ORIGINAL RESEARCH (AR) DOI: 10.4274/jems.2024.09216

Journal of ETA Maritime Science 2024;12(4):446-465

# Close Contact Tracing and Risky Area Identification Using Alpha Shape Algorithm and Binary Contact Detection Model Based on Bluetooth 5.1

Qianfeng Lin[1], Jooyoung Son[2]

[1]National Korea Maritime and Ocean University, Department of Computer Engineering, Busan, South Korea
[2]National Korea Maritime and Ocean University, Department of Marine IT Engineering, Busan, South Korea

## Abstract

This study presents a novel approach to managing disease outbreaks on cruise ships by integrating Bluetooth 5.1 technology, a Binary Contact Detection Model, and an alpha shape algorithm. By harnessing the precise data capture capabilities of Bluetooth 5.1, this study accurately tracks interpersonal interactions and delineated high-risk areas, effectively enhancing close contact tracing and disease surveillance efforts. The Binary Contact Detection Model utilizes In-phase (I) and Quadrature (Q) data to identify close contacts with high accuracy, while the alpha shape algorithm helps in mapping out areas most susceptible to disease transmission. The combined use of these technologies represents a significant advancement in public health surveillance, offering a method to enhance safety and mitigate the spread of infections on cruise ships.

**Keywords:** Bluetooth 5.1 technology, Binary contact detection model, Alpha shape algorithm, Cruise ship disease management

## 1. Introduction

The motivation for this paper stems from the urgent need to enhance infectious disease management on cruise ships, as illustrated by recent outbreaks, such as the one at the Celebrity Summit.

Recent incidents, such as the outbreak onboard Celebrity Cruises' Celebrity Summit during its May 24, 2024, voyage, have highlighted the urgent need for effective disease management strategies on cruise ships. According to the Centers for Disease Control and Prevention, 68 out of the 2,264 passengers and five crew members contracted the virus, highlighting the complexity of controlling infectious diseases in such environments [1]. Cruise ships are unique environments that combine the characteristics of residential communities and transient hubs, presenting distinct challenges for disease control. The motivation for this study stems from the urgent need to enhance infectious disease management on cruise ships, as illustrated by recent outbreaks, such as the one at the Celebrity Summit.

The confined and highly interactive nature of cruise ships exacerbates the difficulty of traditional epidemiological tracking and containment. For instance, the mobility and interaction patterns of passengers and crew in shared spaces like dining rooms, pools, and theaters significantly increase the risk of widespread exposure [2]. In this environment, precise tools are required to monitor and control the spread of pathogens.

Effective disease management on cruise ships requires two critical strategies: close contact tracing and identification of high-risk areas. Close contact tracing involves accurately identifying individuals who have interacted with infected persons, which is essential for implementing effective quarantine measures and preventing further spread [3]. Additionally, identifying areas aboard ships that present higher transmission risks can guide targeted sanitization efforts and the implementation of specific restrictions to manage outbreaks efficiently [4].

In response to these challenges, this paper proposes an integrated technological approach that utilizes Bluetooth

**Address for Correspondence:** Jooyoung Son, National Korea Maritime and Ocean University, Department of Marine IT Engineering, Busan, South Korea
**E-mail:** mmlab@kmou.ac.kr
**ORCID iD:** orcid.org/0000-0001-7851-9214

**Received:** 18.08.2024
**Last Revision Received:** 15.10.2024
**Accepted:** 12.11.2024

**To cite this article:** Q. Lin, and J. Son, "Close Contact Tracing and Risky Area Identification Using Alpha Shape Algorithm and Binary Contact Detection Model Based on Bluetooth 5.1." *Journal of ETA Maritime Science*, vol. 12(4), pp. 446-465, 2024.

5.1 technology. The proposed technology enhances position tracking capabilities with high accuracy, which is crucial for the automatic logging of spatial interactions among passengers and crew [5]. The Binary Contact Detection Model proposed in this paper operates using In-phase and Quadrature (IQ) data from Bluetooth 5.1 signals to capture precise interaction data. This model is further complemented by the alpha shape algorithm, which facilitates the sophisticated delineation of high-risk areas on ships [6].

The combination of these technologies represents a significant advancement in public health surveillance on cruise ships. By employing these tools, we aim to not only rapidly identify potentially infected individuals and areas of high risk but also improve the precision and efficiency of public health interventions. This proactive approach to health management is vital for ensuring passenger safety, maintaining safe navigation, and enhancing the overall customer service experience on cruise ships. Moreover, it supports cruise companies by fostering a sense of corporate responsibility and enhancing consumer trust, which is crucial in today's competitive tourism market [7].

Finally, by implementing these strategies, cruise companies can not only effectively control and prevent the spread of disease but also foster a sense of corporate responsibility and enhance consumer trust, thus maintaining a competitive edge in the intense tourism market. This proactive approach to health management may also become a crucial factor for passengers when choosing a cruise line, especially in today's context of prominent global health issues [8].

Therefore, close contact tracing and risky area identification on cruise ships are not only practical necessities for addressing public health emergencies but are also integral parts of cruise operations, vital for ensuring public health, maintaining safe navigation, and enhancing customer service experiences [9].

Bluetooth 5.1 enhances position tracking capabilities with high accuracy, facilitating the automatic logging of spatial interactions among passengers and crew. Bluetooth 5.1 technology significantly enhances the use of IQ data in signal processing [10]. The IQ data are crucial for defining the amplitude and phase of the Bluetooth signal, which can be used to derive more precise information.

To effectively utilize Bluetooth 5.1 technology in close contact tracing, we propose the implementation of a Binary Contact Detection Model. The proposed model operates by utilizing the IQ data from the Bluetooth 5.1 signal, which is pivotal for capturing precise and granular interaction data. Alongside the Binary Contact Detection Model, integrating the alpha shape algorithm allows for sophisticated delineation of high-risk areas on the ship. This integrated approach not only facilitates the rapid identification of potentially infected individuals and high-risk areas and significantly improves the precision and efficiency of public health interventions. By implementing these advanced technological tools, we can better manage and contain outbreaks in complex environments like cruise ships, thereby ensuring passenger safety and public health. To address these pressing challenges, technological innovations that provide more accurate and automated solutions. This necessity forms the basis for the proposed Binary Contact Detection Model, which is designed to harness advanced Bluetooth 5.1 technology to realize more precise interaction tracking.

The primary contribution of this paper is the proposed Binary Contact Detection Model, which is designed to enhance disease management on cruise ships. This model is then innovatively integrated with Bluetooth 5.1 technology and the alpha shape algorithm, enabling precise tracking of interpersonal interactions and detailed mapping of high-risk areas. By leveraging Bluetooth 5.1's advanced signal processing capabilities, including IQ data, this approach offers a novel method for accurately identifying close contacts and potential transmission hotspots. Consequently, it provides a more effective strategy for outbreak containment and management onboard cruise ships, significantly improving public health surveillance.

The paper is structured to provide a comprehensive understanding of the study, and its findings as follows: Section 2, entitled "Previous Works", reviews the existing literature to highlight past methodologies and technologies used in tracking and managing disease outbreaks, setting the stage for the innovations introduced in this paper. Section 3, "Close Contact Tracing and Risky Area Identification", details the implementation of Bluetooth 5.1 technology integrated with the Binary Contact Detection Model and the alpha shape algorithm, describing how these tools are specifically applied to cruise ships. Section 4, "Performance Evaluation", presents an analysis of the effectiveness of the proposed models in real-world scenarios. Section 5, "Discussion", briefly addresses the implications of the results and identifies potential areas for improvement. Finally, Section 6 presents the conclusions, summarizing the key findings and suggesting directions for future research.

## 2. Previous Works

In ship environments, Bluetooth is favored over alternative wireless technologies for health monitoring and contact tracing because of its numerous benefits, such as low power usage, affordability, easy setup, spatial flexibility, precision, immediacy in data transmission, and robust privacy and security features. In particular, the Bluetooth Low Energy (BLE) variant is tailored for short-distance communications and consumes minimal power, enabling prolonged operation of devices without the need for frequent recharging [11]. These devices are cost-effective and can be extensively implemented throughout a cruise ship without significant upfront investment. Unlike more complex technologies like Wi-Fi, Bluetooth is easier to configure and maintain, which is essential for swift deployment and usability on cruise ships by staff without technical expertise [12].

Bluetooth technology enables direct device-to-device communication without relying on internet connectivity, thereby safeguarding data privacy and security-critical for handling sensitive personal health information. These attributes make Bluetooth the optimal choice for executing health surveillance and epidemic tracking onboard ships [13].

Bluetooth-enabled devices such as wristbands or badges are employed to track interaction patterns among passengers and crew, thereby facilitating the identification of close contacts. This technology's capability for quick and automated data collection diminishes the need for manual documentation and enhances the accuracy and efficiency of tracking efforts. Furthermore, Bluetooth can monitor and analyze movements and congregations on the ship, thereby identifying areas of high risk. When infections are identified, these data becomes essential for swift implementation of control measures like quarantines and sanitation procedures. Employing this technology not only aids in curbing the spread of infectious diseases but also bolsters the ability to manage public health crises, significantly supporting health and safety management on ships [14].

In ship environments, Bluetooth is chosen over other wireless technologies as the primary tool for health monitoring and close contact tracing primarily due to its advantages in low power consumption, cost-effectiveness, ease of deployment, spatial adaptability, accuracy, real-time capabilities, and privacy and security. Bluetooth, especially the BLE version, is designed for short-range communication and consumes very little power, allowing devices to operate for extended periods without frequent charging. Moreover, Bluetooth devices are generally low-cost and can be deployed on a large scale across a cruise ship without substantial investment. Compared to Wi-Fi, Bluetooth is simpler to configure and maintain, which is crucial for rapid deployment and operation on cruise ships, even by non-technical personnel [15]. Bluetooth technology allows direct communication between devices without the need for internet connectivity, enhancing the privacy and security of data, which is especially important for applications involving personal health data [16]. Collectively, these factors make Bluetooth an ideal choice for health surveillance and epidemic tracking in ship environments.

Research on Bluetooth technology in ship environments for health primarily involves real-time monitoring and data collection to support epidemic management and control measures. By using Bluetooth devices, such as wristbands or badges, the contact patterns of passengers and crew aboard a ship can be tracked, thereby aiding in the identification of close contacts [17]. The application of this technology allows for the rapid and automatic collection of contact data, thus reducing reliance on manual recording and improving the accuracy and efficiency of tracking. Additionally, Bluetooth technology can be used to monitor and analyze the movement and gathering of people on the ship, thereby identifying high-risk areas. When cases are detected, this information is crucial for quickly implementing control measures such as lockdowns and disinfection. The use of this technology not only helps control the spread of infectious diseases but also improves the capacity to handle public health emergencies, providing significant support for health and safety management on ships [18].

Studies on the spatial transmission of contagious diseases have extensively utilized Global Positioning System data from mobile devices to track human movement and identify infectious sites, proving critical in monitoring and predicting virus spread based on subjects' position histories during key infection periods [19]. Explores the use of BLE technology for close contact tracing, highlighting challenges due to varying signal strengths influenced by handset models, orientations, physical obstructions, and environmental reflections, which complicate accurate proximity detection necessary for effective coronavirus disease-2019 (COVID-19) contact tracing efforts [20]. Focuses on leveraging BLE technology for proximity detection, emphasizing the need for efficient, privacy-preserving methods and exploring various machine learning classifiers to enhance the accuracy and reliability of detecting high-risk contacts based on proximity data collected from smartphones [21]. Investigates smartphone-based applications utilizing both geolocation and Bluetooth technologies aimed at identifying and mitigating the spread of COVID-19, raising concerns regarding their practical implementation in densely populated areas and potential privacy issues associated with the tracking technologies [22]. Explores Bluetooth-based and decentralized approaches to balance effectiveness with privacy concerns, focusing on minimizing reliance on central authorities while addressing device compatibility and the operational complexities introduced by decentralized models [23]. Primarily focused on using BLE technology to estimate proximity through signal attenuation, with many studies highlighting its limitations under various environmental conditions and proposing enhancements involving machine learning models to leverage additional signal features and contextual data from smartphones to improve accuracy in both indoor and outdoor settings [24]. Utilized device-to-device interactions via technologies like Bluetooth, facing challenges such as low interoperability, modest user adoption, and privacy concerns [25]. There is an increasing focus on enhancing these systems through epidemiological modeling and integration of environmental data to improve effectiveness and adoption rates.

Previous studies have laid a solid foundation in the use of technology for disease surveillance; however, they also reveal significant gaps in current methodologies, particularly in environments as complex as cruise ships. We provide the Binary Contact Detection Model. The integration of the alpha shape algorithm with the Binary Contact Detection Model represents an evolution in the field, aimed at overcoming this limitation.

Bluetooth 5.1 technology has been increasingly adopted in consumer electronic products, especially smartphones and wearable devices. However, considering the diversity of devices carried by cruise ship passengers, our system design does not require passengers to purchase Bluetooth 5.1 devices. Instead, the system simply needs to attach low-cost tags to the smartphones of passengers to enable positioning functionality. This approach not only reduces the barrier for passengers but also ensures broad device compatibility.

Implementing Bluetooth 5.1 AoA positioning technology involves certain costs, primarily based on the procurement of locators and tags. The cost of locators is relatively high; however, the cost of tags is relatively low. In a ship environment, cruise companies can significantly reduce overall costs by purchasing bulk tags. Additionally, the lightweight and easy-to-distribute nature of the tags simplifies logistics management, allowing cruise companies to distribute tags uniformly to passengers upon boarding, thereby ensuring efficient system operation.

The placement of antenna arrays plays a crucial role in the positioning accuracy of Bluetooth 5.1 AoA. However, antenna array optimization typically requires hardware adjustments and upgrades. Although optimizing the locator deployment positions can improve the positioning accuracy, the need to frequently adjust the locator positions in the highly dynamic environment of a ship presents certain operational challenges and limitations.

Although Bluetooth 5.3 introduces several enhancements, such as improved power efficiency, faster data transmission, and better connection stability, it does not offer specific improvements in indoor positioning accuracy compared to Bluetooth 5.1. The key feature for high-precision indoor positioning, namely the AoA functionality, is introduced in Bluetooth 5.1 and remains the core technology for accurate position tracking.

To address this issue, we adopt machine learning algorithms, such as the Random Forest (RF) algorithm, to mitigate the impact of environmental changes on positioning accuracy. Specifically, by filtering out noise features before model training, we can enhance the quality of the input data to ensure that the model can maintain high positioning accuracy in dynamic environments. The proposed method reduces the reliance on frequent hardware adjustments, thereby allowing the system to more flexibly adapt to environmental changes while still providing reliable positioning results.

## 3. Close Contact Tracing and Risky Area Identification

In this section, we present our comprehensive approach for tracing close contacts and identifying risky areas on cruise ships. Given the unique challenges posed by the cruise ship environment-such as high-density populations, enclosed spaces, and complex interpersonal interactions-it is essential to employ advanced technologies and models to effectively monitor and control the spread of infectious diseases. Our method integrates Bluetooth 5.1 technology for precise in-ship positioning, a Binary Contact Detection Model for accurate identification of close contacts, and an alpha shape algorithm to delineate high-risk areas based on the spatial distribution of individuals. The following subsections describe each component of the proposed approach.

### 3.1. Problem Definition

The rapid spread of infectious diseases in cruise ship environments poses a significant public health challenge. As a closed and densely populated setting, a cruise ship's unique environmental conditions-such as high-density interpersonal interactions and frequent social activities-greatly increase the risk of disease transmission. Traditional disease monitoring and control methods often fail to address the rapid spread of diseases within high-density, enclosed spaces, making efficient tracking and control difficult. Specifically, the critical issue that needs to be addressed is how to effectively and accurately track close contacts and identify high-risk transmission areas on a cruise ship. In the cruise ship environment, traditional contact tracing methods primarily rely on manual records or low-precision technological means, which are insufficient for handling the complex and variable patterns of interpersonal interactions inherent to cruise ships. Additionally, the enclosed environment facilitates the widespread dissemination of pathogens, further complicating disease control efforts.

To address these challenges, this study proposes an innovative approach that integrates Bluetooth 5.1 technology, a Binary Contact Detection Model, and the alpha shape algorithm to enhance disease monitoring and response capabilities on cruise ships. Specifically, the proposed method leverages the high-precision data capture capabilities of Bluetooth 5.1 in combination with a Binary Contact Detection Model to achieve accurate tracking of interactions between individuals. Concurrently, the alpha shape algorithm helps identify areas most susceptible to disease transmission, providing an effective means to quantify the risk of infectious disease spread on cruise ships.

Through this comprehensive technological approach, our research aims to play a crucial role in disease prevention and control by slowing the rate of virus transmission and containing the spread within defined limits, thereby ensuring the health and safety of cruise passengers and staff. This not only offers an advanced disease monitoring solution for the cruise industry but also provides valuable insights for controlling infectious diseases in other high-density, enclosed environments.

### 3.2. In-Ship Positioning Using Bluetooth 5.1

Bluetooth 5.1's AoA indoor positioning technology is an advanced wireless positioning method that primarily determines a device's position by measuring the angles of incoming wireless signals [26]. In an AoA positioning system, a positioning tag (e.g., a device equipped with a Bluetooth transmitter) emits signals at specific frequencies. The signals are captured by a receiver equipped with multiple antennas. The receiver utilizes its antenna array to measure the angles of arrival of the signals, which includes both the azimuth angle (the angle on the horizontal plane) and the elevation angle (the angle on the vertical plane) [27]. This measurement typically involves advanced signal processing techniques, such as phase-difference measurements, where the receiver calculates the differences in signal arrival times across various antennas to infer the angles of arrival. By integrating multiple angle measurements, the system can accurately calculate the tag position in 3D space.

The internal environment of ships is complex, with many metallic structures and devices that can affect the propagation of wireless signals. Bluetooth 5.1 technology can provide more accurate positioning in such environments because it does not rely on signal strength. Known for its low energy consumption, Bluetooth technology is well-suited for long-term operation in environments with limited power sources, which is a significant advantage for devices on ships. Compared to other advanced positioning technologies like Wi-Fi or Ultra-Wideband, Bluetooth devices generally have lower costs and are easier to deploy on a large scale. Bluetooth technology supports a wide range of devices and applications, is easy to integrate into existing ship management systems, and can support future expansions and upgrades.

Figure 1's left side shows a direction-finding antenna array labeled with specific antenna numbers from 1 to 16. This array is used to capture signals sent from various angles. Each antenna's position and angle are designed to maximize the capability to receive signals from different directions. When a tag with a Bluetooth transmitter emits a signal, the antenna receives the signal and calculates the angle of arrival based on the differences in the signal arrival times. The middle panel of Figure 1 displays a three-dimensional view of the positioning principle.

Figure 1 illustrates the AoA positioning technique used in Bluetooth 5.1 for indoor positioning. The left schematic in Figure 1 shows the direction-finding antenna array used to detect the AoA of the Bluetooth signal. The schematic on the right of Figure 1 shows how the system determines the tag's position in a three-dimensional space (X, Y, Z planes) using the elevation and azimuthal angles. By measuring the angle between the Bluetooth tag and multiple locators, the system can calculate the tag's precise position. The locators are placed at different positions, which allows the system to triangulate the tag's position using these angle measurements. Figure 1 highlights the tag's positioning in relation to the locators and explains how the azimuthal and elevation angles contribute to determining the position of the tag within the indoor environment.

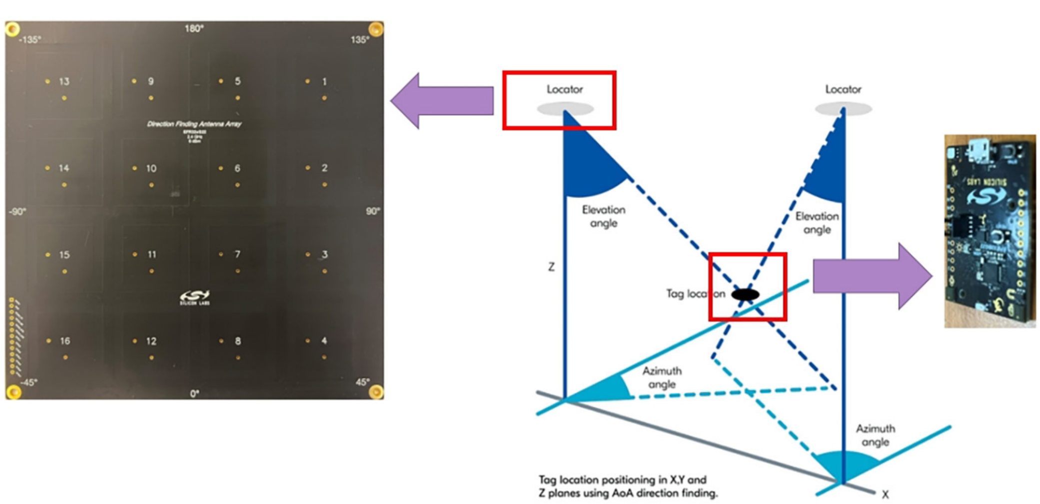


***Figure 1.*** *Illustration of how Bluetooth 5.1 uses the angle of arrival technique for precise indoor positioning by measuring the arrival angles of signals at a receiver antenna array*

The hardware shown on the right side of Figure 1 is actually a tag equipped with a Bluetooth chip, which is used to emit signals containing specific frequencies and coding information. This transmitter is likely designed to have low power consumption and is therefore suitable for applications requiring long run times. Its main function is to continuously emit signals that are received by the direction-finding antenna array. With the combination of precise hardware and complex signal processing algorithms, high-precision position information can be obtained in complex environments. This technology is particularly suitable for scenarios requiring precise positioning, such as on ships, in large factories, or in other environments with complex physical structures. With Bluetooth 5.1 indoor positioning technology, devices can achieve efficient spatial position monitoring and management with low energy consumption [28].

### 3.3. Methods to Obtain IQ Data

Constant Tone Extension (CTE) is a technology used in Bluetooth signal transmission, specifically designed to support AoA indoor positioning functions [29]. Essentially, CTE is a continuous single-frequency tone appended at the end of a data packet. The primary function of the proposed system is to provide a stable reference point, enabling receiving devices to precisely measure the signal angle of arrival by analyzing its phase information [30]. A tag emits a Bluetooth signal containing CTE, which, following conventional data, such as device ID and other communication information, includes the additional CTE portion, as shown in Figure 2. Locators equipped with multiple antennas receive the transmitted signal. Each antenna records the signal arrival time and phase information upon receipt. The processing system within the locator calculates the phase differences of the signals received by each antenna.

Because CTE is a continuous single-frequency tone, it makes the phase information very clear and stable, thereby facilitating high-precision measurements. By analyzing the phase differences received from different antennas, the locator can determine the signal's angle of arrival, including both azimuth and elevation angles. These angles are determined based on the relative positions of the signal to each antenna. By combining the angle information provided by more locators, the system uses geometric triangulation to calculate the exact position of the tag. If the locators are fixed and their positions are known, then the exact coordinates of the tag in 3D space can be accurately determined.

Through this method, CTE enables Bluetooth technology to achieve precision positioning similar to that of radar and sonar. This high-precision angle measurement capability is particularly suitable for complex environments such as ships and large factories. Therefore, CTE not only improves the accuracy of Bluetooth positioning technology and expands its application scenarios, making it a versatile and efficient positioning tool.

In Bluetooth technology, the relationship between IQ data and CTE is central to the implementation of high-precision positioning technologies, such as AoA positioning. IQ data represent the I and Q components of a signal. These components are used to describe the signal's amplitude and phase and are fundamental to signal analysis in wireless communication. Using these data, the signal's waveform can be reconstructed and its transmission characteristics analyzed. Thus, IQ data form the basis for implementing various signal processing techniques, including frequency modulation, phase modulation, and other complex modulation-demodulation methods [31].

The antenna arrays capture CTE signals, recording the I and Q components of the signal received by each antenna.

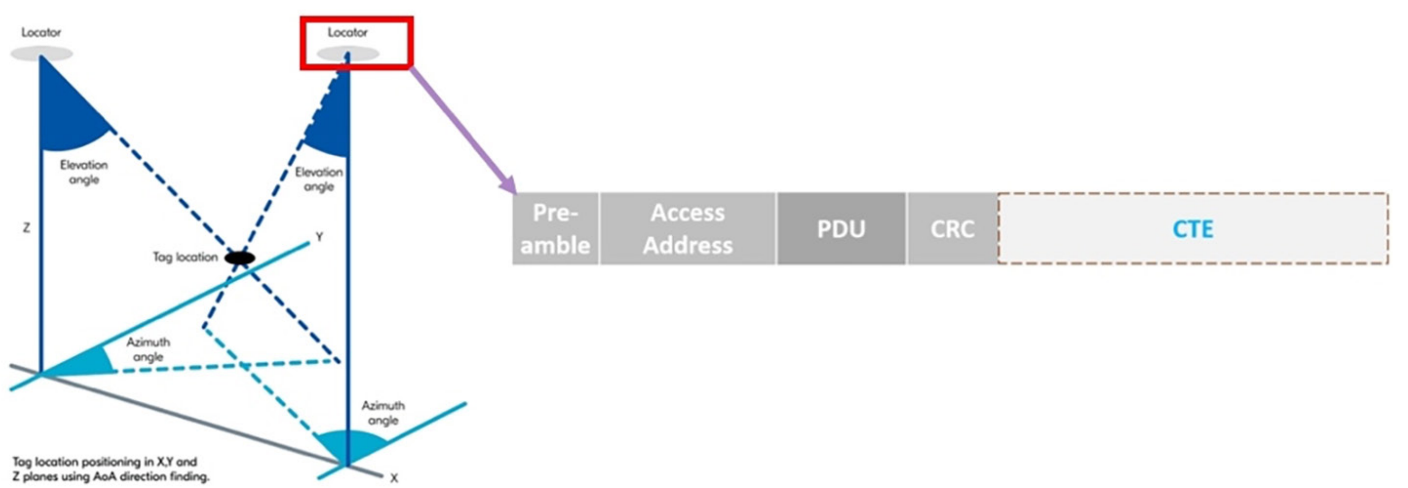


***Figure 2.*** *IQ sampling upon receiving CTE packets*

*IQ: In-phase and Quadrature, CTE: Constant Tone Extension*

Analysis software or hardware devices use the collected IQ data to calculate the signal's phase and amplitude [32]. For AoA positioning, phase differences are particularly important because they can reveal the angle of the transmitter's position relative to the receiver. Because CTE provides a continuous and stable signal, the receiver can perform multiple measurements, thereby enhancing the angle estimation accuracy. Accurate capture of IQ data is crucial to this process. In real environments, multipath effects and environmental noise may interfere with the accuracy of IQ data; thus, high-quality signal processing algorithms and hardware are necessary to ensure correct signal interpretation and precise calculation of phase differences. The combination of IQ data and CTE achieves high-precision three-dimensional positioning in Bluetooth technology, providing an effective solution for tracking and locating devices in complex environments.

The right panel of Figure 3 illustrates the data points of the I and Q components changing over time. The blue curve represents the I data, indicating changes in the I component over time, while the orange curve represents the Q data, indicating changes in the quadrature component. The fluctuations of these two curves depict the dynamic characteristics of the signal, with each point's vertical position representing the amplitude of the component at specific moments in time.

### 3.4. Pairwise Combined IQ Data

The Binary Contact Detection Model identifies close contacts using Pairwise Combined IQ Data. The proposed method combines the identification results from the Binary Contact Detection Model and the corresponding positions with the alpha shape algorithm to identify risky areas. Positions are precisely determined by affixing BLE tags to user smartphones and leveraging Bluetooth 5.1 technology for data acquisition. The core concept of directly converting IQ data into features acceptable by a model lies in fully utilizing the rich information contained within the IQ data to simplify the data processing workflow and enhance model performance. IQ data can comprehensively describe the amplitude and phase information of signals, which is crucial for many applications. By directly using these raw IQ data as input features for models, more useful information can be retained, thereby avoiding information loss during the data preprocessing stage and simplifying the complexity of data processing. Directly using IQ data as features can also improve the predictive performance of the model, as these data contain complete signal information that helps in more accurately capturing signal characteristics.

Figure 4 illustrates the entire process of transmitting IQ data from a BLE Tag to a Locator via Bluetooth 5.1 and eventually

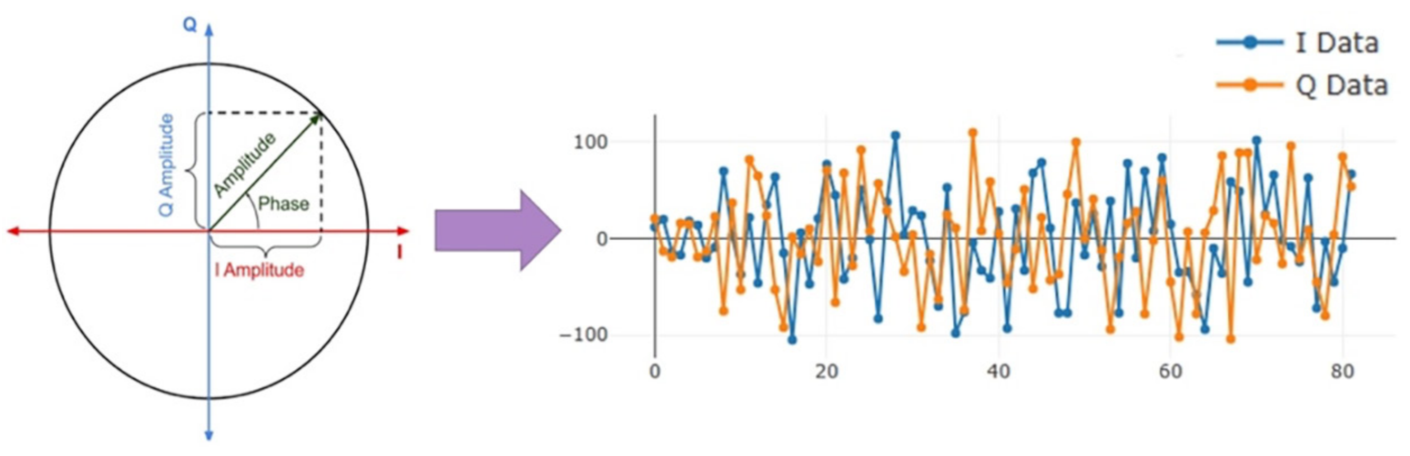


***Figure 3.*** *Visualization of IQ data*

*IQ: In-phase and Quadrature*

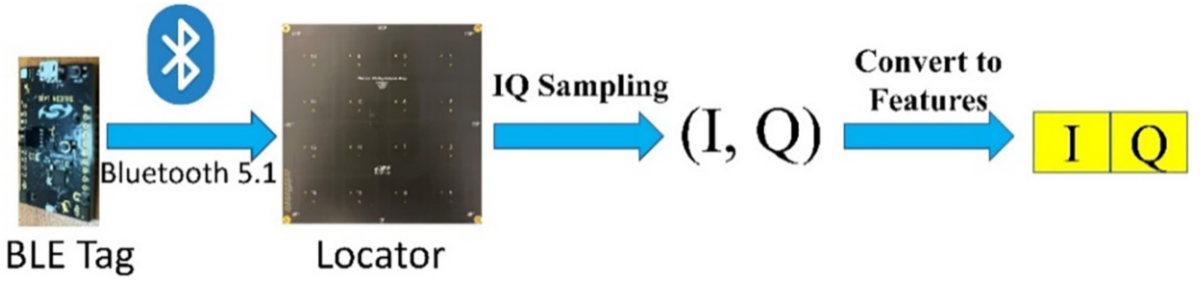


***Figure 4.*** *The IQ data processing flow from BLE tag to model-ready features*

*IQ: In-phase and Quadrature, BLE: Bluetooth Low Energy* converting it into features acceptable to the model. The icon on the left represents the BLE Tag, which transmits data via Bluetooth 5.1. The icon in the middle panel of Figure 4 represents the locater. The locater receives the signals transmitted from the BLE Tag and performs IQ sampling. IQ sampling is achieved by extracting the I and Q components from the received signal. The arrow labeled "IQ Sampling" in Figure 4 illustrates this process, with the (I, Q) symbols next to it representing the collected IQ data.

Subsequently, these IQ data are converted into features that can be directly used as model input. The arrow labeled "Convert to Features" in Figure 4 indicates this conversion process. The converted feature data are shown in the yellow box on the right (labeled I and Q, indicating the directly used feature data. The key to this step is that it does not require complex feature extraction methods; it simply involves using raw IQ data directly as model input, which simplifies the data processing workflow.

Figure 4 shows the process of transmitting signals from the Tag device via Bluetooth 5.1, performing IQ sampling with the locater, and eventually converting these IQ data into features for direct use in model input. This approach not only simplifies the data processing workflow but also retains crucial signal information, potentially enhancing model performance and enabling functionalities such as localization and signal classification.

IQ data between close contacts may be similar or exhibit certain patterns. Therefore, assuming that we already know that an individual is a COVID-19 patient or has close contact with other individuals, we can combine the IQ data of other individuals with the IQ data of the known individual in pairs to provide input features to the model. If the model outputs a result of 1, then the individual and the known close contact are also closely connected. Until the identification results of all individuals are obtained, all individuals are considered potential close contacts.

Specifically, the first step is to identify and confirm COVID-19 patients or known close contacts whose IQ data served as the reference baseline. Then, the IQ data of other individuals are combined with the IQ data of a known individual in pairs to form new feature pairs. These paired IQ data can be used as input features for the machine learning model, which makes predictions based on them. If the model outputs a result of 1, it indicates that the individual has close contact with a known close contact; if the result is 0, it indicates no close contact. Based on the model prediction results, new close contacts can be identified and marked. Until the identification results of all individuals are obtained, all individuals are considered potential close contacts.

During data collection and preprocessing, IQ data can be normalized or standardized to reduce noise and interference effects on the model, thereby improving the prediction accuracy and stability of the model. Using the data of known COVID-19 patients and close contacts, the machine learning model can be trained to recognize patterns in the IQ data. An RF classifier can be used to train and optimize the model. After training the model, it can be tested using a validation dataset to evaluate its accuracy. By continuously adjusting the model parameters and improving the algorithms, the model's performance can be enhanced.

Deploy the trained model into practical applications and monitor its prediction results in real time to ensure that it accurately identifies potential close contacts. Continuously collect new IQ data to update and optimize the model, and address any changes or challenges that may arise. Through these steps and methods, an efficient COVID-19 close-contact identification system can be established to help promptly detect and isolate potential infected individuals, effectively controlling the spread of the pandemic. This approach not only simplifies the data processing workflow but also improves identification accuracy and efficiency, thus contributing to public health management and pandemic prevention.

Figure 5 illustrates the concept of using the IQ data of already labeled close contacts to identify potential close contacts. At the top of Figure 5, the IQ data of close contacts (represented by blue dots) are displayed. The middle panel of Figure 5 presents two sets of IQ data: the upper panel presents the IQ data of known close contacts (blue dots), and the lower panel presents the IQ data of potential close contacts (red dots). The bottom section of Figure 5 provides a solution, indicating that the IQ data of known and potential close contacts should be considered as one group and input into a classifier for identification. This process is further emphasized by a yellow arrow, explaining the specific operation of combining the two sets of IQ data and inputting them into the model for classification. To expand on this, the workflow depicted in Figure 5 involves several key steps. First, the IQ data from known close contacts, represented by blue dots, are used as baseline data. Then, the IQ data of potential close contacts (red dots) are paired with these known data to form new data pairs. Next, these paired data are input into a classification model for training and prediction. By learning the features within these IQ data pairs, the model can identify which individuals are likely to be new, close contacts.

The underlying logic of the proposed method is that the IQ data of close contacts may exhibit similar feature patterns. By using a machine learning model, these feature patterns can be recognized automatically, thereby allowing accurate identification of potential close contacts.

### 3.5. Binary Contact Detection Model

The sigmoid function is a commonly used activation function in machine learning and neural networks. The characteristics of the sigmoid function make it particularly suitable for binary classification problems, where it can map input real numbers to a range of (0,1), representing the probability of a particular class. For example, in binary classification models, the output processed by the sigmoid function can be interpreted as the probability that a sample belongs to a certain class. However, in some practical applications, the standard sigmoid function may not sufficiently meet the requirements; thus, we propose the Shifted-Scaled Sigmoid Function.

The proposed Shifted-Scaled Sigmoid Function is proposed because the standard sigmoid function may have limitations in some application scenarios and may not fully meet the actual requirements. The standard sigmoid function is centered at 0, which means that its symmetry is centered around 0. The standard sigmoid function has an output range of (0,1) with a smooth transition at 0.5, which is an ideal threshold. However, in many cases, the distribution of input data is not symmetric, or the characteristics of the data are not suitable for being centered around 0.

For example, in some detection and classification tasks, the mean or median of the input data may be offset from 0. In such cases, using the standard sigmoid function would lead to biased classification results and would not accurately reflect the actual situation. Therefore, adjusting the position of the sigmoid function so that its symmetry is centered around the median of the data can improve the model's prediction accuracy. Furthermore, the steepness of the standard sigmoid function is fixed; thus, it may not provide sufficient discriminative power for tasks requiring more sensitive threshold transitions. In these tasks, we want the function to have a more rapid transition near a specific point, which allows the model to make clearer classifications for data points close to the threshold. This requires adjusting the steepness of the function to accommodate different application requirements. Table 1 provides a detailed description of the symbols and notations used throughout this paper.

In Equation (1), we define "prediction$_i$" as the models predicted output for the *i*-th sample, which is expressed as

$$prediction_i = f(x_i) \tag{1}$$

where *f* represents the model's prediction function, and *x i* is the input feature vector of the *i*-th sample.

By employing the Shifted-Scaled Sigmoid Function, we can shift the center of the sigmoid function from 0 to other specific values in the data, thereby making the classification threshold more reasonable and accurate. Before introducing the Shifted-Scaled Sigmoid Function, we first need to understand two important parameters: *k* and *x* (Equation 2).

$$x = \frac{\sum_{i=1}^{k} prediction_i}{k} \tag{2}$$

where *k* is the amount of input data. *x* is the cumulative average of the model's output values. After introducing *k* and *x*, we introduce the Shifted-Scaled Sigmoid Function. The Shift-Scaled Sigmoid Function is formulated as Equation (3).

$$Shifted - Scaled\ Sigmoid\ Function = \frac{1}{1+e^{-k(x-0.5)}} \tag{3}$$

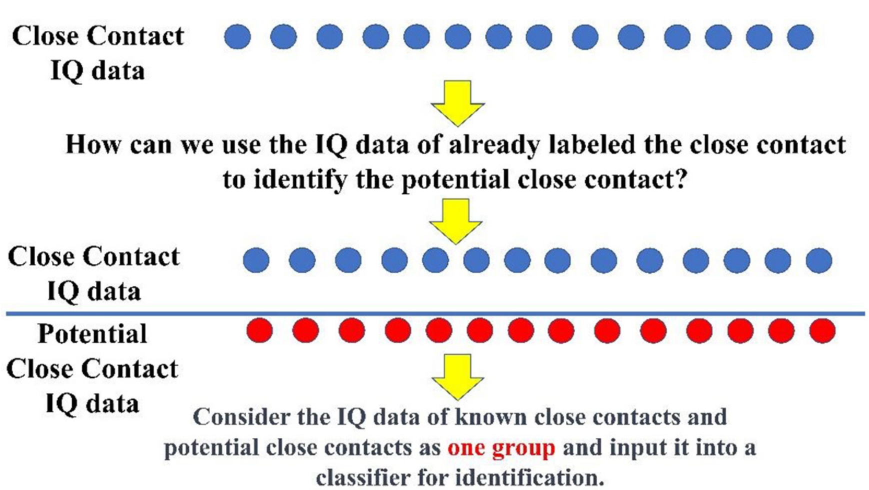


***Figure 5.*** *Using IQ data on labeled close contacts to identify close contacts*

*IQ: In-phase and Quadrature*

where the shift parameter 0.5 moves the center from 0 to 0.5, thereby making the function symmetrical around 0.5. The choice of 0.5 as the shift parameter is intended to align the decision threshold of the sigmoid function with the standard for binary classification problems, thereby enhancing classification accuracy and interpretability while ensuring model consistency and natural symmetry. The scaling parameter $k$ controls the function steepness. Larger $k$ values make the function steeper at approximately 0.5, which makes the distinction between 0 and 1 in classification tasks more clear. Therefore, if the output value is greater than 0.5, it is classified as 1; otherwise, it is classified as 0.

For example, the distribution of Pairwise Combined IQ Data may be asymmetric, and the feature center of the data may deviate from 0. By using the Shifted-Scaled Sigmoid Function, we can ensure that the model's classification threshold is closer to the actual data feature center, thereby avoiding classification errors caused by the standard sigmoid function's center deviation. At the same time, by adjusting the scaling parameter $k$, the model can more flexibly adapt to different data distributions, which improves the accuracy and stability of the classification results.

In summary, the shift-scaled sigmoid function overcomes the limitations of the standard sigmoid function in terms of center position and steepness by shifting and scaling it, making it better suited for various practical applications. By adjusting the values of $x$ and $k$, the model's classification accuracy and stability can be improved, making it suitable for tasks that require flexible adjustment of classification thresholds and rapid response to changes.

Cruise ships are constructed with extensive metal structures, including hulls, decks, and bulkheads, which can cause significant signal reflection, absorption, and multipath propagation. These metal surfaces can distort Bluetooth signals, leading to errors in the angle estimation and reduced reliability of AoA measurements. Additionally, the complex layout of ships, with numerous confined spaces and obstacles such as walls, furniture, and equipment, contributes to signal obstructions. The dynamic movement of passengers and crew members adds another layer of complexity, as human bodies can absorb and reflect signals, causing fluctuations in signal strength and quality.

To mitigate these environmental factors and enhance the accuracy of Bluetooth 5.1 AoA technology on cruise ships, this study employs a method that involves optimizing the features of the input signal. Signal data are used as features and input to machine learning algorithms, such as the Light Gradient Boosting Machine (LightGBM), to perform feature filtering. This process forms a feature filter that removes noisy features, effectively addressing the inherent interference problem. By filtering out these noisy features-which represent manifestations of environmental interference-the overall impact of interference in the ship environment can be mitigated. This optimization enhances the quality of the data used for angle estimation, leading to more reliable AoA measurements and improved positioning accuracy. Moreover, our approach does not require the assistance of the Wireless Fidelity (WiFi) and the Radio Frequency Identification (RFID) to improve accuracy, thereby avoiding increased hardware costs.

The Binary Contact Detection Model is a machine learning model designed to identify close contacts. Its core idea is to use Pairwise Combined IQ Data as features, undergo a series of feature selection and classification processing

***Table 1.*** *Notations and symbol descriptions*

| Symbol | Description |
|---|---|
| $f(x_i)$ | The function represents the model's prediction function, and is the input feature vector of the $i$-th sample |
| $\text{prediction}_i$ | The models predicted output for the $i$-th sample |
| $k$ | Amount of input data used to calculate the average prediction |
| $x$ | Cumulative average of the model's output values |
| $h(p)$ | Prediction function for close contact between devices |
| $e_i$ | The positional coordinate of the $i$-th device pair is predicted as "1" (close contact) |
| $E$ | Set of positional coordinates of all device pairs predicted as "1" |
| $\sigma$ | A simplex with a circumcircle radius in the Delaunay triangulation |
| $Del(E)$ | Delaunay triangulation of position set $E$, a geometric structure that connects points to form triangles where no point is inside the circumcircle of any triangle |
| $r(\sigma)$ | The radius of the circumcircle |
| Alpha | The parameter that controls the tightness of the shape |
| $n$ | Number of positions in the input set $E$ |
| $\gamma$ | Adjustment factor for the alpha value (typically set to 1) |

steps, and finally use the Shifted-Scaled Sigmoid Function to output the result, determining whether an individual is in close contact. The specific process is as follows: first, the IQ data are paired to form Pairwise Combined IQ Data, which generates the initial feature set. Then, through Controllable Feature Selection (CFS), including feature importance calculation using LightGBM, sorting features by importance in descending order, forward feature selection, and RF classifier validation, the optimal feature set is finally output. The optimal feature set forms a feature filter that screens the most important features. The features processed by the Feature Filter are then input into a trained RF classifier for initial classification, and the final classification result is input into the Shift-Scaled Sigmoid Function. If the output is 1, the individual is considered to have close contact; if the output is 0, the individual is considered to have no close contact.

Figure 6 illustrates the overall process of the Binary Contact Detection Model, including two main stages: the offline and online stages, used to identify close contacts. In the offline stage, IQ data are paired to form Pairwise Combined IQ Data, generating the initial feature set. Next, CFS is performed, which includes feature importance calculation, sorting features by importance in descending order, forward feature selection, and RF classifier validation. These steps ultimately output the optimal feature set.

In the online stage, the input Pairwise Combined IQ Data is represented by green dots. First, this data is screened through a feature filter formed using the optimal feature set obtained in the offline stage to filter out important features. Then, the features processed by the Feature Filter are input to a trained RF classifier for initial classification. The classification result is then processed by the Shift-Scaled Sigmoid Function, which ultimately outputs 0 or 1. If the result is 1, the individual is judged to have close contact; if the result is 0, the individual is judged to have no close contact.

Through this process, the Binary Contact Detection Model can effectively identify close contacts, improving the model's classification accuracy and stability, making it suitable for public health management and epidemic prevention.

### 3.6. System Model

In this study, a risky area is defined as the convex hull polygon formed by a set of positions of close contacts. Therefore, we propose an innovative model that combines an alpha shape algorithm and the output of a Binary Contact Detection Model to precisely identify and mark risky areas. This method is particularly suitable for disease transmission and public health management, especially in environments where rapid identification and response to infectious disease outbreaks. The Binary Contact Detection Model, which is based on IQ data received from the locater, predicts whether

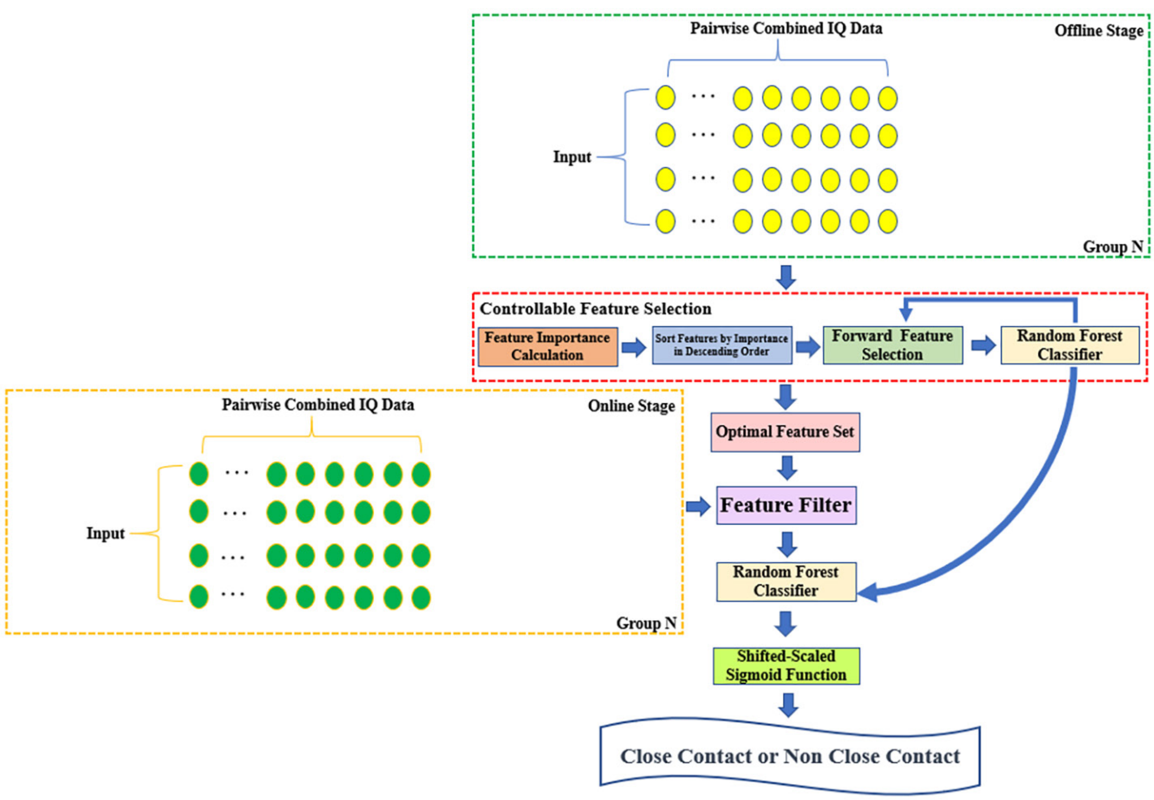


*Figure 6. Binary Contact Detection Model*

there is close contact between individuals based on the BLE Tags they carry. The model outputs a binary value-1 or 0. Here, “1” indicates that there is close contact between two devices, and “0” indicates non-close contact. This step is crucial for identifying risky areas.

All device pairs marked as “1” in the Binary Contact Detection Model are considered close contacts. The positional coordinates of these devices are extracted as inputs to the subsequent alpha shape algorithm. This selection process ensures that the alpha shape algorithm is applied only to the points most likely to be mediators of disease transmission, thereby enhancing the efficiency and focus of the overall analysis.

The Binary Contact Detection Model predicts the likelihood of close contact between each pair of individuals based on the IQ data from their devices. The output value of “1” indicates close contact, while “0” indicates non-close contact. Let $\{p_1, p_2, \ldots, p_n\}$ be the set of positions, and the prediction function $h(p)$ is given by Equation (4).

$$h(p) = \begin{cases} 1, & \textit{if close contact between devices} \\ 0, & \textit{if non}-\textit{close contact} \end{cases} \tag{4}$$

The positional coordinates of all device pairs predicted as “1” form the input position set $E$ for the alpha shape algorithm, that is Equation (5).

$$E = \{e_i | h(p_i) = 1\} \tag{5}$$

Given a set of positions E the alpha shape algorithm is used to identify the shape formed by these positions. An appropriate alpha value is selected that determines the tightness of the algorithm boundary. The convex hull contains all the positions in the set. The Convex Hull is a special case of the alpha shape; specifically, it is the alpha shape when the alpha tends toward infinity. In this limit, all positions are included in a single convex shape without considering any internal structure or holes within the set of positions. The larger the alpha value, the more the resulting alpha shape tends to approximate the convex hull of the position set.

Alpha shape is used to determine the shape of a position set by analyzing its Delaunay triangulation. The construction of the alpha shape is based on the Delaunay triangulation of the position set. The Delaunay triangulation provides a well-defined circumcircle for each triangle, characterized by not containing any other positions. Alpha shapes compare the radius of these circumcircles with a given threshold alpha to decide which triangles should be included in the final shape [33]. Technically, if the circumcircle radius of a simplex (an edge or triangle) in the Delaunay triangulation is less than or equal to alpha, then that simplex is included in the alpha shape [34]. When the alpha value is small, only positions that are very close to each other are connected, resulting in a tighter, more detailed shape that may exhibit more non-convex features. As the alpha value increases, more simplices are included, and the alpha shape gradually expands until it becomes the convex hull of the position set when the alpha value is sufficiently large to exceed the maximum distance between any positions.

Using position set $E$ as input, the alpha shape algorithm is used to determine the boundary of the area formed by these close contacts. Let $Del(E)$ represent the Delaunay triangulation of the position set $E$, and let simplex $\sigma \in Del(E)$ have a circumcircle with radius $r(\sigma)$. The alpha shape can be described by Equation (6) [35].

$$A_{Alpha} = \bigcup\{\sigma \in Del(E) | r(\sigma) \leq Alpha\} \tag{6}$$

where $\sigma$ represents a simplex (an edge or triangle) within the Delaunay triangulation, $r(\sigma)$ is the radius of the circumcircle, and alpha is the parameter that controls the tightness of the shape. In the alpha shape algorithm, triangles whose circumcircle radius is less than or equal to alpha are selected as part of the alpha shape. Each triangle in Delaunay triangulation undergoes a filtering process, ultimately forming a shape that describes the original set of positions.

In the traditional alpha shape algorithm, alpha represents the radius of the largest circle (or sphere) used in the construction of the shape. As the number of positions involved in the calculation increases, the complexity of the Delaunay triangulation also increases. If the alpha value is kept constant, new positions may be added that form new simplices whose circumcircle radii exceed the current alpha value, thereby refining the alpha shape. However, if the alpha value increases with the number of positions, more triangles can be maintained within the alpha shape, making the shape more likely to encompass the Convex Hull, which includes all positions. Here, $n$ is the number of positions in input set $E$.

$$n = |E| \tag{7}$$

$$Alpha = \gamma \times n \tag{8}$$

where $\gamma$ is an adjustment factor that can be modified based on the specific characteristics of the dataset, which is typically set to 1. This dynamic adjustment of the alpha shape’s size adapts to changes in the size of the position set, thus more effectively reflecting the spatial structure and characteristics of the dataset. This setting allows the value of alpha to grow linearly with the number of positions while maintaining sufficient flexibility and coverage while avoiding overly complex shapes. In many practical application scenarios, the collection of positions is not static, but can change over time or under different conditions. Setting alpha proportional to the number of positions, scaled by $\gamma$, provides a natural mechanism by which the alpha shape dynamically adapts to changes in the number of positions.

Through this approach, the combination of the alpha shape algorithm and Binary Contact Detection Model not only provides precise delineation of risky areas but also makes the management of these areas more scientific and systematic, thereby effectively supporting the formulation and implementation of disease prevention and control measures.

Figure 7 illustrates a process for identifying risky areas, where the output of the Binary Contact Detection Model is used in the alpha shape algorithm to calculate risky areas. In this model, the output of the Binary Contact Detection Model is 0 or 1, indicating whether close contact exists between device pairs. If the model output is 1, then the two devices are determined to be in close contact, and their positions are marked and extracted.

Next, the positions of all devices marked as 1 by the Binary Contact Detection Model are collected from position set $E$, which is then inputted into the alpha shape algorithm. The alpha shape algorithm determines the shape boundaries by computing the Delaunay triangulation of these positions and by filtering out triangles or edges whose circumcircle radii do not exceed alpha.

In particular, Delaunay triangulation is first performed for position set $E$. Delaunay triangulation is a special type of triangular mesh in which no positions from the set are inside the circumcircle of any triangle. Next, alpha is calculated based on the number of positions $n$ in the set $E$ and the factor $\gamma$. Then, the alpha shape is constructed by filtering out triangles and edges from the Delaunay triangulation whose circumcircle radii are less than the alpha radii. The filtered simplices form the alpha shape.

Through these steps, the alpha shape algorithm generates a shape boundary that includes all positions from point set $E$. The final output risky area is indicated by the red boundary in Figure 7. This method, by combining the output of the Binary Contact Detection Model with the alpha shape algorithm, not only dynamically adjusts the shape to adapt to changes in the dataset but also precisely identifies and marks the risky areas. This is of significant importance for public health management and disease prevention and control, as it helps to detect and isolate close contacts to prevent the spread of epidemics.

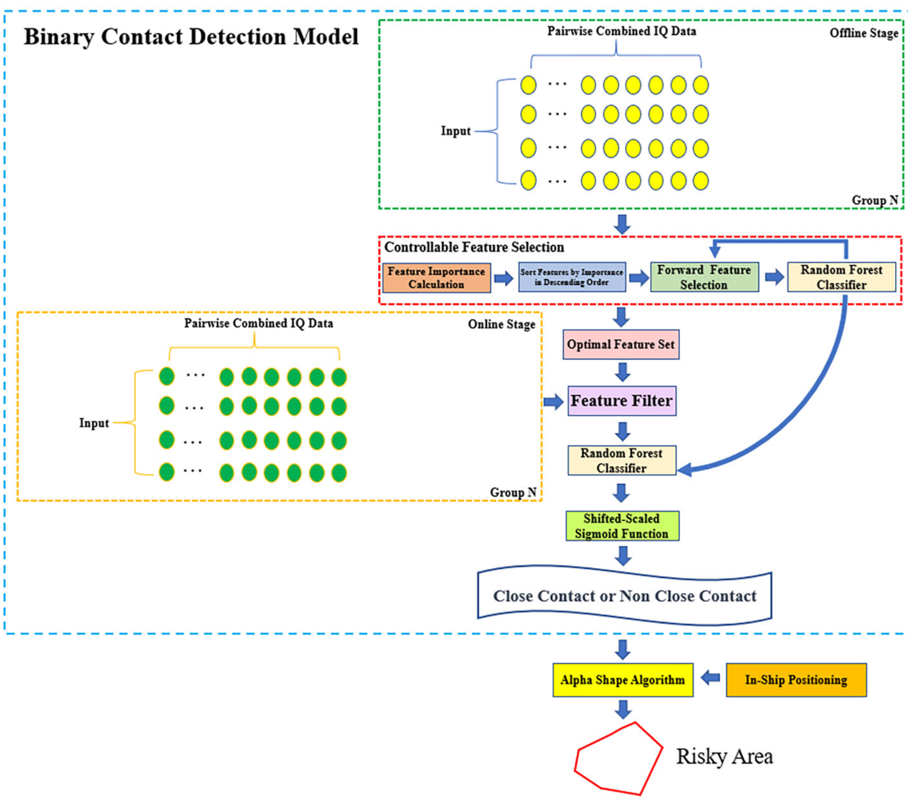


***Figure 7.*** *Risky area identification process based on alpha shape algorithm and binary contact detection model*

## 4. Performance Evaluation

Figure 8 shows the layout of the upper deck of the HANNARA ship used to collect IQ data for the close-contact experiments. In this experiment, BLE Tags are placed in various positions on the deck, and user positions (blue markers) are recorded, allowing the capture and recording of individual BLE tag IQ data. The experiment focused on pairs of individuals within a distance of 1.5 m. These individuals are considered to have close contacts because their physical distance falls within the possible range of disease transmission. Through this experimental layout and data collection, we can use the collected IQ data to train and validate the Binary Contact Detection Model.

In Figure 8, "2-12" indicates that individuals 2 and 12 form a group called group 1, which is abbreviated as "g_1". Therefore, there are 8 groups of close contacts and 8 groups of non-close contacts, totaling 16 groups. The 8 groups of close contacts are "g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8", while the 8 groups of non-close contacts are "g_9, g_10, g_11, g_12, g_13, g_14, g_15, g_16". Table 2 summarizes the key experimental parameters and model configurations used in this study, including the environment, model types, dataset size, and performance metrics.

In Figure 9, the x-axis represents the relative importance calculated by the Binary Contact Detection Model, indicating the relative importance of each feature. This value is typically computed based on how the feature improves the model's predictive performance, for example, in decision trees, it can be based on the gain during node splits. The y-axis lists the names of the features used for learning and prediction in the Binary Contact Detection Model. This graph shows the contribution of IQ data to the decisions made by the Binary Contact Detection Model. The Binary Contact Detection Model uses 144 IQ data points as features. For ease of display, the graph only lists the top 20 features. For example, "IQ_data_1_I_24" represents the I data of the 24th IQ data point from BLE Tag 1. In a ship scenario, these features can represent data points used to identify close contacts. The

***Table 2.*** *Experimental setup*

| Parameter | Description |
|---|---|
| Environment | Python |
| Model type | LightGBM, Random Forest |
| Number of samples | 23,932 |
| Feature count | 144 |
| Train-test split ratio | 70% training and 30% testing |
| Performance metrics | Accuracy, AUC, precision, recall, F1 score, receiver operating characteristic |

LightGBM: Light Gradient Boosting Machine, AUC: Area Under the Curve

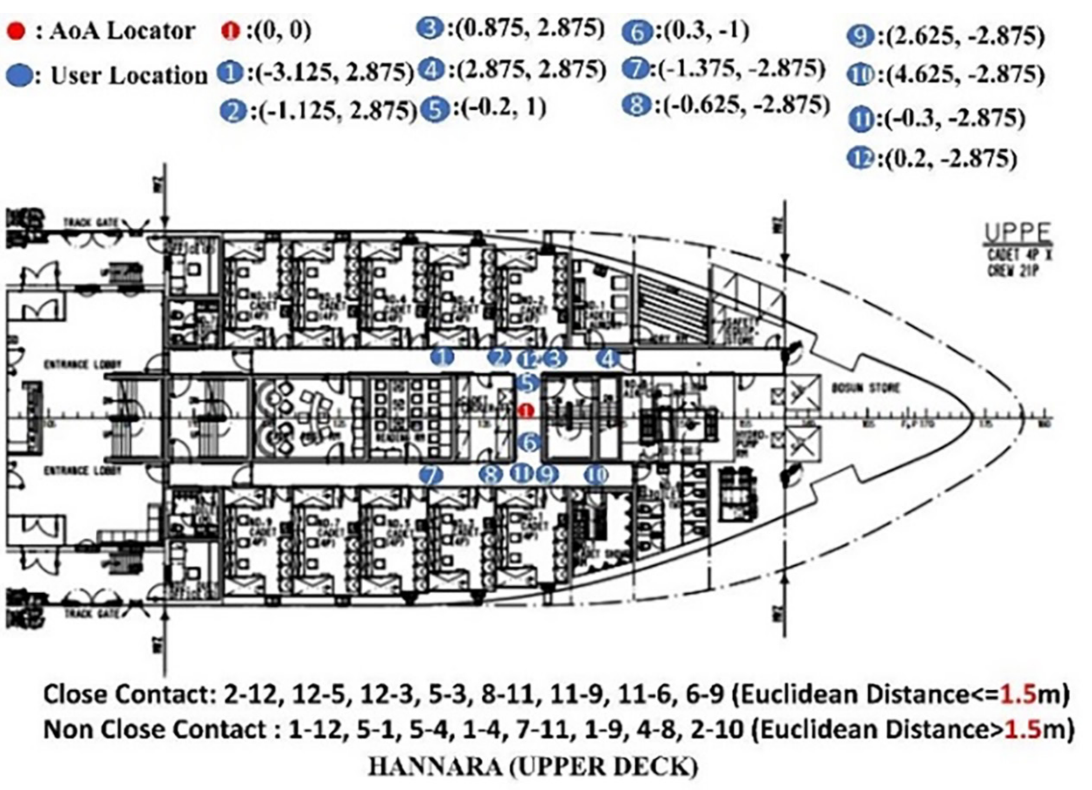


***Figure 8.*** *HANNARA ship upper deck layout for a close contact experiment*

features are sorted by their relative importance, with the most important features playing a more significant role in the predictions made by the Binary Contact Detection Model. This means that these features provide the most information when identifying close contacts.

An in-depth examination of Figure 9 reveals that the top-ranked features have a significantly higher relative importance than the others. This indicates that a small subset of IQ data points strongly influences the model's predictive ability. By focusing on these key features, we can potentially streamline the model for faster computation without sacrificing accuracy. Additionally, understanding the most impactful features can provide insights into the underlying patterns that signify close contacts, thereby improving feature engineering and data collection strategies in future implementations.

As shown in Figure 10, as the number of sorted features increases, the key evaluation metrics such as test accuracy,

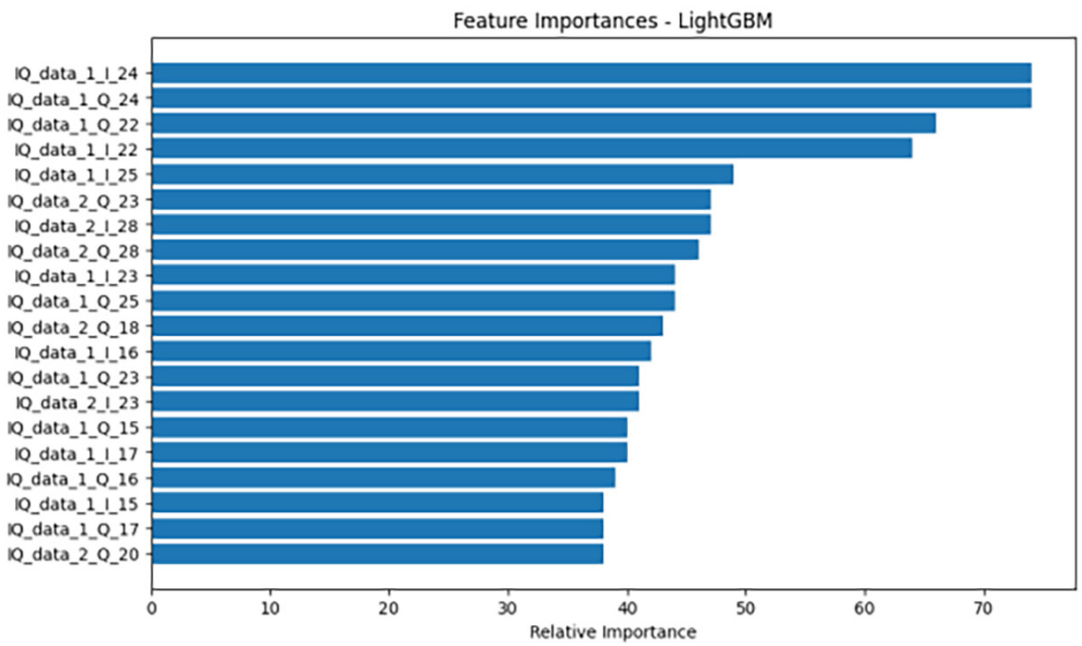


*__Figure 9.__ A graph showing the contribution of IQ data to the model's decision*

*IQ: In-phase and Quadrature*

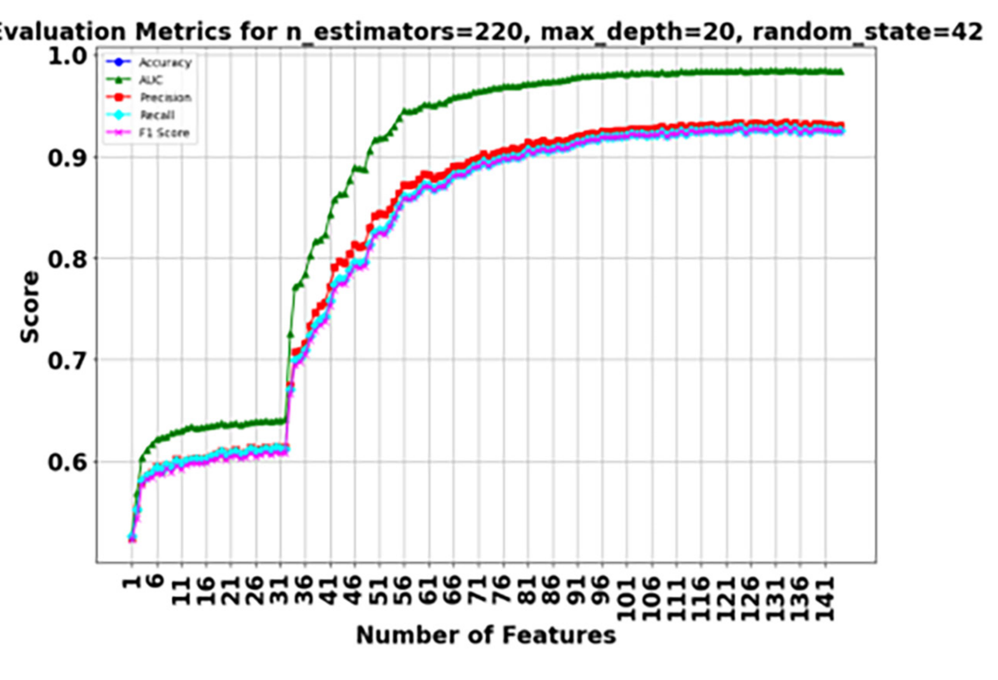


*__Figure 10.__ Impact of sorted number of features on model evaluation metrics*

Area Under the Curve (AUC), precision, recall, and F1 score, of the model initially show signs of improvement; however, these performance metrics stabilize after reaching a certain point. The x-axis represents the "Number of Features" used by the Binary Contact Detection Model for prediction. These features may be columns of the dataset, which the model uses to make predictions. The y-axis represents the "Score", which measures the performance of the model. The score ranges from 0 to 1, with a score closer to 1 indicating better model performance.

In particular, the data demonstrate that when the number of features reached above 33, the model performance significantly improved. Starting from the 33 features, all major performance metrics (accuracy, AUC, precision, recall, F1 score) increased significantly. This trend indicates that initially increasing the number of features can provide more information to the model, thereby enhancing its predictive capability. However, after reaching a certain number of features, performance improvement becomes very limited. In particular, when the number of features is small, model performance improved significantly, especially as the number of features started to increase from a lower range. In particular, model performance improved noticeably before reaching a certain threshold. This threshold represents the turning point at which the model transitions from acquiring essential information to approaching its performance ceiling. The test accuracy and AUC initially improved with an increase in the number of features, indicating that the model's ability to distinguish between different categories is enhanced. However, after the number of features increases to a certain level, the improvement in these metrics diminishes and eventually stabilizes, which indicates that the model has reached its potential limit in distinguishing capability.

The precision, recall, and F1 scores also demonstrate noticeable improvements when the number of features is small, followed by a slowdown in performance enhancement and stabilization as the number of features continues to increase. As the model processes more information, its ability to classify positive and negative samples reaches a certain balance. The results demonstrate that adding features significantly improves model performance when the number of features is small. However, beyond a certain threshold, the performance improvement diminishes and stabilizes, emphasizing the importance of fine feature selection to add necessary information while avoiding excessive noise or irrelevant information.

Ultimately, Figure 10 underscores the principle of the "curse of dimensionality", where adding too many features can lead to diminishing returns or even degrade model performance due to overfitting or increased noise. This highlights the necessity of an optimal feature set that balances the amount of information provided and the potential for introducing irrelevant data. This balance ensures that the model remains both efficient and effective, maximizing the predictive accuracy while minimizing the computational complexity.

Figure 11 shows the receiver operating characteristic (ROC) curve, which shows the performance of the classification model at all possible classification thresholds. The AUC is a metric used to measure the performance of a classifier, with higher AUC values indicating better classifier performance. An ideal ROC curve bends toward the top left corner of the plot, which means that the classifier can achieve a high True Positive Rate (TPR) with a low False Positive Rate (FPR).

In Figure 11, the ROC curve is displayed in orange, starting from (0,0), rapidly rising to near the (0,1) point, and then

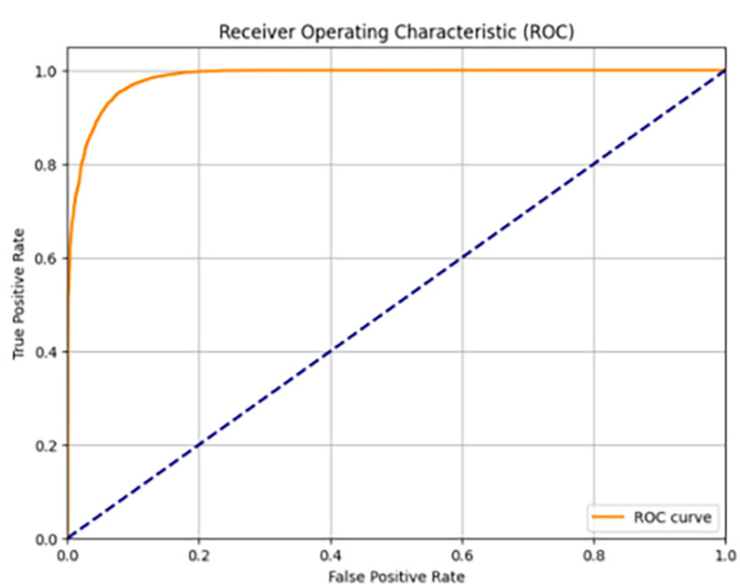


***Figure 11**. Receiver operating characteristics*

gradually approaching (1,1) to the right. The blue dashed line represents the performance of a Random Classifier (RF), essentially the result of random guessing, and the slope of this line is 1, indicating that the TPR and FPR are equal for the classifier. Ideally, the ROC curve should be above the blue dashed line, indicating that the classifier's performance is better than random guessing. In Figure 11, the ROC curve is clearly above the blue dashed line, indicating that the classifier exhibits good classification performance.

The sharp ascent of the ROC curve in Figure 11 toward the top left corner demonstrates the model's strong ability to distinguish between the positive and negative classes. The substantial AUC indicates that the model performs significantly better than random guessing. This high AUC value signifies that the model exhibits a high TPR while maintaining a low FPR across various thresholds. Such performance is crucial in applications like close contact detection on ships, where accurately identifying potential risks without generating excessive false alarms is essential for effective disease prevention and control measures.

Figure 12 shows the results of 8 overlapping risky areas on the HANNARA upper deck. The x-axis represents the pixel width, which can be understood as the horizontal resolution of the image, and the y-axis represents the pixel height, corresponding to the vertical resolution of the image. This is the visual overlap of 8 risky areas identified by the model. The risky areas are determined based on the likelihood of close contact, which involves disease prevention and ship safety monitoring. In the daily operation of a ship, these results can be used to monitor the flow and activities of personnel to ensure the safety of passengers and crew. If a specific area is marked as high-risk, measures can be taken to restrict access to that area or increase the frequency of cleaning and disinfection. This not only provides an intuitive representation of the risky areas but also helps managers take preventive measures to reduce health risks on the ship. By monitoring and adjusting the usage patterns of the ship, the spread of diseases onboard can be effectively controlled and prevented.

## 5. Discussion

Our experimental results demonstrate that after filtering out noise features, the model's performance improved significantly as the number of features increased from a lower range [36]. This indicates that utilizing more relevant information greatly enhances the model's ability to identify close contacts. This trend emphasizes the importance of providing the model with sufficient and relevant information to ensure accurate identification of close contacts and areas of transmission risk.

Our goal is to first filter out noise features and optimize the feature set, thereby ensuring that the model receives only effective information that contributes to improved identification accuracy. Filtering out noise features not only reduces the model's complexity but also enhances its robustness when handling high-density, complex interpersonal interaction environments [37]. Subsequently, as the number of features increases, the model leverages more relevant information to further enhance its ability to identify close contacts and transmission risky areas. The proposed method effectively combines feature selection and feature expansion to ensure that the model remains efficient and precise while being information rich.

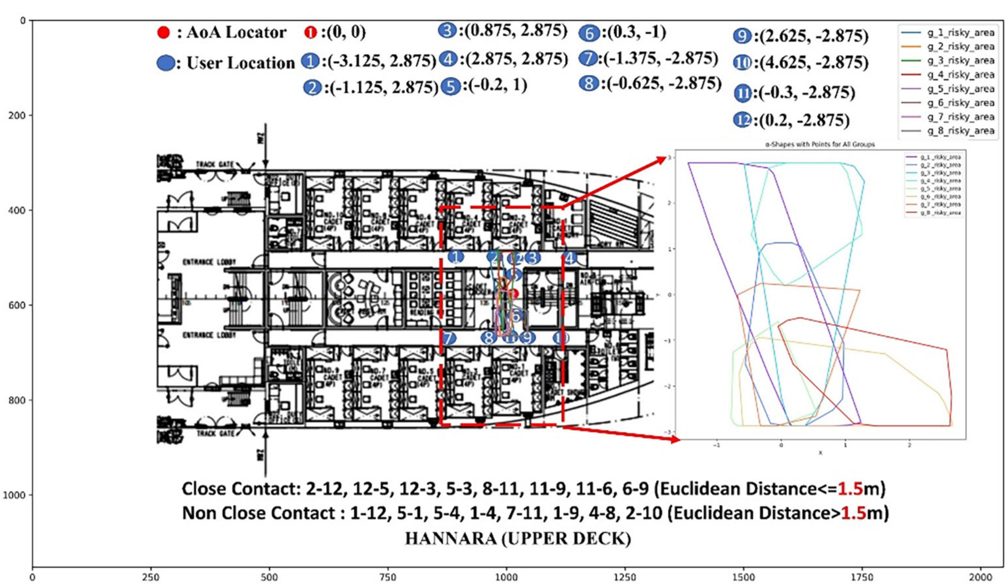


*Figure 12*. *Results of 8 risky areas on the HANNARA upper deck*

The proposed method primarily relies on Bluetooth 5.1 technology to track close contact between individuals and employs a Binary Contact Detection Model to analyze these data. Although this technology is highly effective at detecting direct interpersonal contact, it does not capture pathogens transmitted through the air. However, timely identification of close contacts can still significantly reduce the speed of virus transmission and help control the spread of infection.

Noise primarily causes errors in positioning systems, which is a challenge in any environment [38]. The proposed model has a strong ability to filter out noise, which improves accuracy in any setting. While testing in different ship environments is desirable and can provide additional insights, it is not essential for demonstrating the effectiveness of our noise-filtering approach. The key contribution of our work lies in the model's capability to handle noise effectively, which we have thoroughly validated through our experiments. The HANNARA training ship accommodates hundreds of students, each with their own living space. This high-density passenger environment closely resembles actual operational scenarios of cruise ships, effectively simulating the movements and close contact among passengers of real cruise ships.

Compared to existing technologies for contact tracing and risky area identification, such as WiFi-based positioning systems and RFID tracking, the proposed solution using Bluetooth 5.1 technology combined with machine learning-based feature filtering offers superior accuracy, energy efficiency, and ease of deployment in the complex environment of a cruise ship. WiFi-based positioning systems, while utilizing existing infrastructures, consume significant power, making them unsuitable for the long-term tracking of contacts. High power consumption poses challenges for continuous operation over extended periods. In contrast, Bluetooth technology offers lower power consumption and wider device coverage, facilitating easy deployment for continuous monitoring. Bluetooth devices are energy-efficient, enabling long-term tracking without frequent battery replacements or recharging [39].

In addition, ordinary devices can be converted into Bluetooth Tags by modifying the communication protocols, thus eliminating the need for specialized hardware and simplifying deployment. This adaptability allows for cost-effective implementation and scalability across different ship sizes and configurations. Traditional BLE methods without AoA capabilities rely on the Received Signal Strength Indicator values and are highly sensitive to environmental factors, leading to unreliable proximity detection. The proposed method mitigates these limitations by effectively filtering out noisy features that represent environmental interference, thereby enhancing the quality of angle estimations and improving the positioning accuracy, which is crucial for close contact detection [40]. While the initial investment and requirement for specific hardware are considered, the increased accuracy, robustness to interference, energy efficiency, and ease of deployment of our system present significant advantages over existing solutions, providing a robust and adaptable approach tailored to the unique challenges of the ship environment.

In subsequent practical applications, data anonymization techniques are employed to encrypt and hash passengers' unique identifiers (such as device IDs) and remove sensitive information related to personal identities, thereby ensuring data privacy. All data are securely stored in high encrypted databases using advanced encryption algorithms (such as AES-256), and only authorized personnel have access. Additionally, the system design strictly adheres to relevant data protection regulations, including the European Union's General Data Protection Regulation, to ensure that data processing activities are lawful, transparent, and fair [41]. Upon boarding, we ensure that passengers fully understand the purposes, scope, and rights of data collection and usage through clear notifications and easy-to-understand consent processes, allowing them to voluntarily participate in the system's use. To prevent potential misuse of the technology, the system design incorporates multiple protective measures, such as strict access controls, data usage monitoring, and anomaly detection, to ensure that data are used solely for legitimate purposes. Implementing the proposed solution requires an initial investment in Bluetooth 5.1 devices, specifically locators and tags, and the necessary infrastructure on cruise ships [42]. While this upfront cost is a consideration, the system's ability to rapidly identify and isolate close contacts can significantly reduce the spread of diseases, leading to substantial savings from preventing outbreaks and minimizing disruptions to cruise operations. Ongoing operational costs, such as maintenance and staff training, can be offset by these potential savings. Furthermore, the system is designed to be scalable and adaptable to ships of different sizes and passenger capacities; by adjusting the number of locators and tags deployed, the solution can be efficiently scaled, ensuring both cost-effectiveness and optimal performance across various ship configurations.

## 6. Conclusion

This study effectively integrates Bluetooth 5.1 technology with a Binary Contact Detection Model and an alpha shape algorithm to enhance disease surveillance on cruise ships. The key results demonstrate that when the model uses more than 33 features, there is a significant improvement in performance metrics-including accuracy, AUC, precision, recall, and F1 score-leading to more accurate detection of close contacts. In addition, the alpha shape algorithm successfully identified eight overlapping high-risk areas on the ship's upper deck. The main contributions of this paper are the development of a precise and efficient real-time monitoring system that reduces false detections and enhances disease control capabilities in complex, high-density environments, ensuring passenger safety through targeted interventions. Future work will focus on refining the algorithms to further enhance accuracy and reduce false positives, integrating machine learning techniques to predict outbreak patterns, and expanding the system's application to other high-density settings. Collaborations with public health authorities will be essential to develop standardized protocols for data sharing and privacy protection, maximizing the system's effectiveness while ensuring ethical considerations are met.

## Footnotes

### Authorship Contributions

Concept design: Q. Lin, and J. Son, Data Collection or Processing: Q. Lin, Analysis or Interpretation: Q. Lin, and J. Son, Literature Review: Q. Lin, and J. Son, Writing, Reviewing and Editing: Q. Lin, and J. Son.

**Funding:** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (grant number: 2021R1I1A3056125).

## References

[1] N. Diller, "Nearly 70 celebrity cruises guests sick in norovirus outbreak". *USA TODAY*, Jun 2024.

[2] E. Dahl, "Coronavirus (COVID-19) outbreak on the cruise ship Diamond Princess". *International Maritime Health*, vol. 71, pp. 5-8, Mar 2020.

[3] K. Ritos, D. Drikakis, and I. W. Kokkinakis, "Virus spreading in cruiser cabin". *Physics of Fluids*, vol. 35, pp. 1-16, Oct 2023.

[4] W. Ahmed et al., "Detection of SARS-CoV-2 RNA in commercial passenger aircraft and cruise ship wastewater: a surveillance tool for assessing the presence of COVID-19 infected travelers". *Journal of Travel Medicine*, vol. 27, pp. 1-11, Aug 2020.

[5] T. Sekizuka, et al., "Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak". *Proceedings of the National Academy of Sciences*, vol. 117, pp. 20198-20201, Aug 2020.

[6] K. Gravningen, et al., "Risk factors, immune response and whole-genome sequencing of SARS-CoV-2 in a cruise ship outbreak in Norway". *International Journal of Infectious Diseases*, vol. 118, pp. 10-20, May 2022.

[7] L.-S. Huang, L. Li, L. Dunn, and M. He, "Taking account of asymptomatic infections: A modeling study of the COVID-19 outbreak on the Diamond Princess cruise ship". *PLoS ONE*, vol. 16, pp. 1-10, Mar 2021.

[8] A. J. Kucharski, et al., "Early dynamics of transmission and control of COVID-19: a mathematical modelling study". *The Lancet Infectious Diseases*, vol. 20, pp. 553-338, Mar 2020.

[9] X. Liu, and Y. C. Chang, "An emergency responding mechanism for cruise epidemic prevention—taking COVID-19 as an example". *Marine Policy*, vol. 119, pp. 1-11, Sep 2020.

[10] Q. Lin, J. Son, and H. Shin, "A self-learning mean optimization filter to improve Bluetooth 5.1 AoA indoor positioning accuracy for ship environments". *Journal of King Saud University-Computer and Information Sciences*, vol. 35, pp. 59-73, Mar 2023.

[11] S. J. Curtis, et al., "Feasibility of bluetooth low energy wearable tags to quantify healthcare worker proximity networks and patient close contact: A pilot study". *Infection, Disease & Health*, vol. 27, pp. 66-70, May 2022.

[12] L. Reichert, S. Brack, and B. Scheuermann, "A survey of automatic contact tracing approaches using bluetooth low energy". *ACM Transactions on Computing for Healthcare*, vol. 2, pp. 1-33, Mar 2021.

[13] F. Daniş, and A. Cemgil, "Model-based localization and tracking using bluetooth low-energy beacons". *Sensors*, vol. 17, pp. 1-23, Oct 2017.

[14] Z. Su, K. Pahlavan, and E. Agu, "Performance evaluation of COVID-19 proximity detection using bluetooth LE signal". *IEEE Access*, vol. 9, pp. 38891-38906, Oct 2021.

[15] Q. Lin, and J. Son, "A close contact tracing method based on bluetooth signals applicable to ship environments". *KSII Transactions on Internet and Information Systems*, vol. 17, pp. 1-19, Feb 2023.

[16] Q. Lin, and J. Son, "A close contact identification algorithm using kernel density estimation for the ship passenger health". *Journal of King Saud University - Computer and Information Sciences*, vol. 35, pp. 101564-101564, Jun 2023.

[17] V. Bianchi, P. Ciampolini, and I. De Munari, "RSSI-based indoor localization and identification for ZigBee wireless sensor networks in smart homes". *IEEE Transactions on Instrumentation and Measurement*, vol. 68, pp. 566-575, Feb 2019.

[18] M. S. Munir, D. H. Kim, A. K. Bairagi, and C. S. Hong, "When CVaR meets with bluetooth PAN: a physical distancing system for COVID-19 proactive safety". *IEEE Sensors Journal*, vol. 21, pp. 13858-13869, Jun 2021.

[19] H. Ghayvat, M. Awais, P. Gope, S. Pandya, and S. Majumdar, "ReCognizing SUspect and PredictiNg ThE SpRead of Contagion Based on Mobile Phone LoCation DaTa (COUNTERACT): a system of identifying COVID-19 infectious and hazardous sites, detecting disease outbreaks based on the internet of things, edge computing, and artificial intelligence". *Sustainable Cities and Society*, vol. 69, pp. 1-11, Jun 2021.

[20] D. J. Leith and S. Farrell, "Coronavirus contact tracing". *ACM SIGCOMM Computer Communication Review*, vol. 50, pp. 66-74, Oct 2020.

[21] P. C. Ng, P. Spachos, and K. N. Plataniotis, "COVID-19 and your smartphone: BLE-based smart contact tracing". *IEEE Systems Journal*, pp. 1-12, Dec 2021.

[22] M. T. Rahman, R. T. Khan, M. R. A. Khandaker, M. Sellathurai, and M. S. A. Salan, "An automated contact tracing approach for controlling Covid-19 spread based on geolocation data from mobile cellular networks". *IEEE Access*, vol. 8, pp. 213554-213565, Dec 2020.

[23] A. Berke, K. Larson, A. S. Pentland, M. Bakker, P. Vepakomma, and D. Calacci, "Assessing disease exposure risk with location data: A proposal for cryptographic preservation of privacy". *Mitmedialab,* vol. 1, pp. 1-15, Mar 2020, arXiv preprint arXiv:2003.14412.

[24] P. G. Madoery, et al., "Feature selection for proximity estimation in COVID-19 contact tracing apps based on bluetooth low energy (BLE)". *Pervasive and Mobile Computing*, vol. 77, pp. 1-14, Oct 2021.

[25] G. Barthe, et al., "Listening to bluetooth beacons for epidemic risk mitigation". *Scientific Reports*, vol. 12, pp. 1-14, Apr 2022.

[26] Q. Lin, and J. Son, "Sustainable ship management post COVID-19 with in-ship positioning services". *Sustainability*, vol. 14, pp. 1-23, Dec 2021.

[27] Q. Lin, J. Son, and H. Shin, "A mean optimization filter to improve bluetooth aoa indoor positioning accuracy for ship environments". *IPIN-WiP*, vol. 1, pp. 1-16, Sep 2022.

[28] Q. Lin, and J. Son, "An in-ship localization algorithm for close contact identification". *Journal of Theoretical and Applied Information Technology*, vol. 100, pp. 1706-1717, Mar 2022.

[29] N. B. Suryavanshi, K. Viswavardhan Reddy, V. R. Chandrika, "Direction finding capability in Bluetooth 5.1 standard". *Ubiquitous Communications and Network Computing: Second EAI International Conference*, Feb 2019.

[30] P. Sambu, and M. Won, "An experimental study on direction finding of Bluetooth 5.1: indoor vs outdoor". *2022 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr 2022.

[31] Y. Zhuang, C. Zhang, J. Huai, Y. Li, L. Chen, and R. Chen, "Bluetooth localization technology: principles, applications, and future trends". *IEEE Internet of Things Journal*, vol. 9, pp. 23506-23524, Dec 2022.

[32] F. A. Toasa, L. Tello-Oquendo, Carlos Ramiro Peñafiel-Ojeda, and G. Cuzco, "Experimental demonstration for indoor localization based on AoA of Bluetooth 5.1 using software defined radio". *IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)*, Jan 2021.

[33] H. Edelsbrunner, and E. P. Mücke, "Three-dimensional alpha shapes". *ACM Transactions on Graphics*, vol. 13, pp. 43-72, Jan 1994.

[34] H. Edelsbrunner, "Alpha shapes-a survey". *In: Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings,* vol. 1, pp. 1-25, Dec 2011.

[35] N. Akkiraju, H. Edelsbrunner, M. Facello, P. Fu, E. P. Mücke, and C. Varela, "Alpha shapes: definition and software". In *Proceedings of the 1st international computational geometry software workshop,* Sep 1995.

[36] G. K. Rajbahadur, S. Wang, G. A. Oliva, Y. Kamei, and A. E. Hassan, "The impact of feature importance methods on the interpretation of defect classifiers". *IEEE Transactions on Software Engineering*, vol. 48, pp. 2245-2261, Feb 2021.

[37] D. Rengasamy, et al., "Feature importance in machine learning models: A fuzzy information fusion approach". *Neurocomputing*, vol. 511, pp. 163-174, Sep 2022.

[38] X. Zhu, J. Yi, J. Cheng, and L. He, "Adapted error map based mobile robot UWB indoor positioning". *IEEE Transactions on Instrumentation and Measurement*, vol. 69, pp. 6336-6350, Jan 2020.

[39] B. Etzlinger, B. Nußbaummüller, P. Peterseil, and K. A. Hummel, "Distance estimation for BLE-based contact tracing–a measurement study". *2021 Wireless Days (WD)*, Aug 2021.

[40] B. Nußbaummüller, B. Etzlinger, and K. A. Hummel, "BLE-based contact tracing: characterization of distance estimation errors and mitigation options". *IEEE Pervasive Computing*, vol. 22, pp. 7-14, Nov 2023.

[41] C. Badii, P. Bellini, A. Difino, and P. Nesi, "Smart city IoT platform respecting GDPR privacy and security aspects". *IEEE Access*, vol. 8, pp. 23601-23623, Jan 2020.

[42] D. Antonioli, N. O. Tippenhauer, K. B. Rasmussen, and M. Payer, "BLURtooth: exploiting cross-transport key derivation in bluetooth classic and bluetooth low energy". In *Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security,* Jun 2022.