outlook on forensic microbiology.

KEYWORDS

forensic microbiology, forensic medicine, machine learning, deep learning, artificial intelligence

Introduction

Microorganisms, including viruses, bacteria, and fungi, are ubiquitously distributed and form diverse and rich communities (Oliveira and Amorim, 2018). Forensic microbiology is a discipline that deals with the study of microbiology in the context of forensic investigation (Kuiper, 2016). Its applications mainly focus on the diagnosis of cause of death, inference of crime location, estimation of postmortem interval (PMI), and individual identification.

Numerous studies have explored the application of microorganisms in forensic practice. Certain specific microorganisms have been proven to contribute to the determination of various causes of death, including drowning (Lucci et al., 2008; Kakizaki et al., 2010; Tie et al., 2010; Huys et al., 2012; Lee et al., 2016, 2017; Rácz et al., 2016; Wang et al., 2020), poisoning (Butzbach, 2010; Skopp, 2010; Grad et al., 2012; Han et al., 2012; Butzbach et al., 2013; Sastre et al., 2017), and hospital-acquired infections (Klevens et al., 2007; Sodhi et al., 2013; Lax et al., 2015; Khan et al., 2017). Considering the distinctive microbial community in soil obtained from a crime scene and other intermediary sites, soil microbiome provides evidence for inferring the geolocation (Meyers and Foran, 2008; Costello et al., 2009; Jesmok et al., 2016; Habtom et al., 2019; Yang et al., 2021). Moreover,

since microorganisms contribute to postmortem decomposition and have a succession that follows a predictable pattern (Hauther et al., 2015; Javan et al., 2016; DeBruyn and Hauther, 2017; Lutz et al., 2020; Scott et al., 2020), forensic researchers began to explore the feasibility of microbiome succession for inferring the postmortem interval over the past few years. Some studies used machine learning and other technologies to establish models for PMI estimation based on microorganisms, which actually improve the accuracy of PMI inference (Pechal et al., 2014; Johnson et al., 2016; Liu et al., 2020; Zou et al., 2020). In addition, the microbiome composition in humans varies in body location (Kong and Segre, 2012); host characteristics, such as sex, age, and lifestyle (Ross et al., 2017); and skin care status (Bouslimani et al., 2019). Therefore, individuals may be identified using skin, hair, and body fluid microbiomics (Lax et al., 2014; Bäckhed et al., 2015; Wu et al., 2016; Schmedes et al., 2017; Willis et al., 2018).

In this review, we aim to summarize the developmental progress of forensic microbiology from classical methods to high-throughput data combined with artificial intelligence technologies and discuss the outlook for the future.

Development of forensic microbiology

Forensic microbiology first gained global recognition in 2001 as a result of the *Bacillus anthracis* attacks through the USA postal service. In previous studies on forensic microbiology, forensic microbiological technologies were not specifically described, except for agar cultures for bacteria and fungi combined with PCR for certain microorganisms (Aoyagi et al., 2009; Huys et al., 2012; Uchiyama et al., 2012; Tuomisto et al., 2013; Can et al., 2014; Hauther et al., 2015; Yu et al., 2021a,b). To date, though more than 2,460 different species are presented in The Ribosomal Database Project stores (Maidak et al., 2000), most microbes in the environment have not been described and accessed for biotechnology. Few viable bacteria can be cultivated on artificial media (Kimura and Nobutada, 2006; Cecchini et al., 2012). Furthermore, traditional microbial culture highly relies on culture conditions and has limitations in the analysis of microbial community composition.

The 16/18S rRNA is the common marker for microbial classification and identification. Fluorescence *in situ* hybridization (FISH) (Langendijk, 1995) and denaturing gradient gel electrophoresis (DGGE) (Muyzer et al., 1993) are used to detect the specific 16S rRNA and gain access to unculturable microbes. Sanger sequencing technology (Sanger et al., 1977), also known as first-generation sequencing technology, allows for the wide use of 16S rRNA gene sequencing in bacterial taxonomy and leads to the discovery of a large number of new microbial taxa. Next-generation sequencing technology (NGS) can measure tens of thousands to millions of DNA simultaneously and can provide a high-throughput microbiome database (Metzker, 2010). With the development of sequencing technology, third-generation sequencing technology has more advantages in the study of community diversity due to its ultra-long sequencing reading (Franzén et al., 2015). However, the 16S rRNA test results are assembled into operational taxonomic units (OTUs). Most high-throughput sequencing microbiome data could not be identified and classified due to the limitation of the referenced genomes and genetic datasets (Yooseph et al., 2013; Afshinnekoo et al., 2015). Using targeted 16S rRNA with short amplifiers could not achieve reliable resolution at the species level, and full-length 16S rRNA sequences do not necessarily reduce this limitation (Forney et al., 2004). In addition, whole-genome shotgun (WGS) targets all gene content in microbial ecosystems and can differentiate microbial species and taxa to a greater extent than 16S rRNA amplicons (Schloissnig et al., 2013; Franzosa et al., 2015).

Though the expanding role of NGS and WGS combined with artificial intelligence will likely be a routine tool in forensic microbiology, the isolation, detection, and confirmation of specific microbes (pathogens or colonies) and the use of nucleic acid sequencing remain less relevant in resolving forensic challenges. Some investigators suggest the use of nanotechnology to design biosensors for the identification of foreign pathogens. Recently, microbial computer image analysis technology combined with machine learning and deep learning elucidated the identification of specific microbiomes (Ma et al., 2022). Moreover, some challenges related to standardization, validation, and procedural and bioinformatic pipelines persist in the study of forensic microbiology (Roesch et al., 2009; Lauber et al., 2010; Wu et al., 2010). Further advancements in technology will continue to improve the application of microbiology in forensic medicine. The trend in forensic microbiology is summarized in Figure 1.

Applications of machine learning in forensic microbiology

Establishing a prognostic model and predicting the dynamic succession of microbial communities have improved research in forensic microbiology. High-throughput microbiotic datasets combined with machine learning extend the application of microbiomes in forensic issues. Traditional statistical methods could only determine the general composition of a microbial community and its basic succession, whereas machine learning models achieve quantitative analysis and accurate prediction (Zou et al., 2020). Currently, different machine learning models, including random forest (RF), support vector machine (SVM), and AdaBoost, are used in the field of forensic microbiology, and they have become a promising strategy for several forensic events.

One main use of forensic microbiology is human individual identification (Fierer et al., 2010; Lax et al., 2015; Leake et al., 2016; Williams and Gibson, 2017; Yang et al., 2019). In cases without blood or tissue evidence, a microbiological approach provides a breakdown of situations with the detection of the microbiome on skin, saliva, hair, or objects that have been touched. UniFrac metric (Fierer et al., 2010) and hierarchical clustering (Leake et al., 2016) are used in determining skin-associated bacteria and salivary microbiomes. In an RF model used for investigating pubic hair bacteria, the accuracy of individual identification was 2.7 $\pm$ 5.8%, and the gender accuracy was 1.7 $\pm$ 5.2% (Williams and Gibson, 2017).

In addition, machine learning can be used for PMI estimation in microbial forensics (Johnson et al., 2016; Liu et al., 2020, 2021; Cao et al., 2021; Cartozzo et al., 2021; Randall et al., 2021; Zhang et al., 2021; Kaszubinski et al., 2022; Zhao et al., 2022).

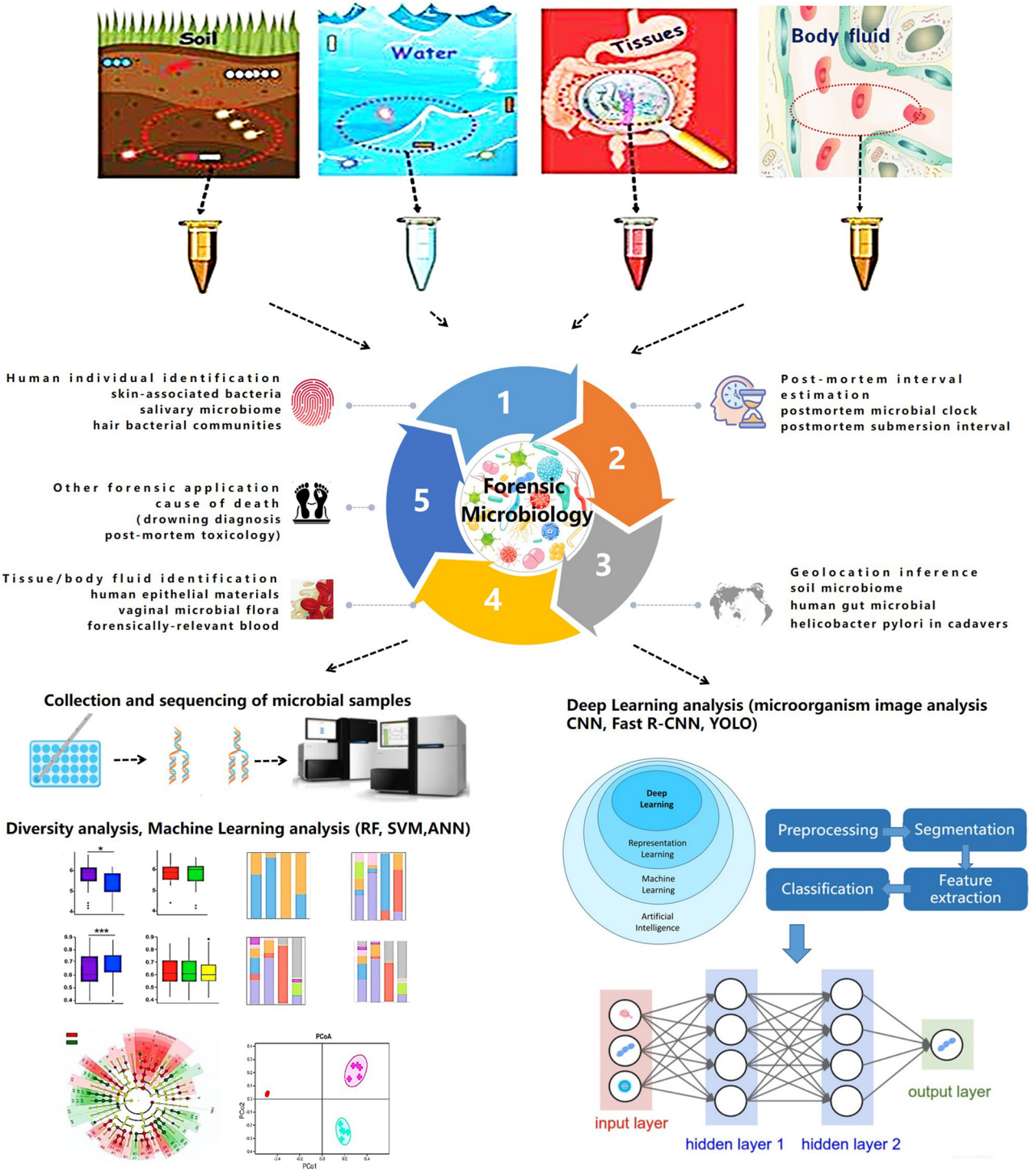


FIGURE 1
Trends in forensic microbiology.

RF models are widely used for PMI estimation and postmortem submersion interval (PMSI) diagnosis (Metcalf et al., 2013; Pechal et al., 2014; Cartozzo et al., 2021; Randall et al., 2021; Zhang et al., 2021; Kaszubinski et al., 2022; Zhao et al., 2022). In 2013, Metcalf et al. (2013) first proposed the concept of a "microbial clock" to infer PMI by setting up regression models using an RF. In addition, artificial neural network (ANN) models are used to infer PMI based on the microbiotic OTU data. In 2020, Liu et al. combined microbial community characterization and machine learning algorithms (RF, SVM, and ANN) to investigate microbial succession patterns during corpse decomposition and estimate PMI with a mean absolute error (MAE) of 1.5 ± 0.8 h within 24 h.

Available evidence demonstrates that forensic microbiology with machine learning can be used to infer geolocation to a certain extent (McNulty et al., 2004; Kersulyte et al., 2010; Blaser et al., 2013; Nagasawa et al., 2013; Tyakht et al., 2013; Escobar et al., 2014; Li et al., 2014; Suzuki and Worobey, 2014; Huang et al., 2020). Human gut microbiota plays a significant role in geolocation inference, which is supported by correlations based on traditional statistical methods, such as Wilcoxon's rank-sum test (Suzuki and Worobey, 2014) and analysis of similarities (ANOSIM) (Escobar et al., 2014). In 2014, UniFrac, network, and ANOSIM were used to analyze the human saliva microbiome. The variance between individuals was 6.75–10.21% (Li et al., 2014). In 2020, Huang et al. (2020) applied a machine learning framework to determine geolocations with an accuracy of 86%.

As previously stated, machine learning can be used to establish models for tissue/body fluid identification (Costello et al., 2009; Benschop et al., 2012; Lopez et al., 2019, 2020; Salzmann et al., 2021). Lopez et al. (2019) used ANN to identify different human epithelial biomaterials with AUC values of 0.99, 0.99, and 1 for skin, oral, and vaginal secretions, respectively. Later in 2020, deep neural network (DNN) was additionally used to identify human forensically relevant blood samples successfully (Lopez et al., 2020).

Investigative forensic microbiology with machine learning has practical applications in other aspects. Concerns related to drowning diagnosis (Huys et al., 2012; Wang et al., 2020), postmortem toxicology (Kaszubinski et al., 2020), and disease diagnosis are summarized in Table 1. Although the expanding effect of machine learning on NGS and WGS databases is without much doubt, some models derived from different machine learning models still mean less to the actual situation, which limits their application in forensic practice. The current challenges of machine learning are insufficient training samples and an imperfect microbial databank.

Deep learning for microorganisms

Compared with machine learning, deep learning can realize automatic feature learning through advanced network structure by combining several simple modules (Lecun et al., 2015; Ma et al., 2022). The deep learning process generally needs a large amount of training data followed by the formation of suitable neural networks. Previous studies have proven that convolutional neural networks (CNNs) and their derivative models allow for accurate tissue-type classification and microbiome identification, which could further extend the diagnosis of microbially caused death and personal identification in forensic science. A recent study constructed a CNN model through micorbiome analysis to classify three different human epithelial materials of skin, oral, and vaginal origins (Lopez et al., 2019). Other researchers proposed and designed the CNN based on a system for detecting fungi, parasite ova, and bacteria (Akintayo et al., 2016; Tahir et al., 2018). The accuracy for identifying fungi, soybean cyst nematode eggs, and bacteria was 94.8% (Tahir et al., 2018), 94.33% (Akintayo et al., 2016), and 96% (Treebupachatsakul and Poomrittigul, 2019), respectively.

Recent studies have demonstrated that CNN models could differentiate bacteria and algae based on microbiome images. The application of microbial computer image analysis mainly focuses on the segmentation, clustering, classification, and counting of microorganisms (Ma et al., 2022). The CNN models based on images for microbiome identification are supported by the study of Panicker (Oomman et al., 2018), which could detect tuberculosis in microscopic sputum smear images with a precision of 78.4%. In addition, deep learning on image analysis is mainly focused on the identification of diatoms and algae, which would provide a potential strategy for the diagnosis of drowning. In 2018, Bueno et al. (2018) compared the functions of R-CNN and you only look once (YOLO) in diatom detection; the F-measure of YOLO was 84%. In 2019, Huang et al. applied and trained a CNN model based on the GoogLeNet Inception V3 architecture to identify diatoms with a validation rate of 97.33% (Zhou et al., 2019, 2020), which indicates that DL is an efficient and low-cost automated diatom detection technology. In 2020, YOLO and SegNet were compared by Salido et al. (2020). In 2020, Ruis-Santaquiteria et al. (2020) found that Mask-RCNN and SegNet models are capable of segment diatoms from the same raw images used for manual identification, without any cropping or preprocessing step. With the assistance of necessary operation systems, such as the Center for Microbial Ecology Image Analysis System (Dazzo and Niccum, 2015), image analysis of microorganisms would contribute more to forensic microbiology due to its convenience and operational speed (Ma et al., 2022).

Future outlook

Forensic microbiology is still in its infancy. Advanced technologies, such as NGS and WGS, have provided a sufficient dataset that could not be imagined previously. Further advancements in technology will continue to improve the capacity for forensic microbial investigations. In addition, artificial intelligence has promoted the technological innovation of forensic microbiology in the past decades. With the accumulation of genomic sequencing datasets and microbial images, deep learning will exert greater power in achieving better prognostic models with higher accuracy. However, there are several issues that deserve further discussion.

1. Advantages and defects of AI

Based on the current microbiota database, ML and DL have the quick and automatic capability of relatively accurate prediction compared with the conditional methods. The "black box" of AI methods would not completely and theoretically characterize the real results. Although the new algorithms of AI will shed light on microbiotic analysis for forensic purposes, the basic strategy is focused on the bigger microbiota database to promote AI to work better.

2. Key points for establishing the forensic microbiota database

Considering the variation of microbial communities, the development of a forensic microbial bank is urgently needed. The sampling and sequencing procedures of forensic microorganisms still need to be further standardized. In

TABLE 1 Applications of artificial intelligence in forensic microbiology.

Forensic issue	Animal model	Sampling location	AI model	Model performance	References
Human individual identification					
	Human	Skin, Keyboard, Smartphone screen	RF	Accuracy around 90%	Yang et al., 2019
	Human	Shoes and phones	RF	Error ratio 3.6	Lax et al., 2015
	Human	Skin, computer keyboards	ANOSIM	Unweighted PC_O1 (17%) PC_O2 (6.5%)	Fierer et al., 2010
				Weighted PC_O1 (61%) PC_O2 (19%)	
	Human	Saliva	Hierarchical clustering		Leake et al., 2016
	Human	Hair	RF	Individual 2.7% ± 5.8%	Williams and Gibson, 2017
				Gender 1.7% ± 5.2%	
Post-mortem interval estimation					
	Mice	Cecum	RF	MAE 20.01 h	Liu et al., 2021
	Mice	Brain, heart, and cecum	RF, SVM, ANN	MAE within 24 h	Liu et al., 2020
	Swine	Bone	RF	Variation >80%	Kaszubinski et al., 2022
	Pig	Bone	RF	RMSE ± 104 days Rib	Randall et al., 2021
				RMSE ± 63 days Scapulae	
	Pig	Bone	RF	RMSE ± 27 days Rib	Cartozzo et al., 2021
				RMSE ± 29 days Scapulae	
	Rat	Oral	RF	$R^2 = 93.94\%$	Zhao et al., 2022
	Rat	Gravesoil, rectum and skin	RF	MAE 1.82 days in Gravesoil	Zhang et al., 2021
				MAE 2.06 days in Rectum	
	Human	Nasal cavity and ear canal	K-neighbors regression (KNR), Ridge regression (RR), Lasso regression (LR), Elastic net regression (ENR), Random forest regression (RFR), Bayesian ridge regression (BRR)	MAE ± 55ADD	Johnson et al., 2016
	Rat	Cecum	Partial least squares (PLS)	RMSE within 9 days	Cao et al., 2021
Geolocation inference					
	Human	Gut	Spearman's correlations	Positive correlation between Firmicutes abundance and latitude ($\rho = 0.857$, $p < 0.0001$)	Suzuki and Worobey, 2014
			Wilcoxon rank sum test	Negative correlation between Bacteroidetes and latitude ($\rho = -0.637$, $p = 0.001$)	
	Human	Gut	ANOSIM	ADONIS: $R^2 = 0.22$, $P = 0.001$	Escobar et al., 2014
				ANOSIM: $R = 0.78$, $P = 0.001$	
	Human	Gut	ANOSIM	Russian and the US, Danish and Chinese groups $R = 0.74$, 0.50 and 0.26, respectively $P = 9.999 * 10^{-5}$	Tyakht et al., 2013
	Human	Skin	ANOVA, ANOSIM, PERMANOVA	US vs. VZ; ANOSIM $P < 0.001$; PERMANOVA $P < 0.001$ for both unweighted and weighted measures	Blaser et al., 2013
	Human	Saliva	UniFrac, network, ANOSIM	Variance between individuals: 6.75–10.21% within individuals: 89.79–93.25%	Li et al., 2014

(Continued)

TABLE 1 (Continued)

Forensic issue	Animal model	Sampling location	AI model	Model performance	References
				ANOSIM statistic: $R = -0.0935$, P-value = 0.7386	
	Human	Subways and urban biomes	Logistic regression model with L_2 regularization	Accuracy 86%	Huang et al., 2020
Tissue/body fluid identification					
	Human	Skin, oral and vaginal	ANN	AUC values of 0.99 for skin, 0.99 for oral, and 1 for vaginal secretion	Lopez et al., 2019
	Human	Blood	DNN	0.978 for nasal blood	Lopez et al., 2020
				0.978 for finger-prick blood	
	Human	Vagina	Microarray analysis	121 of the 389 probes detected	Benschop et al., 2012
	Human	Blood, menstrual blood, saliva, semen, and vaginal secretion	Lasso regression analysis	26 taxa showed high predictive value for TsD	Salzmann et al., 2021
Other forensic application					
Drowning diagnosis					
	Rat	Skin, cardiac blood, lung, and liver	Unweighted UniFrac-based PCoA	PCo1 60.27% PCo2 19.15% (skin) PCo1 52.66% PCo2 15.98% (lung) PCo1 50.52% PCo2 18.37% (blood)	Wang et al., 2020
Post-mortem toxicology					
	Human	Nose, mouth, rectum, ears, eyes	Beta-dispersion	Cardiovascular disease and drug-related deaths correctly classified in 79%	Kaszubinski et al., 2020
Microorganism image analysis					
	Fungus		CNN	Accuracy 94.8%	Tahir et al., 2018
	Bacteria		CNN	Accuracy 96%	Treebupachatsakul and Poomrittigul, 2019
	Actino		CNN, ResNet	Accuracy 80.8% to 80.1%	Sajedi et al., 2019
	Diatoms		R-CNN, YOLO	F-measure of YOLO 84%	Bueno et al., 2018
	Diatoms		Inception V3	Identification rate 89.6%	Zhou et al., 2019
	Diatoms		YOLO, SegNet	Specificity, sensitivity, precision	Salido et al., 2020
	Diatoms		SegNet, Mask R-CNN	Precision, sensitivity, specificity	Ruis-Santaquiteria et al., 2020
	Diatoms		CNN	Validation set accuracy 97.33%	Zhou et al., 2019
	Cyanobacteria		Fast R-CNN	Average precision 0.929	Baek et al., 2020
				R^2 value of 0.775	
	Soybean cyst nematode eggs		CNN	Average accuracy 94.33%	Akintayo et al., 2016
	Tuberculosis		CNN	Precision 78.4%	Oomman et al., 2018
	Cell in blood		Faster R-CNN	Total accuracy 98%	Hung et al., 2017

addition, due to the complex influence of sample types, locations, environmental factors, and postmortem changes, combining the microbiome data from the experimental animals, human samples, and certain materials from the crime scene should be considered. Investigators should carry out multi-center cooperation.

3. Microbiome combined with other methods to solve forensic problems

The selection of differential microbes by bioinformatic technologies could better disclose microbial markers. AI could establish models for predicting forensic issues. Based on the specific microbes that could characterize the different body fluids and

environments, even individual identification, targeted PCR testing of the selected microorganisms could be explored to improve the efficiency and accuracy of forensic problems in future.

Author contributions

RZ and DG designed the manuscript and edited the manuscript. HY and ZiW wrote the manuscript. ZhW and FZ searched, edited, and reviewed the literature. All authors have read and commented on the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (Grant No: 81971793), the Natural Science Foundation of Liaoning Province (Grant No: 2022-YGJC-74), and the Development Program of China (Grant No: 2022YFC3302002).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., et al. (2015). Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Syst.* 1, 72–87. doi: 10.1016/j.cels.2015.01.001

Akintayo, A., Lee, N., Chawla, V., Mullaney, M., and Sarkar, S. (2016). An end-to-end convolutional selective autoencoder approach to Soybean Cyst Nematode eggs detection. *arXiv [Preprint]*. arXiv: 1603.07834. doi: 10.48550/arXiv.1603.07834

Aoyagi, M., Iwadate, K., Fukui, K., Abe, S., Sakai, K., Maebashi, K., et al. (2009). A novel method for the diagnosis of drowning by detection of *Aeromonas sobria* with PCR method. *Leg. Med.* 11, 257–259. doi: 10.1016/j.legalmed.2009.07.003

Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., et al. (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 17, 690–703. doi: 10.1016/j.chom.2015.04.004

Baek, S. S., Pyo, J. C., Pachepsky, Y., Park, Y., and Cho, K. H. (2020). Identification and enumeration of cyanobacteria species using a deep neural network. *Ecol. Indicat.* 115, 106395. doi: 10.1016/j.ecolind.2020.106395

Benschop, C. C. G., Quaak, F. C. A., Boon, M. E., Sijen, T., and Kuiper, I. (2012). Vaginal microbial flora analysis by next generation sequencing and microarrays; can microbes indicate vaginal origin in a forensic context? *Int. J. Legal Med.* 126, 303–310. doi: 10.1007/s00414-011-0660-8

Blaser, M. J., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Estrada, I., et al. (2013). Distinct cutaneous bacterial assemblages in a sampling of South American Amerindians and US residents. *ISME J.* 7, 85–95. doi: 10.1038/ismej.2012.81

Bouslimani, A., da Silva, R., Kosciolek, T., Janssen, S., Callewaert, C., Amir, A., et al. (2019). The impact of skin care products on skin chemistry and microbiome dynamics. *BMC Biol.* 17, 47. doi: 10.1186/s12915-019-0660-6

Bueno, G., Déniz, O., Ruiz-Santaquiteria, J., Olenici, A., and Borrego-Ramos, M. (2018). "Lights and pitfalls of convolutional neural networks for diatom identification," in *Optics, Photonics and Digital Technologies for Imaging Applications* (Washington, DC: SPIE). doi: 10.1117/12.2309488

Butzbach, D. M. (2010). The influence of putrefaction and sample storage on post-mortem toxicology results. *For. Sci. Med. Pathol.* 6, 35–45. doi: 10.1007/s12024-009-9130-8

Butzbach, D. M., Stockham, P. C., Kobus, H. J., Sims, D. N., Byard, R. W., Lokan, R. J., et al. (2013). Bacterial degradation of risperidone and paliperidone in decomposing blood. *J. For. Sci.* 58, 90–100. doi: 10.1111/j.1556-4029.2012.02280.x

Can, I., Javan, G. T., Pozhitkov, A. E., and Noble, P. A. (2014). Distinctive thanatomicrobiome signatures found in the blood and internal organs of humans. *J. Microbiol. Methods* 106, 1–7. doi: 10.1016/j.mimet.2014.07.026

Cao, J., Li, W. J., Wang, Y. F., An, G. S., Lu, X. J., Du, Q. X., et al. (2021). Estimating postmortem interval using intestinal microbiota diversity based on 16S rRNA high-throughput sequencing technology. *Fa Yi Xue Za Zhi* 37, 621–626. doi: 10.12116/j.issn.1004-5619.2020.400708

Cartozzo, C., Singh, B., Swall, J., and Simmons, T. (2021). Postmortem submersion interval (PMSI) estimation from the microbiome of sus scrofa bone in a freshwater lake. *J. For. Sci.* 66, 1334–1347. doi: 10.1111/1556-4029.14692

Cecchini, F., Iacumin, L., Fontanot, M., Comi, G., and Manzano, M. (2012). Identification of the unculturable bacteria *Candidatus arthromitus* in the intestinal content of trouts using dot blot and Southern blot techniques. *Vet. Microbiol.* 156, 389–394. doi: 10.1016/j.vetmic.2011.11.020

Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326, 1694–1697. doi: 10.1126/science.1177486

Dazzo, F. B., and Niccum, B. C. (2015). Use of CMEIAS image analysis software to accurately compute attributes of cell size, morphology, spatial aggregation and color segmentation that signify in situ ecophysiological adaptations in microbial biofilm communities. *Computation* 3, 72–98. doi: 10.3390/computation3010072

DeBruyn, J. M., and Hauther, K. A. (2017). Postmortem succession of gut microbial communities in deceased human subjects. *Peerj* 5, 3437. doi: 10.7717/peerj.3437

Escobar, J. S., Klotz, B., Valdes, B. E., and Agudelo, G. M. (2014). The gut microbiota of Colombians differs from that of Americans, Europeans and Asians. *BMC Microbiol.* 14, 311. doi: 10.1186/s12866-014-0311-6

Fierer, N., Lauber, C. L., Zhou, N., McDonald, D., Costello, E. K., and Knight, R. (2010). Forensic identification using skin bacterial communities. *Proc. Natl. Acad. Sci. U. S. A.* 107, 6477–6481. doi: 10.1073/pnas.1000162107

Forney, L. J., Zhou, X., and Brown, C. J. (2004). Molecular microbial ecology: land of the one-eyed king. *Curr. Opin. Microbiol.* 7, 210–220. doi: 10.1016/j.mib.2004.04.015

Franzén, O., Hu, J., Bao, X., Itzkowitz, S. H., Peter, I., and Bashir, A. (2015). Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. *Microbiome* 3, 43. doi: 10.1186/s40168-015-0105-6

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M., et al. (2015). Identifying personal microbiomes using metagenomic codes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E2930–E2938. doi: 10.1073/pnas.1423854112

Grad, Y. H., Lipsitch, M., Feldgarden, M., Arachchi, H. M., Cerqueira, G. C., FitzGerald, M., et al. (2012). Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3065–3070. doi: 10.1073/pnas.1121491109

Habtom, H., Pasternak, Z., Matan, O., Azulay, C., Gafny, R., and Jurkevitch, E. (2019). Applying microbial biogeography in soil forensics. *For. Sci. Int. Genet.* 38, 195–203. doi: 10.1016/j.fsigen.2018.11.010

Han, E., Kim, E., Hong, H., Jeong, S., Kim, J., In, S., et al. (2012). Evaluation of postmortem redistribution phenomena for commonly encountered drugs. *For. Sci. Int.* 219, 265–271. doi: 10.1016/j.forsciint.2012.01.016

Hauther, K. A., Cobaugh, K. L., Jantz, L. M., Sparer, T. E., and DeBruyn, J. M. (2015). Estimating time since death from postmortem human gut microbial communities. *J. For. Sci.* 60, 1234–1240. doi: 10.1111/1556-4029.12828

Huang, L. H., Xu, C. Q., Yang, W. X., and Yu, R. S. (2020). A machine learning framework to determine geolocations from metagenomic profiling. *Biol. Direct* 15, 278. doi: 10.1186/s13062-020-00278-z

Hung, J., Ravel, D., Lopes, S. C. P., Rangel, G., Nery, O. A., Malleret, B., et al. (2017). Applying faster R-CNN for object detection on malaria images. *Confer. Comput. Vis. Pattern Recogn. Workshops* 2017, 808–813. doi: 10.1109/CVPRW.2017.112

Huys, G., Coopman, V., Van Varenbergh, D., and Cordonnier, J. (2012). Selective culturing and genus-specific PCR detection for identification of Aeromonas in tissue samples to assist the medico-legal diagnosis of death by drowning. *For. Sci. Int.* 221, 11–15. doi: 10.1016/j.forsciint.2012.03.017

Javan, G. T., Finley, S. J., Can, I., Wilkinson, J. E., Hanson, J. D., and Tarone, A. M. (2016). Human thanatomicrobiome succession and time since death. *Sci. Rep.* 6, 598. doi: 10.1038/srep29598

Jesmok, E. M., Hopkins, J. M., and Foran, D. R. (2016). Next-generation sequencing of the bacterial 16S rRNA gene for forensic soil comparison: a feasibility study. *J. For. Sci.* 61, 607–617. doi: 10.1111/1556-4029.13049

Johnson, H. R., Trinidad, D. D., Guzman, S., Khan, Z., Parziale, J. V., DeBruyn, J. M., et al. (2016). A machine learning approach for using the postmortem skin microbiome to estimate the postmortem interval. *PLoS ONE* 11, 370. doi: 10.1371/journal.pone.0167370

Kakizaki, E., Kozawa, S., Matsuda, H., Muraoka, E., Uchiyama, T., Sakai, M., et al. (2010). Freshwater bacterioplankton cultured from liver, kidney and lungs of a decomposed cadaver retrieved from a sandy seashore: possibility of drowning in a river and then floating out to sea. *Leg. Med.* 12, 195–199. doi: 10.1016/j.legalmed.2010.03.008

Kaszubinski, S. F., Pechal, J. L., Smiles, K., Schmidt, C. J., Jordan, H. R., Meek, M. H., et al. (2020). Dysbiosis in the dead: human postmortem microbiome beta-dispersion as an indicator of manner and cause of death. *Front. Microbiol.* 11, 5347. doi: 10.3389/fmicb.2020.555347

Kaszubinski, S. F., Receveur, J. P., Nestle, E. D., Pechal, J. L., and Benbow, M. E. (2022). Microbial community succession of submerged bones in an aquatic habitat. *J. For. Sci.* 67, 1565–1578. doi: 10.1111/1556-4029.15036

Kersulyte, D., Kalia, A., Gilman, R. H., Mendez, M., Herrera, P., Cabrera, L., et al. (2010). Helicobacter pylori from peruvian amerindians: traces of human migrations in strains from remote Amazon, and genome sequence of an amerind strain. *PLoS ONE* 5, 5076. doi: 10.1371/journal.pone.0015076

Khan, H. A., Baig, F. K., and Mehboob, R. (2017). Nosocomial infections: epidemiology, prevention, control and surveillance. *Asian Pac. J. Trop. Biomed.* 7, 478–482. doi: 10.1016/j.apjtb.2017.01.019

Kimura, N., and Nobutada, S. (2006). Metagenomics: access to unculturable microbes in the environment. *Microb. Environ.* 21, 201–215. doi: 10.1264/jsme2.21.201

Klevens, R. M., Edwards, J. R., Richards, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., et al. (2007). Estimating health care-associated infections and deaths in US hospitals, 2002. *Public Health Rep.* 122, 160–166. doi: 10.1177/003335490712200205

Kong, H. H., and Segre, J. A. (2012). Skin microbiome: looking back to move forward. *J. Invest. Dermatol.* 132, 933–939. doi: 10.1038/jid.2011.417

Kuiper, I. (2016). Microbial forensics: next-generation sequencing as catalyst. *Embo Rep.* 17, 1085–1087. doi: 10.15252/embr.201642794

Langendijk, P. S. (1995). Quantitative fluorescence in situ hybridization of *Bifidobacterium* spp. with genus-specific 16S rRNA-targeted probes and its application in fecal samples. *Appl. Environ. Microbiol.* 61, 3069–3075. doi: 10.1128/aem.61.8.3069-3075.1995

Lauber, C. L., Zhou, N., Gordon, J. I., Knight, R., and Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiol. Lett.* 307, 80–86. doi: 10.1111/j.1574-6968.2010.01965.x

Lax, S., Hampton-Marcell, J. T., Gibbons, S. M., Colares, G. B., Smith, D., Eisen, J. A., et al. (2015). Forensic analysis of the microbiome of phones and shoes. *Microbiome* 3, 82. doi: 10.1186/s40168-015-0082-9

Lax, S., Smith, D. P., Hampton-Marcell, J., Owens, S. M., Handley, K. M., Scott, N. M., et al. (2014). Longitudinal analysis of microbial interaction between humans and the indoor environment. *Science* 345, 1048–1052. doi: 10.1126/science.1254529

Leake, S. L., Pagni, M., Falquet, L., Taroni, F., and Greub, G. (2016). The salivary microbiome for differentiating individuals: proof of principle. *Microb. Infect.* 18, 399–405. doi: 10.1016/j.micinf.2016.03.011

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436. doi: 10.1038/nature14539

Lee, D.-G., Yang, K. E., Hwang, J. W., Kang, H.-S., Lee, S.-Y., Choi, S., et al. (2016). Degradation of kidney and psoas muscle proteins as indicators of post-mortem interval in a rat model, with use of lateral flow technology. *PLoS ONE* 11, 557. doi: 10.1371/journal.pone.0160557

Lee, S. Y., Woo, S. K., Lee, S. M., Ha, E. J., Lim, K. H., Choi, K. H., et al. (2017). Microbiota composition and pulmonary surfactant protein expression as markers of death by drowning. *J. For. Sci.* 62, 1080–1088. doi: 10.1111/1556-4029.13347

Li, J., Quinque, D., Horz, H. P., Li, M. K., Rzhetskaya, M., Raff, J. A., et al. (2014). Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiol.* 14, 316. doi: 10.1186/s12866-014-0316-1

Liu, R., Gu, Y., Shen, M., Li, H., Zhang, K., Wang, Q., et al. (2020). Predicting postmortem interval based on microbial community sequences and machine learning algorithms. *Environ. Microbiol.* 22, 2273–2291. doi: 10.1111/1462-2920.15000

Liu, R., Wang, Q., Zhang, K., Wu, H., Wang, G., Cai, W., et al. (2021). Analysis of postmortem intestinal microbiota successional patterns with application in postmortem interval estimation. *Microb. Ecol.* 12, 1–16. doi: 10.21203/rs.3.rs-411962/v1

Lopez, C. D., Gonzalez, D. M., Haas, C., Vidaki, A., and Kayser, M. (2020). Microbiome-based body site of origin classification of forensically relevant blood traces. *For. Sci. Int. Genet.* 47, 2280. doi: 10.1016/j.fsigen.2020.102280

Lopez, C. D., Vidaki, A., Ralf, A., Gonzalez, D. M., Radjabzadeh, D., Kraaij, R., et al. (2019). Novel taxonomy-independent deep learning microbiome approach allows for accurate classification of different forensically relevant human epithelial materials. *For. Sci. Int. Genet.* 41, 72–82. doi: 10.1016/j.fsigen.2019.03.015

Lucci, A., Campobasso, C. P., Cirnelli, A., and Lorenzini, G. (2008). A promising microbiological test for the diagnosis of drowning. *For. Sci. Int.* 182, 20–26. doi: 10.1016/j.forsciint.2008.09.004

Lutz, H., Vangelatos, A., Gottel, N., Osculati, A., Visona, S., Finley, S. J., et al. (2020). Effects of extended postmortem interval on microbial communities in organs of the human cadaver. *Front. Microbiol.* 11, 569630. doi: 10.3389/fmicb.2020.569630

Ma, P., Li, C., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., et al. (2022). A state-of-the-art survey of object detection techniques in microorganism image analysis: from classical methods to deep learning approaches. *Artif. Intell. Rev.* 12, 1–72. doi: 10.1007/s10462-022-10209-1

Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T. Jr., Saxman, P. R., Stredwick, J. M., et al. (2000). The RDP (ribosomal database project) continues. *Nucl. Acids Res.* 28, 173–174. doi: 10.1093/nar/28.1.173

McNulty, S. L., Mole, B. M., Dailidiene, D., Segal, I., Ally, R., Mistry, R., et al. (2004). Novel 180- and 480-base-pair insertions in African and African–American strains of *Helicobacter pylori*. *J. Clin. Microbiol.* 42, 5658–5663. doi: 10.1128/JCM.42.12.5658-5663.2004

Metcalf, J. L., Parfrey, L. W., Gonzalez, A., Lauber, C. L., Knights, D., Ackermann, G., et al. (2013). A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. *Elife* 2, 1104. doi: 10.7554/eLife.01104

Metzker, M. L. (2010). Sequencing technologies: the next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626

Meyers, M. S., and Foran, D. R. (2008). Spatial and temporal influences on bacterial profiling of forensic soil samples. *J. For. Sci.* 53, 652–660. doi: 10.1111/j.1556-4029.2008.00728.x

Muyzer, G., Dewaal, E. C., and Uitterlinden, A. G. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59, 695–700. doi: 10.1128/aem.59.3.695-700.1993

Nagasawa, S., Motani-Saitoh, H., Inoue, H., and Iwase, H. (2013). Geographic diversity of Helicobacter pylori in cadavers: forensic estimation of geographical origin. *For. Sci. Int.* 229, 7–12. doi: 10.1016/j.forsciint.2013.02.028

Oliveira, M., and Amorim, A. (2018). Microbial forensics: new breakthroughs and future prospects. *Appl. Microbiol. Biotechnol.* 102, 10377–10391. doi: 10.1007/s00253-018-9414-6

Oomman, P. R., Kalmady, K. S., Jeny, R., and Sabu, M. K. (2018). Automatic detection of tuberculosis bacilli from microscopic sputum smear images using deep learning methods. *Biocybernet. Biomed. Eng.* 38, 691–699. doi: 10.1016/j.bbe.2018.05.007

Pechal, J. L., Crippen, T. L., Benbow, M. E., Tarone, A. M., Dowd, S., and Tomberlin, J. K. (2014). The potential use of bacterial community succession in forensics as described by high throughput metagenomic sequencing. *Int. J. Legal Med.* 128, 193–205. doi: 10.1007/s00414-013-0872-1

Rácz, E., Könczöl, F., Tóth, D., Patonai, Z., Porpáczy, Z., Kozma, Z., et al. (2016). PCR-based identification of drowning: four case reports. *Int. J. Legal Med.* 130, 1303–1307. doi: 10.1007/s00414-016-1359-7

Randall, S., Cartozzo, C., Simmons, T., Swall, J. L., and Singh, B. (2021). Prediction of minimum postmortem submersion interval (PMSImin) based on eukaryotic community succession on skeletal remains recovered from a lentic environment. *For. Sci. Int.* 323, 784. doi: 10.1016/j.forsciint.2021.110784

Roesch, L. F. W., Casella, G., Simell, O., Krischer, J., Wasserfall, C. H., Schatz, D., et al. (2009). Influence of fecal sample storage on bacterial community diversity. *Open Microbiol. J.* 3, 40–46. doi: 10.2174/1874285800903010040

Ross, A. A., Doxey, A. C., and Neufeld, J. D. (2017). The skin microbiome of cohabiting couples. *mSystems* 2, 43. doi: 10.1128/mSystems.00043-17

Ruis-Santaquiteria, J., Bueno, G., Deniz, O., Vallez, N., and Cristobal, G. (2020). Semantic versus instance segmentation in microscopic algae detection-ScienceDirect. *Eng. Appl. Artif. Intell.* 87, 103271. doi: 10.1016/j.engappai.2019.103271

Sajedi, H., Mohammadipanah, F., and Rahimi, S. (2019). Actinobacterial strains recognition by machine learning methods. *Multimedia Tools Appl.* 78, 20285–20307. doi: 10.1007/s11042-019-7379-9

Salido, J., Sánchez, C., Ruiz-Santaquiteria, J., Cristóbal, G., Blanco, S., and Bueno, G. (2020). A low-cost automated digital microscopy platform for automatic identification of diatoms. *Appl. Sci.* 10, 6033. doi: 10.3390/app10176033

Salzmann, A. P., Arora, N., Russo, G., Kreutzer, S., Snipen, L., and Haas, C. (2021). Assessing time dependent changes in microbial composition of biological crime scene traces using microbial RNA markers. *For. Sci. Int. Genet.* 53, 2537. doi: 10.1016/j.fsigen.2021.102537

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Sastre, C., Bartoli, C., Baillif-Couniou, V., Leonetti, G., and Pelissier-Alicot, A. L. (2017). Post mortem redistribution of drugs: current state of knowledge. *Curr. Pharm. Des.* 23, 5530–5541. doi: 10.2174/1381612823666170622111739

Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., et al. (2013). Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50. doi: 10.1038/nature11711

Schmedes, S. E., Woerner, A. E., and Budowle, B. (2017). Forensic human identification using skin microbiomes. *Appl. Environ. Microbiol.* 83, 1672. doi: 10.1128/AEM.01672-17

Scott, L., Finley, S. J., Watson, C., and Javan, G. T. (2020). Life and death: a systematic comparison of antemortem and postmortem gene expression. *Gene* 731, 144349. doi: 10.1016/j.gene.2020.144349

Skopp, G. (2010). Postmortem toxicology. *For. Sci. Med. Pathol.* 6, 314–325. doi: 10.1007/s12024-010-9150-4

Sodhi, K., Shrivastava, A., Arya, M., and Kumar, M. (2013). Knowledge of infection control practices among intensive care nurses in a tertiary care hospital. *J. Infect. Public Health* 6, 269–275. doi: 10.1016/j.jiph.2013.02.004

Suzuki, T. A., and Worobey, M. (2014). Geographical variation of human gut microbial composition. *Biol. Lett.* 10, 1037. doi: 10.1098/rsbl.2013.1037

Tahir, M. W., Zaidi, N. A., Rao, A. A., Blank, R., Vellekoop, M. J., and Lang, W. (2018). A fungus spores dataset and a convolutional neural network based approach for fungus detection. *IEEE Trans. Nanobiosci.* 17, 281–290. doi: 10.1109/TNB.2018.2839585

Tie, J., Uchigasaki, S., Haseba, T., Ohno, Y., Isahai, I., and Oshida, S. (2010). Direct and rapid PCR amplification using digested tissues for the diagnosis of drowning. *Electrophoresis* 31, 2411–2415. doi: 10.1002/elps.200900754

Treebupachatsakul, T., and Poomrittigul, S. (2019). "Bacteria classification using image processing and deep learning," in *Proceedings of the 2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)* (Jeju: IEEE). doi: 10.1109/ITC-CSCC.2019.8793320

Tuomisto, S., Karhunen, P. J., Vuento, R., Aittoniemi, J., and Pessi, T. (2013). Evaluation of postmortem bacterial migration using culturing and real-time quantitative PCR. *J. For. Sci.* 58, 910–916. doi: 10.1111/1556-4029.12124

Tyakht, A. V., Kostryukova, E. S., Popenko, A. S., Belenikin, M. S., Pavlenko, A. V., Larin, A. K., et al. (2013). Human gut microbiota community structures in urban and rural populations in Russia. *Nat. Commun.* 4, 3469. doi: 10.1038/ncomms3469

Uchiyama, T., Kakizaki, E., Kozawa, S., Nishida, S., Imamura, N., and Yukawa, N. (2012). A new molecular approach to help conclude drowning as a cause of death: simultaneous detection of eight bacterioplankton species using real-time PCR assays with TaqMan probes. *For. Sci. Int.* 222, 11–26. doi: 10.1016/j.forsciint.2012.04.029

Wang, L. L., Zhang, F. Y., Dong, W. W., Wang, C. L., Liang, X. Y., Suo, L. L., et al. (2020). A novel approach for the forensic diagnosis of drowning by microbiological analysis with next-generation sequencing and unweighted UniFrac-based PCoA. *Int. J. Legal Med.* 134, 2149–2159. doi: 10.1007/s00414-020-02358-1

Williams, D. W., and Gibson, G. (2017). Individualization of pubic hair bacterial communities and the effects of storage time and temperature. *For. Sci. Int. Genet.* 26, 12–20. doi: 10.1016/j.fsigen.2016.09.006

Willis, J. R., González-Torres, P., Pittis, A. A., Bejarano, L. A., Cozzuto, L., Andreu-Somavilla, N., et al. (2018). Citizen science charts two major "stomatotypes" in the oral microbiome of adolescents and reveals links with habits and drinking water composition. *Microbiome* 6, 218. doi: 10.1186/s40168-018-0592-3

Wu, G. D., Lewis, J. D., Hoffmann, C., Chen, Y.-Y., Knight, R., Bittinger, K., et al. (2010). Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags. *BMC Microbiol.* 10, 206. doi: 10.1186/1471-2180-10-206

Wu, J., Peters, B. A., Dominianni, C., Zhang, Y., Pei, Z., Yang, L., et al. (2016). Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* 10, 2435–2446. doi: 10.1038/ismej.2016.37

Yang, J. Y., Tsukimi, T., Yoshikawa, M., Suzuki, K., Takeda, T., Tomita, M., et al. (2019). Cutibacterium acnes (*Propionibacterium acnes*) 16S rRNA genotyping of microbial samples from possessions contributes to owner identification. *Msystems* 4, 594. doi: 10.1128/mSystems.00594-19

Yang, T., Shi, Y., Zhu, J., Zhao, C., Wang, J., Liu, Z., et al. (2021). The spatial variation of soil bacterial community assembly processes affects the accuracy of source tracking in ten major Chinese cities. *Sci. China Life Sci.* 64, 1546–1559. doi: 10.1007/s11427-020-1843-6

Yooseph, S., Andrews-Pfannkoch, C., Tenney, A., McQuaid, J., Williamson, S., Thiagarajan, M., et al. (2013). A metagenomic framework for the study of airborne microbial communities. *PLoS ONE* 8, 862. doi: 10.1371/journal.pone.0081862

Yu, Z., Xie, Q., Zhao, Y., Duan, L., Qiu, P., and Fan, H. (2021a). NGS plus bacterial culture: a more accurate method for diagnosing forensic-related nosocomial infections. *Leg. Med.* 52, 101910. doi: 10.1016/j.legalmed.2021.101910

Yu, Z., Xu, Q., Xiao, C., Li, H., Wu, W., Du, W., et al. (2021b). SYBR green real-time qPCR method: diagnose drowning more rapidly and accurately. *For. Sci. Int.* 321, 720. doi: 10.1016/j.forsciint.2021.110720

Zhang, J., Wang, M., Qi, X., Shi, L., Zhang, J., Zhang, X., et al. (2021). Predicting the postmortem interval of burial cadavers based on microbial community succession. *For. Sci. Int. Genet.* 52, 2488. doi: 10.1016/j.fsigen.2021.102488

Zhao, X., Zhong, Z., and Hua, Z. (2022). Estimation of the post-mortem interval by modelling the changes in oral bacterial diversity during decomposition. *J. Appl. Microbiol.* 133, 3451–3464. doi: 10.1111/jam.15771

Zhou, Y., Cao, Y., Huang, J., Deng, K., Ma, K., Zhang, T., et al. (2020). Research advances in forensic diatom testing. *For. Sci. Rese.* 5, 98–105. doi: 10.1080/20961790.2020.1718901

Zhou, Y., Zhang, J., Huang, J., Deng, K., and Huang, P. (2019). Digital whole-slide image analysis for automated diatom test in forensic cases of drowning using a convolutional neural network algorithm. *For. Sci. Int.* 302, 109922. doi: 10.1016/j.forsciint.2019.109922

Zou, Y., Zhuang, C., Fang, Q., and Li, F. (2020). Big data and artificial intelligence: new insight into the estimation of postmortem interval. *Fa yi xue za zhi* 36, 86–90. doi: 10.12116/j.issn.1004-5619.2020.01.017

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact