JOURNAL OF BUILDING PERFORMANCE SIMULATION
https://doi.org/10.1080/19401493.2024.2395269

Enhancing healthcare facility resilience: utilizing machine learning model for airborne disease infection prediction

Kangkang Tang

Department of Civil and Environmental Engineering, Brunel University London, London, UK

ABSTRACT

During this pandemic, advanced epidemiological models have been widely employed to determine intervention strategies for controlling the spread of the disease in public and healthcare settings. These models play a crucial role by providing predictive insights into disease transmission dynamics, informing resource allocation and guiding policy decisions. However, the accuracy of these predictions depends on a substantial amount of input data, which was not readily available at the onset of the pandemic. Another concern with the existing models is their inability to adequately account for the complex indoor built environments, which has been shown to significantly impact infection risk. To tackle these issues, this paper discusses the potential of developing a joint modelling technique that integrates machine learning models, building information models and agent-based models to assess the risk of nosocomial airborne infections. With limited available data, machine learning models can determine infection risk with high confidence. By incorporating these risk estimates into agent-based models within a more comprehensive building model, we conducted a thorough investigation of nosocomial infections. This approach also considers human movement patterns across various indoor conditions, enhancing the depth of our analysis.

ARTICLE HISTORY
Received 2 February 2024
Accepted 17 August 2024

KEYWORDS
Artificial neural network modelling (ANN); Respiratory syndrome coronavirus 2 (SARS-CoV-2); Agent-based modelling (ABM); Building information modelling (BIM); Modern methods of construction (MMC)

1. Research background

Significant instances of nosocomial transmissions occurred at the early stage of the COVID-19 pandemic. In the UK, approximately one-seventh to one-fifth of COVID-19 patients, as well as the majority of infected healthcare workers, contracted the disease in healthcare facilities during the first wave (Campbell and Barr 2021; Evans et al. 2020; Read et al. 2021). Rickman et al. (2020) investigated nosocomial COVID-19 transmission cases in London hospitals and reported that 55% of the patient-to-patient infections took place in the same bay; 14% were in different bays but on the same floor; 12% of infections actually happened in the single-occupancy rooms. Besides the serious challenges posed by nosocomial infections, the COVID-19 pandemic has underscored another critical aspect: the resilience of healthcare facilities in coping with a surge of patients. Many hospitals found it challenging to deliver adequate healthcare services while managing the overwhelming patient influx. This pandemic has highlighted the existing issues regarding resilience of healthcare systems on a significant scale, capturing widespread attention.

The readiness and responsiveness of the healthcare system rely on multifaceted strategies and approaches. Preparedness plans, surge capacity planning, and multi-department collaboration form the foundation for effective outbreak response, ensuring the allocation of resources and coordination of efforts. Among these, epidemiological modelling has played a crucial role by providing predictive insights into disease transmission dynamics, informing resource allocation and guiding policy decisions. During this pandemic, advanced epidemiological models have been widely used to determine intervention strategies for controlling the spread of the disease in public and healthcare settings (Currie et al. 2020; Swallow et al. 2022). It is important to note that the accuracy of these predictions typically depends on a significant volume of input data to validate the models and verify the results. However, the inadequate and sometimes conflicting data available at the onset of the pandemic has posed significant challenges for achieving accurate predictions. The following chapter examines the evidence and data concerning COVID-19 transmission. As the pandemic has evolved, discussion have expanded to

CONTACT Kangkang Tang kangkangtang@gmail.com; kangkang.tang@brunel.ac.uk

consider the impact of a growing variety and scope of data sources, all of which hold significant importance for computer modelling.

2. Factors influencing human-to-human transmission of COVID-19

This chapter examines challenges associated with the modelling of COVID-19 within hospital environments, including uncertainties in transmission characteristics, early-stage debates, evolving viral properties and inflexible policies. Collectively, these factors posed significant challenges to accurately predicting disease transmission in the early stages.

2.1. Primary modes of transmission

A respiratory disease typically spreads through droplet, direct and indirect contact, as well as airborne routes. During the initial stages of the pandemic, it was believed, as per the March 2020 WHO report (2020), that the SARS-CoV-2 virus was primarily transmitted through respiratory droplets, direct and indirect contact, with a relatively lower likelihood of airborne transmission. Since droplet transmission typically takes place when an individual is in close proximity of an infected person, it was believed that practicing social distancing such as 2 m would be an effective measure in controlling the spread of the disease. Nevertheless, increasing scientific evidence suggests that the transmission of SARS-CoV-2 could occur through airborne spread which became recognized as the predominant mode of transmission in indoor environments (Greenhalgh et al. 2021; Morawska and Milton 2020). According to Qureshi et al. (2020), SARS-CoV-2 can remain viable in smaller aerosols capable of travelling distances of up to 8 m or more, even in the absence of ventilation or airflow. This extended its infection risk over long distances is further corroborated by experimental evidence demonstrating the virus's viability in aerosols for up to 16 hours (Daphne et al. 2022).

Understanding the transmission characteristics of a new disease like COVID-19 is crucial for epidemiological modelling, which plays a vital role in informing healthcare authorities' efforts to devise effective strategies for prevention and control. Yet, defining these characteristics in the initial phases of the pandemic proved challenging due to divergent perspectives among researchers. This uncertainty presented significant challenges for healthcare authorities, impacting the development of protocols and guidelines to manage nosocomial infections. Despite advancements in healthcare research, similar challenges arising from conflicting information may persist in the future.

2.2. Viral properties

The reproductive success of a virus, also known as its fitness for replication, refers to its ability to effectively replicate within a host cell and produce new virus particles capable of infecting other cells and hosts. Assessing the fitness for replication of SARS-CoV-2 in humans is crucial in determining its infectiousness. SARS-CoV-2 has been found to replicate more effectively in the upper respiratory tract when compared to SARS-CoV, which led to the 2003 SARS outbreak (V'kovski et al. 2021). Different strains of SARS-CoV-2 however demonstrate varying preferences for localized replication within either the upper or lower respiratory tract (Ulrich et al. 2022). Understanding these preferences is pivotal in developing more effective prevention and control strategies. When a particular strain shows a higher propensity for upper respiratory tract replication, prioritizing preventive measures like mask usage and adherence to social distancing becomes crucial to mitigate transmission via respiratory droplets. Conversely, strains favouring the lower respiratory tract may require additional interventions targeting aerosol transmission to protect individuals in close proximity to infected individuals. Despite the significant disparities observed among various strains of the disease, substantial progress has been achieved in providing estimates to evaluate infection severity and assess the risk of viral transmission. In the context of respiratory diseases, the viral load, quantified as RNA copies per millilitre (RNA/ml) of sputum, serves as a crucial metric for evaluating infection severity and estimating the risk of viral transmission. Elevated viral loads are frequently linked to heightened infectivity and increased disease severity. In the case of SARS-CoV-2 transmissions, a viral load below 10^7 RNA/ml copies/ml is believed to be associated with mild-to-moderate cases, whereas a higher viral load, up to 10^9 RNA copies, might lead to moderate-to-severe cases (Aganovic et al. 2021).

2.3. Impact of close contact distance

Many countries implement contact tracing as part of their public health policies, typically using a distance of up to 2 m or 6 feet to identify close contacts. This 'one-size-fits-all' approach was based on the assumption that respiratory droplets were the primary mode of SARS-CoV-2 transmission. However, as detailed in Section 2.1, extensive research evidence indicate that airborne transmission through smaller aerosols may travel distances of up to 8 m or more, even in the absence of ventilation or airflow (Qureshi et al. 2020). Consequently, this standardized approach may not be well-suited to its intended purpose.

2.4. Effects of indoor environmental conditions on SARS-CoV-2 transmissibility

The transmission potential of SARS-CoV-2 indoors is significantly influenced by various indoor conditions, including air exchange rate, temperature, humidity, and sunlight intensity. Adequate ventilation can lower the indoor concentration of airborne pathogens, thus reducing the risk of infection. Additionally, elevated humidity and high temperatures can decrease the infectivity of influenza viruses.

Dabisch et al. (2021) reported that both room temperature and air humidity can significantly impact the infectivity of SARS-CoV-2. When air humidity is held constant, the duration for a 90% virus decay increases from 11.5 min to 19.5 min as the room temperature decreases from 30°C to 10°C. Under a consistent temperature, the decay rate rises from $0.6 \pm 0.6\%$ to $1.5 \pm 0.5\%$ with an increased relative humidity from 20% to 70% (Dabisch et al. 2021). This indicates that lower temperatures prolong the survival of the virus, suggesting a potential for increased infectivity in colder environments; higher humidity levels however accelerate the decay of the virus, potentially reducing its infectivity. As HVAC (heating, ventilation and air conditioning) systems are essential tools for managing indoor environmental factors such as humidity and temperature, they hold a crucial position in regulating the transmission of diseases within indoor spaces (Saran et al. 2020).

The assessment of transmission risk, especially when it occurs through direct or indirect contact, is significantly influenced by the duration for which viruses remain infectious and capable of propagating on various surfaces or objects. Hirose et al. (2021) conducted a study comparing the stability of SARS-CoV-2 and influenza A virus on human skin, and found that SARS-CoV-2 could remain viable for approximately 9 hours, significantly longer than the survival time of influenza A virus, which was 1.8 h only. SARS-CoV-2 can survive even longer on inanimate surfaces. According to Geng and Wang (2023), SARS-CoV-2 has the capability to maintain viability on solid surfaces, including plastic, glass, and stainless steel, for a duration of up to 28 days at a room temperature of 20°C. Consequently, there is a potential risk of infection through contact with contaminated surfaces or objects (or fomites), although the associated risk is considered low. According to CDC (2021), the probability of infection from each contact with a contaminated surface is estimated to be less than 1 in 10,000. Nevertheless, a review of the literature has not yielded clear evidence of actual infection cases resulting from fomite or hand contact.

3. Epidemiological modelling of airborne disease transmission under indoor environmental conditions

This chapter examines conventional simulation methods for airborne diseases under indoor conditions. As the discussion progresses, attention turns towards the inherent research gaps in these modelling approaches, laying the groundwork for further investigation into the key factors influencing transmission and potential solutions.

3.1. Traditional airborne infection transmission models

Several epidemiological models have been developed to assess the risk of airborne disease transmission in indoor environments, with the Wells-Riley model standing out as a popular model to calculate the probability of infection based on factors such as the number of infectious individuals, the ventilation rate, the duration of exposure and the respiratory rate of the occupants (Riley, Murphy, and Riley 1978). The Wells-Riley equation can be written as:

$$P_I = 1 - \exp\left(-\frac{Iqpt}{Q}\right) \quad (1)$$

where P_I is the probability of infection; I is the number of infectors, p is the pulmonary ventilation rate of the susceptible (m^3/hr); q is the quanta generation rate of patients (quanta/hr/person); t is the exposure time (h) and Q is the room ventilation rate (m^3/h).

The Wells-Riley model offers valuable insights into the potential spread of diseases across different environments, making it a fundamental tool in epidemiological studies focused on airborne pathogens. It should be noted that variable q in Equation (1) is a hypothetical unit which cannot be directly determined through experiments, although it could be obtained through 'unexpected' occurrences. The outbreak of influenza on a commercial aircraft in 1999 led to 20 passenger infections (out of 74 passengers in total), indicating a 27% infection rate (P_I) (Marsden 2003). Considering the standard aircraft cabin air exchange rate and flight hours (t), q was quantitatively decided and this paves the way to assess the risk of outbreak of a similar airborne disease under different environmental conditions (Marsden 2003). In line with the same principle, various researchers (Buonanno, Stabile, and Morawska 2020; de Oliveira et al. 2021) have suggested the anticipated risks of COVID-19 transmission by employing diverse input values, as detailed in Table 1.

The original Wells-Riley model does not consider other crucial indoor conditions such as temperature and

Table 1. Data from previous studies assessing the indoor infection risk of SARS-CoV-2.

Infection risk P_I (%)	Ip (m^3/hr)	V (m^3)	Q (ACH)	T (min)	Temperature (°C)	Humidity (%)	References
4	0.96	75	2.2	10	N/A	N/A	Buonanno, Stabile, and Morawska 2020
2.1	1.38	800	0.5	12	N/A	N/A	Buonanno, Stabile, and Morawska 2020
0.5	1.87	200	5	60	N/A	N/A	de Oliveira et al. 2021
0.1–1	0.94-1.16	100	0	20	N/A	N/A	Schijven et al. 2021
2–30	0.94-1.16	100	0	120	N/A	N/A	Schijven et al. 2021
0.8–13	0.94-1.16	100	6	120	N/A	N/A	Schijven et al. 2021
7.324.476.836.195.18	0.52	180	0.5	60	20	2037537083.5	Aganovic et al. 2021
14.6714.9913.7712.2710.19	0.52	180	0.5	120	20	2037537083.5	
6.306.305.925.444.64	0.52	180	2	60	20	2037537083.5	
12.2212.3211.5810.629.02	0.52	180	2	120	20	2037537083.5	
4.754.754.544.223.79	0.52	180	6	60	20	2037537083.5	
9.399.288.918.277.36	0.52	180	6	120	20	2037537083.5	

humidity. To bridge this deficiency, Aganovic et al. (2021) developed a revised Wells-Riley model to assess the infection risk of SARS-CoV-2 by considering the effect of humidity, or Equation (2).

$$P_I = 1 - \exp\left[-Ip\int_0^T n(t)dt\right] \quad (2)$$

Aganovic et al. (2021) assumed a constant value of 0.52 (m^3/hr) to represent the product of I and p. Alternatively, the respiratory ventilation rate of an infected individual was set at 0.52 (m^3/hr). In Equation (2), $n(t)$ represents the quanta concentration under an indoor environment at time t and it can be calculated according to Equation (3).

$$n(t) = n_0 \cdot \exp\left[-\left(\frac{Q}{V} + D + k\right)t\right] + \frac{S}{V}\left\{\frac{1}{\frac{Q}{V} + D + k} - \frac{1}{\frac{Q}{V} + D + k}\exp\left[-\left(\frac{Q}{V} + D + k\right)t\right]\right\} \quad (3)$$

where V represents the volume of the room (m^3). n_0 is the initial quanta concentration at time $t = 0$. k represents the virus inactivation rate (1/hr), while D stands for the aerosol deposition rate, both of which have shown significant susceptibility to changes influenced by indoor humidity levels. S denotes the quanta source emission rate from infected individuals (quanta/hr) and was determined by the authors based on particle size. The development of Equations (2) and (3) facilitates the evaluation of airborne disease transmission risks by incorporating indoor building conditions, such as room volume (V), air exchange rate (Q), total exposure time (T), room temperature and humidity, utilizing the developed aerosol deposition rate (D). Building on this framework, Aganovic et al. (2021) calculated the infection risks for moderate-to-severe COVID-19 cases, assuming a viral load in the sputum of approximately 10^9 RNA/ml. The outcomes of this calculation are presented in Table 1.

Other advancements in implementing the Wells-Riley model include the development of a tri-component epidemiological model, as proposed by the author of this study (Tang and Chen 2021, 2023). The tri-component epidemiological model incorporates an agent-based model (ABM) and an infection model within a hospital building information model. Such an epidemiological model enables the inclusion of roles of complex hospital infrastructures and human motions in disease transmissions through spatially explicit stochastic simulations. Agent-based modelling (ABM) is a computational modelling technique used to simulate the actions and interactions of individual agents. Within healthcare research, ABM has proven valuable in modelling the transmission of infectious diseases and evaluating the effectiveness of various non-pharmaceutical interventions (Hoertel et al. 2020; Mahmood et al. 2020; Maziarz and Zach 2020; Tang and Chen 2021). Despite these meaningful progresses, there is still a lack of studies to evaluate the transmission of airborne diseases within hospital built environments. As such, evidence obtained through ABM has not been considered in the assessment of healthcare interventions during the pandemic by the Oxford Centre for Levels of Evidence (OCEBM 2009) or the National Institute for Health and Care Excellence Guidelines (NICE 2014).

3.2. Machine learning techniques: artificial neural networks (ANNs) in transmissible disease modelling

Artificial Neural Networks (ANNs), mimicking the human brain, are increasingly utilised in assessing transmission risks, gaining prominence during the COVID-19 outbreak (Kufel et al. 2023). ANNs have proven effective in offering insightful forecasts, including projections of cumulative COVID-19 cases at a national level (Farooq and Bazaz 2021; Mollalo, Rivera, and Vahedi 2020). Such forecasts provide vital information for planning clinical resource readiness.

Compared to traditional transmissible disease models, Artificial Neural Networks (ANNs) present a distinct

advantage, especially in scenarios with limited data availability, a challenge encountered at the outset of the pandemic. This capability stems from the inherent ability of ANNs to learn complex patterns and relationships within datasets. ANNs function as an information processing method capable of learning and predicting the relationships within data, even when the data are scarce or noisy. This sets it apart from conventional mathematical regression analysis, which primarily concerns understanding relationships between a dependent variable and one or more independent variables. With the ability to learn from incomplete or imperfect data and to generalize to new situations, Artificial Neural Networks (ANNs) present a robust solution for transmissible disease modelling, especially in contexts where data availability is limited (Shaikhina and Khovanova 2017).

In the architecture of ANNs, layers serve as fundamental building blocks responsible for processing and modifying input data. These layers play a crucial role in shaping the network's structure and functionality. Typically, ANNs consist of three main types of layers (Kufel et al. 2023):

- Initial Layer: This layer receives the raw input data and communicates to one or more 'hidden layers'.
- Intermediate Layers (hidden layers): Situated between the input and output layers, these tiers facilitate the processing and acquisition of input data representations.
- Output Layer: The layer responsible for producing the ultimate network output, including the prediction of the response variable.

In the hidden layers, the network acquires patterns and relationships in the input data for making predictions. There can be multiple hidden layers, executing intricate transformations of the input data. Throughout the training process, the network adapts weights and biases to minimise prediction errors.

3.3. Concluding remarks

Based on a review of the literature, a research gap emerges in quantifying indoor environments that influence airborne disease transmission using existing epidemiological models. This challenge arises from the complexity of built environments and the lack of comprehensive and reliable research data (Iranzo and Pérez-González 2021). The latter has become particularly evident at the onset of the pandemic, with limited and sometimes contradictory data available (e.g. as discussed in Section 2.1). In comparison to conventional epidemiological models, Artificial Neural Networks (ANNs) can predict infection risk with high confidence levels, even amidst limited or noisy data. This capability stems from ANNs' inherent ability to discern complex patterns and relationships within datasets (Shaikhina and Khovanova 2017). Agent-based modelling (ABM) represents another unconventional computational modelling technique capable of simulating human motions and interactions within complex indoor conditions (Tang and Chen 2023). Despite meaningful progress in each realm, there remains a lack of studies that combine the strengths of both ANNs and ABM modelling approaches to more effectively simulate airborne disease transmission within indoor built environments.

4. Research objectives and research methodologies

The primary objective of this study is to evaluate the risk of nosocomial airborne infections by analysing various indoor built environments. This research approach involves forecasting infection risks using machine learning models (Section 4.1) which incorporates the building information model data derived from a real case study (Section 4.2). The determined infection risks are then integrated into an agent-based model, as discussed in Section 4.3, allowing for comprehensive simulation of nosocomial infections that accounts for the effects of human motion. This investigation and modelling procedure is schematically shown in Figure 1.

4.1. Machine learning approaches for predicting indoor infection risk

The initial focus of this study was to investigate the feasibility of assessing infection risks associated with airborne diseases by thoroughly considering the conditions of the indoor built environment. To achieve this, two predictive models, namely the linear regression model and the neural network infection model, were developed using the machine learning technique. Beginning with a linear regression model provides a baseline and a straightforward interpretation of relationships between input features and the target variable. This assists in developing a good understanding of the data through the application of machine learning techniques, offering initial insights into the contribution of various features to the prediction. As confidence was built through the linear regression model, the transition to a neural network introduces complexity gradually. Both models were developed through programming using Python version 3.12.0 (Python Software Foundation 2023) within the Google Colaboratory (Google 2022) platform. The first model implements a straightforward linear regression model, whereas the second model adopts a more intricate neural

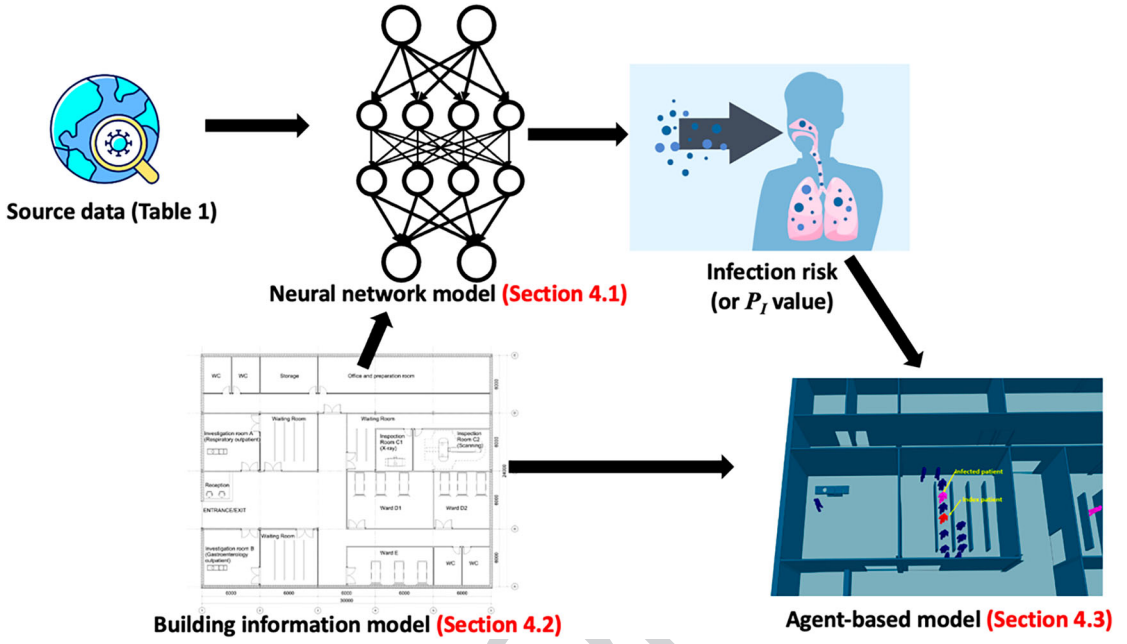


Figure 1. Schematic representation of the epidemiological modelling procedure used in this study.

Q15

network architecture with increased learning rate and epochs. Despite these differences, both models maintain a cohesive structure for predicting infection risk and share the crucial aspect of using the same dataset, outlined in Table 1, for training and testing purposes. This ensures consistency and fair evaluation of model performance. The Python scripts were enhanced using open-source machine-learning libraries specifically designed for data manipulation and analysis in Python. The employed external libraries include Scikit-learn (Cournapeau and Brucher 2023), Pandas (2023) and TensorFlow (Abadi et al. 2016), which have been previously used in data science and healthcare research (Gagliano et al. 2021; Mohammad et al. 2022; Pantic et al. 2023).

4.1.1. Linear regression model

The development of the linear regression model focuses on predicting infection risks through conventional linear regression analysis, utilizing the least squares error method. This implementation was facilitated with the support of external libraries, including Scikit-learn (Cournapeau and Brucher 2023) and Pandas (2023). Pandas facilitates efficient data manipulation, while Scikit-learn provided robust tools for machine learning tasks, including model training and evaluation. All input data were loaded into Python from a CSV file (comma file) created and it incorporates data in Table 1, including room ventilation rate, temperature, humidity, room size and index patient duration and resulting infection risks. A standard 80–20 split strategy (e.g. 80% for training and 20% for testing) has been followed to partition the input data into distinct training and testing sets. Adding the 'seed' parameter in the programme ensures that the split is reproducible, fostering result consistency across multiple runs. This also facilitates robust model evaluation, allowing insights into the model's generalization capabilities beyond the training data. While the training data are utilised for model training purposes, the testing data are employed to evaluate the model's accuracy, quantified by the R_2 or Equation (4).

$$R_2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \quad (4)$$

where n is the sample size, y_i is the actual value of the target variable for observation i, $\widehat{y_i}$ is the actual value of the target variable for a specific instance, and $\bar{y}_i$ denotes the mean of the actual values. R_2 represents the overall goodness of fit of the model based on the test set, but not on a specific prediction. It provides a measure of how well the model's predictions align with the actual values across the entire test set.

4.1.2. Neural network model

The neural network model employs an artificial neural network (ANN) developed through programming using Python version 3.12.0 (2023) within the Google Colaboratory (Google 2022) platform as well. The code was further enhanced with open-source machine learning library TensorFlow (Abadi et al. 2016) which facilitates the construction, training and deployment of neural network models, making it a crucial component in this machine learning project for predicting nosocomial infection risks.

The neural network model's architecture comprises one input layer, three hidden layers and one output layer. The first hidden layer, with 64 neurons, initiates the extraction of complex patterns from the input data, followed by a second layer with 32 neurons and a third layer with 16 neurons, progressively refining the hierarchical representation. These layers collectively constitute the neural network's architecture, with each layer playing a role in learning hierarchical representations from the input data, ultimately predicting the output. Such a layered structure facilitates the neural network's capacity to comprehend and model complex relationships within the data. The datasets were divided following the same 80–20 split strategy (80% for training and 20% for testing) as used in the linear regression model.

To enhance the result interpretability, SHAP (SHapley Additive exPlanations) (Lundberg and Lee 2017) was used in the developed model. SHAP is a Python library that offers insights into how individual features contribute to machine learning model predictions, specifically shedding light on the significance of each feature in predicting infection risks within the scope of this study. The calculated SHAP values provide an effective method for attributing the model's output to each feature. In the context of a hospital infection risk prediction model with a baseline infection rate, a positive SHAP value associated with the total number of infected patients indicates that a higher patient count positively influences an elevated infection risk. As an example, if air exchange rates have a negative SHAP value, it implies that a higher air exchange rate negatively impacts infection risk, resulting in a lower predicted risk. Incorporating SHAP values offers valuable insights into feature importance and individual predictions, proving particularly beneficial in practical applications where understanding and explaining model decisions are essential.

4.2. Case study investigation – a clinical infection unit (CIU)

To investigate the relationship between hospital building layouts and the transmission risks of airborne diseases, a case study investigation of a clinical infection unit (CIU), as seen in Figure 2, was conducted. Figure 2 shows the building layout of a local CIU which delivers specialised services encompassing out-patient clinical diagnosis and in-patient care, comprising various functional units such as consultation rooms, surgeries, ward units, MRI and diagnostic rooms. Patient admission predominantly occurs through Accident & Emergency (A&E) and Urgent Treatment Centres (UTC) referrals. The evaluation of the Clinical Infection Unit (CIU) represents an analysis of a smaller functional unit within a broader and more complex general hospital system. This assessment of the CIU provides a basis for confidently modelling a comprehensive clinical process within the larger hospital, where ongoing research initiatives are currently in progress.

As a rapid response to the COVID-19 outbreak, CIU layout B (Figure 3) was implemented, incorporating a one-way patient traffic system. This design includes two distinct entrances, heightened separation and compartmentalization. Additionally, it facilitates the evaluation of building layouts' influence on airborne disease transmission. The creation of separate patient entrances involved establishing a new patient pathway off the same column gridline (6 m × 6 m) as Layout A (Figure 2). By reconfiguring investigation rooms, waiting areas, and wards, dedicated spaces for suspected or confirmed COVID-19 patients were implemented.

The initial phase of this study involves a systematic examination of presumed indoor environmental factors, including variables such as humidity, temperature and air ventilation rate. The HTM 03–01 standard (DHSC 2021) suggests maintaining a room temperature between 18°C and 22°C in hospital indoor environments, with operating theatres potentially having an elevated room temperature of 25°C. According to the same standard, humidity levels should be kept below 70% to prevent condensation. However, air-change rates (ACH) may vary depending on the facility's purpose: general wards require 6 ACH, infectious disease isolation wards necessitate up to 10 ACH, and operating theatres mandate over 22 ACH (DHSC 2021). In this investigation, a consistent room temperature of 20°C and 60% humidity were employed. Various air-change rates (ACH) ranging from 1 to 6 ACH were considered to examine their influence on the resulting infection rates, utilizing two different prediction models detailed in Section 4.1. This facilitates the estimation of infection risk values (P_I) based on the aforementioned input values. The outcomes are discussed in Chapter 5.

4.3. Impact of building layout on airborne disease transmission: agent-based modelling (ABM)

The infection risk model gives the prediction for the infection risks by considering a variety of indoor conditions, as discussed in previous sections. The infection risks were then incorporated into the agent-based model. The virtual building model of the CIU units (A and B) were first created using Autodesk Revit 2024 (Autodesk 2024) and saved as an IFC file for integration into MassMotion (Oasys 2020), an agent-based modelling (ABM) software. The Agent-based model (ABM) considered all structural and non-structural elements, including furniture,

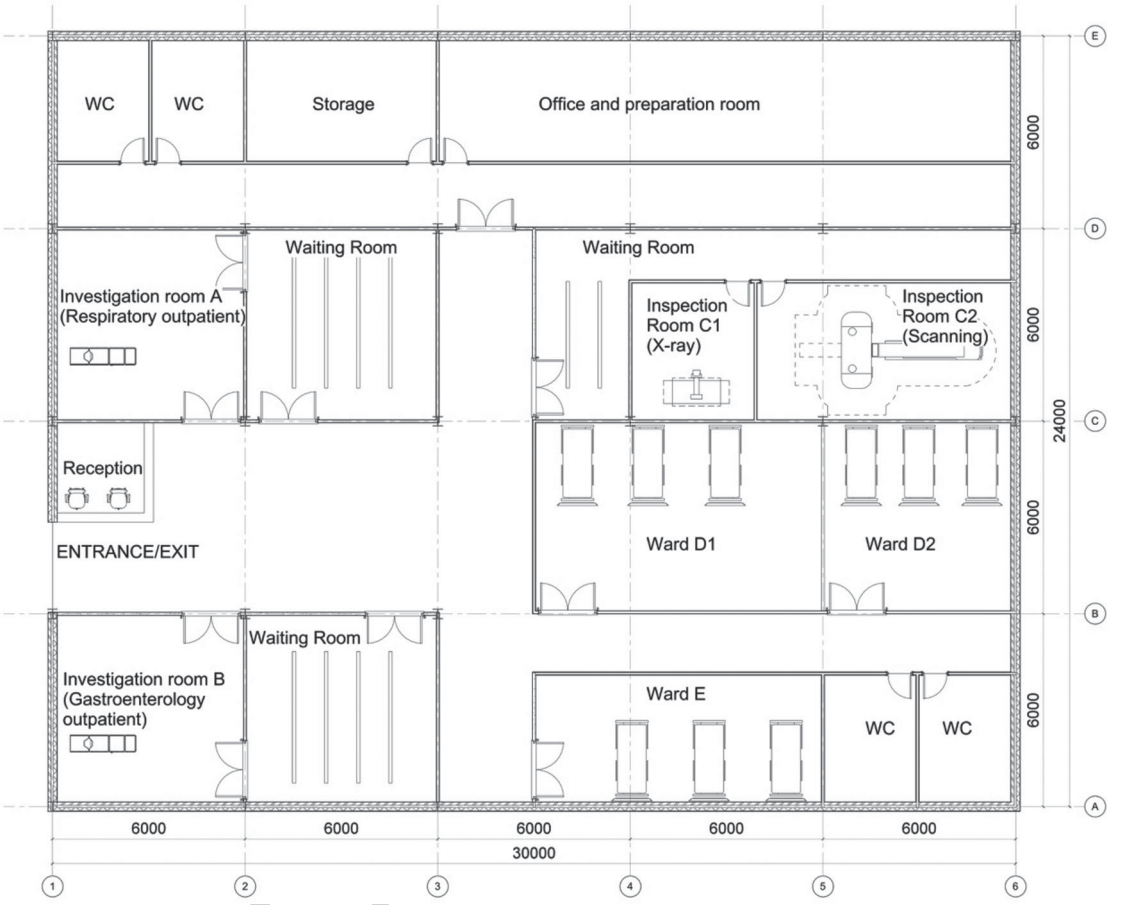


Figure 2. CIU layout – A (all units in mm).

partition walls, and significant medical instruments, since they collectively impact human motion behaviours.

In the ABM analysis, a total of 200 patients (agents) were admitted to the CIU over a 60-minute period, with 40% seeking medical assistance for digestive issues and 60% experiencing respiratory distress syndromes. The clinical process is schematically shown in Figure 4. This simulation did not consider accompanying visitors or any medical and support staff. Patient movement was modelled at an average speed of 1.35 m/sec, with 95% of the distribution falling within 0.25 m/sec standard deviations of the mean speed (Tang and Chen 2023). Each patient's actual speed also relies on agent density in close proximity, with path selection guided by the minimum time cost model (Equation 5) (Kinsey 2015). In the epidemiological analysis, two infected patients (agents) were admitted at specified intervals. The first infected patient, already displaying symptoms of respiratory disease, visited investigation room A after 30 min, while the second patient, with digestive disease symptoms, sought examination in investigation room B at the same time.

$$\text{Agent time cost} = W_D \times \left(\frac{D_G}{V}\right) + W_q \times Q_u + W_L \times L \tag{5}$$

where W_D: Distance weight of each person, D_G: Total distance from the person's present position to the destiny (m), V: Velocity of each person (m/Sec), W_q: Queue weight of each person, Q_u: Expected time in queue before reaching a particular point (Sec), W_L: Geometric component traversal weight, L: Geometric component type cost (Sec).

Airborne transmission risk of SARS-CoV-2 was considered using two distinct prediction models as detailed in Sections 4.1. The outcomes, or the P_I values, were considered in waiting rooms A, B and C – designated for investigation rooms A, B and X-ray/CT scanning rooms, respectively. The entire analysis spanned 120 min in the ABM framework, where patients autonomously determined their pathways following a specified clinical process based on the minimum agent time cost or Equation (5).

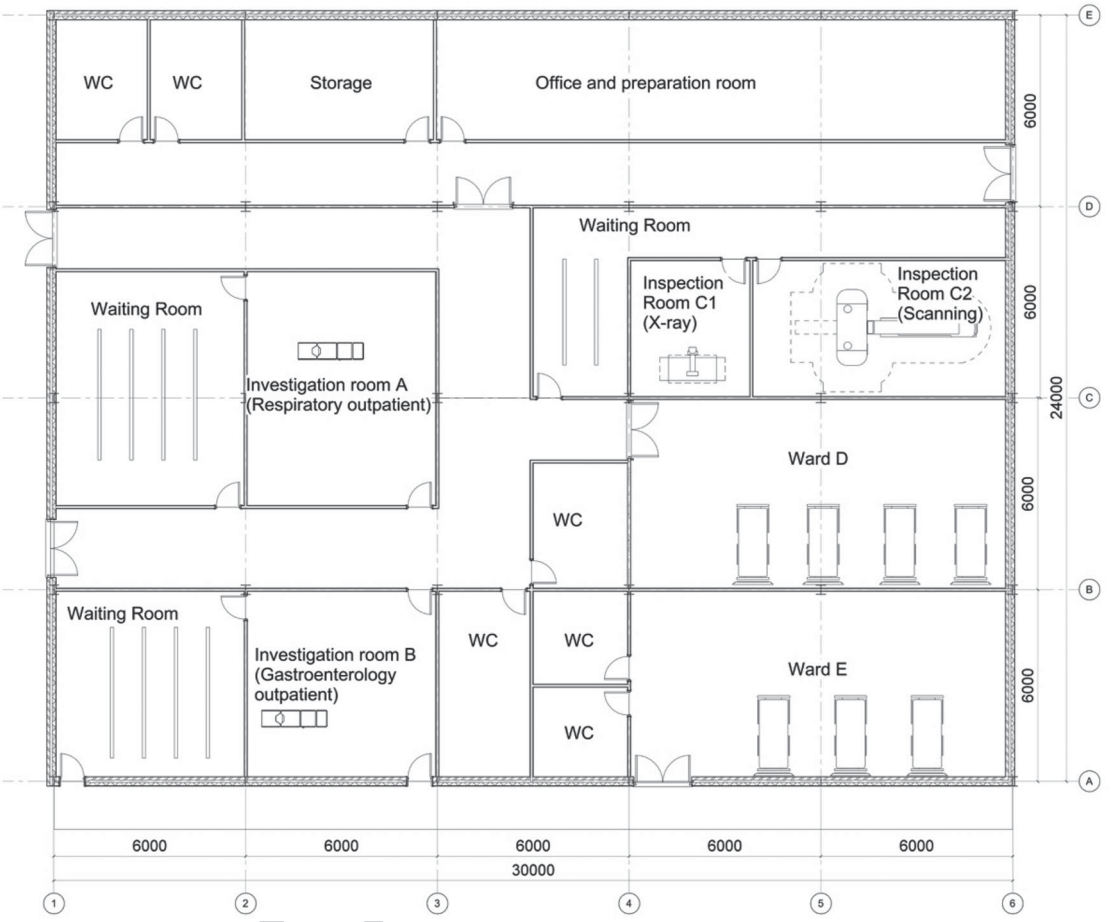
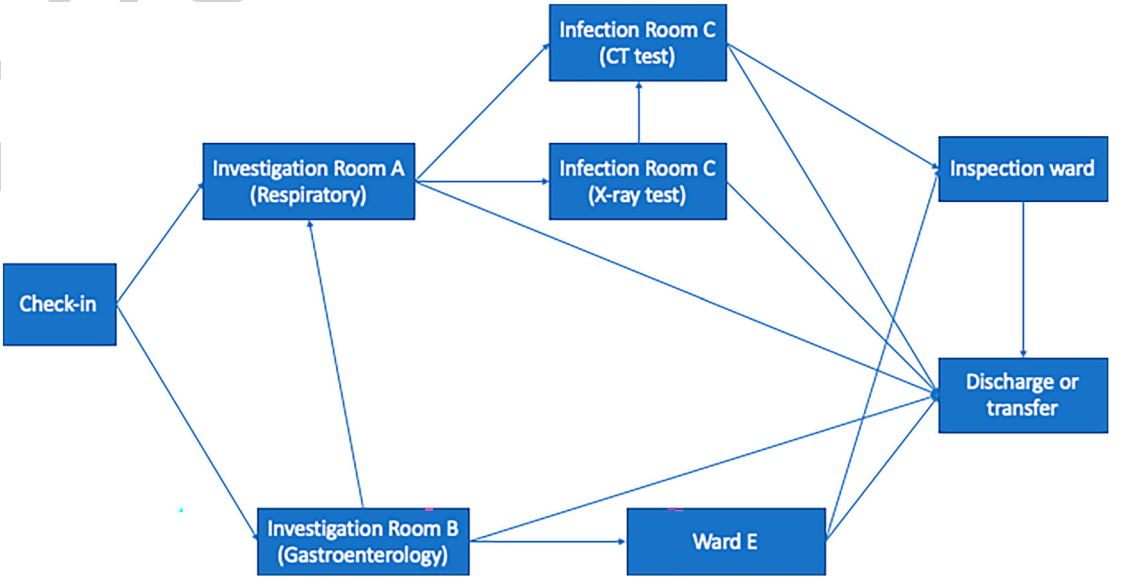


Figure 3. CIU layout – B (all units in mm).



Colour online, B/W in print

Figure 4. Clinical process.

Table 2. Predicted infection risk (P_I) – CIU layout A (linear regression results).

P_I (%)	Ip (m^3/hr)	V (m^3)	Q (ACH)	T (min)	Temperature (°C)	Humidity (%)
4.5	0.52	108 (room A)	1	40.5	20	60
3.4	0.52	108 (room A)	3	40.5	20	60
1.5	0.52	108 (room A)	6	40.5	20	60
3.2	0.52	108 (room B)	1	26.3	20	60
2.0	0.52	108 (room B)	3	26.3	20	60
0.2	0.52	108 (room B)	6	26.3	20	60
1.0	0.52	112 (room C)	1	2.5	20	60
0.0	0.52	112 (room C)	3	2.5	20	60
0.0	0.52	112 (room C)	6	2.5	20	60

To explore the impact of building layouts on airborne disease transmission, both CIU layout A and B were taken into account in the epidemiological analysis.

5. Results and discussion

5.1. Prediction for the infection risk

5.1.1. Linear regression model results

All input data for the linear regression were loaded from a CSV file (comma-separated values), detailed in Table 1. This file includes information on room ventilation rate, temperature, humidity, room size, index patient duration, and the resulting infection risks. A linear regression analysis was then conducted using Python, and the derived equation for the linear regression is presented as Equation (6). The infection risk predictions for CIU layouts A and B are presented in Tables 2 and 3, utilizing the linear regression model discussed in Section 4.1.1. The infectivity of a single COVID-19 patient, represented by the I_p value, was set at 0.52 m^3/hr according to published data (Aganovic et al. 2021).

$$P_I = 4.6058 - 1.4401I_p + 0.0002V - 0.6042Q + 0.0912T - 0.0411H \quad (6)$$

Equation (6) suggests a decrease in infection risk with an increased respiratory ventilation rate (Ip) of the infected patient. The infection risks tend to rise with an extended duration of time (T) while decreasing with higher room air exchange rates (Q) and humidity (H), aligning with the observations in the dataset presented in Table 1. The R_2 serves as a statistical metric illustrating the fraction of the variance in the prediction, as detailed in Section 4.1.1. For this linear regression model, R_2 was found to be 0.91, or approximately 91% of the variability in the testing data is explained or captured by the model. This implies that the model performs well in explaining the variance in the infection risk predictions based on the selected features.

Although distinct in nature, both the linear regression model and the neural network model, as discussed in the following section, offer unique insights to the analytical process. Specifically, the interpretative framework provided by the linear regression model illuminates fundamental relationships within the data, laying the groundwork for the subsequent implementation of neural network models. For example, the linear coefficient obtained from regression analysis plays a similar role to SHAP (SHapley Additive exPlanations) values in neural network modelling, providing insights into variable impacts. It is noteworthy that despite sharing a common dataset, the identification of differing influential factors by each method underscores the complementary nature of their contributions.

5.1.2. Neural network model results

The preceding section, focused on a linear regression model, offers a clear interpretation of the relationships between input characteristics and output values using machine learning techniques. As confidence grows with the linear regression model, the transition to a neural network becomes crucial, as it has the ability to better capture nonlinear relationships within the data. This becomes particularly advantageous when dealing with intricate patterns of transmission risks (Kufel et al. 2023). The flexibility of the neural network model also allows for easy adjustments to the model architecture and learning rate, enhancing adaptability to diverse problem complexities. The infection risk predictions for CIU layouts A and B, corresponding to the input values discussed in Section 4.1 and 4.2, are presented in Tables 4 and 5. For the neural network model, R_2 was found to be 0.86, or approximately 86% of the variability in the testing data is explained or captured by the model. This implies that the model still performs well in explaining the variance in the infection risk predictions based on the selected features.

Figure 5 shows that with an increase in Q (ACH) values from 1 to 6, the corresponding SHAP values show a more negative trend, decreasing from − 0.06 to − 5.75. This implies that elevated Q (ACH) values have a negative influence on the model's predictions for infection risk values, aligning with the observations in the dataset presented in Table 1. Comparable trends are noted in the case of the total exposure time (T), as seen in Figure 6. When the air exchange rate (Q) doubled from 3 to 6

Table 3. Predicted infection risk (P_I) – CIU layout B (linear regression results).

P_I (%)	Ip(m^3/hr)	V(m^3)	Q(ACH)	T(min)	Temperature(°C)	Humidity(%)
4.5	0.52	133 (room A)	1	40.6	20	60
3.3	0.52	133 (room A)	3	40.6	20	60
1.5	0.52	133 (room A)	6	40.6	20	60
1.8	0.52	108 (room B)	1	11	20	60
0.6	0.52	108 (room B)	3	11	20	60
0.0	0.52	108 (room B)	6	11	20	60
1.0	0.52	112 (room C)	1	2.5	20	60
0.0	0.52	112 (room C)	3	2.5	20	60
0.0	0.52	112 (room C)	6	2.5	20	60

Table 4. Predicted infection risk (P_I) – CIU layout A (neural network model results).

P_I (%)	Ip (m^3/hr)	V (m^3)	Q (ACH)	T (min)	Temperature (°C)	Humidity (%)
7.7	0.52	108 (room A)	1	40.5	20	60
6.6	0.52	108 (room A)	3	40.5	20	60
5.4	0.52	108 (room A)	6	40.5	20	60
7.5	0.52	108 (room B)	1	26.3	20	60
6.3	0.52	108 (room B)	3	26.3	20	60
5.1	0.52	108 (room B)	6	26.3	20	60
7.1	0.52	112 (room C)	1	2.5	20	60
6.1	0.52	112 (room C)	3	2.5	20	60
4.8	0.52	112 (room C)	6	2.5	20	60

Table 5. Predicted infection risk (P_I) – CIU layout B (neural network model results).

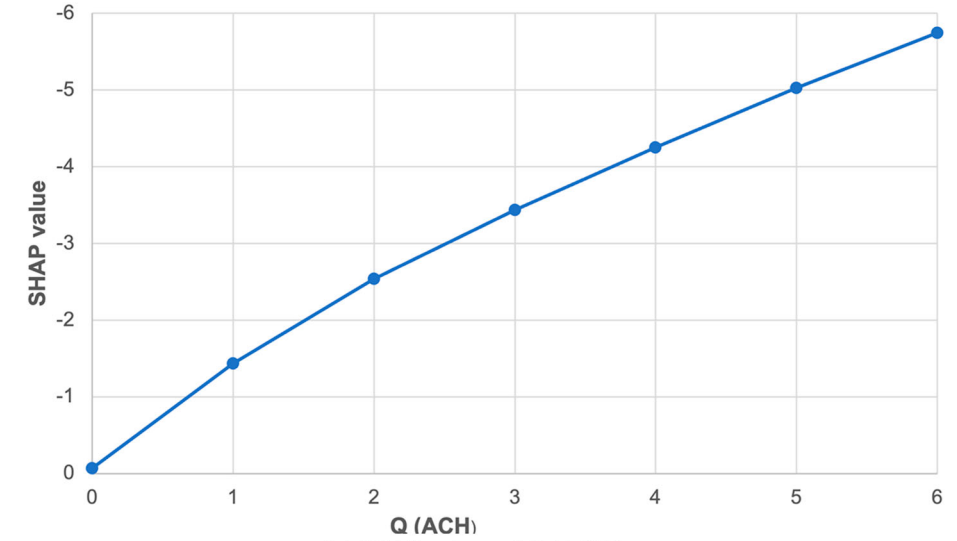
P_I (%)	Ip (m^3/hr)	V (m^3)	Q (ACH)	T (min)	Temperature (°C)	Humidity (%)
7.1	0.52	133 (room A)	1	40.6	20	60
5.4	0.52	133 (room A)	3	40.6	20	60
4.6	0.52	133 (room A)	6	40.6	20	60
7.0	0.52	108 (room B)	1	11	20	60
5.5	0.52	108 (room B)	3	11	20	60
4.8	0.52	108 (room B)	6	11	20	60
6.5	0.52	112 (room C)	1	2.5	20	60
5.3	0.52	112 (room C)	3	2.5	20	60
4.6	0.52	112 (room C)	6	2.5	20	60

ACH, the absolute value of the SHAP values increased by 66%. In contrast, when the exposure time (T) also doubled from 60 to 120 min, the SHAP values increased by 134%. This indicates that exposure time has a greater impact on the predicted infection risk values compared to the air exchange rate. Utilizing the SHAP explainer enhances interpretability by elucidating the contribution of each feature to predictions. With the potential for improved accuracy in capturing intricate relationships, this neural network model serves as a good alternative to linear regression model.

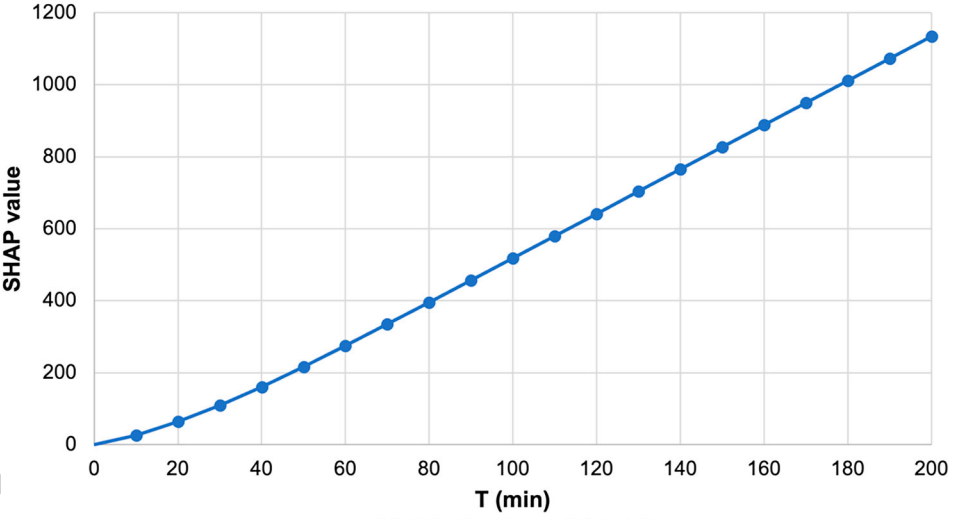
5.2. Case study results: impact of building layout on airborne disease transmission

The effectiveness of various hospital building layouts (A and B) was evaluated using MassMotion (Oasys 2020), which utilized infection risk data presented in Tables 4 and 5. The nosocomial or herein long-range infections are visually demonstrated in Figure 6(a). The number of nosocomial infections are shown in Tables 6 and 7 based on different air ventilation rates (Q). In CIU layout A, with 2 admitted COVID-19 patients among 198 suspected patients, 14 individuals contracted the disease with a low ventilation rate of 1 ACH, resulting in a 7% infection rate. Notably, 10 cases occurred in waiting room A, and 4 in waiting room B (refer to Figure 2). Given the transmission potential of SARS-CoV-2 through droplets and aerosols, the proximity between individuals becomes a crucial factor affecting the risk of transmission. With enhanced ventilation at 6 ACH, 10 individuals contracted the disease, demonstrating the effective control of the spread of this airborne-transmitted disease through improved ventilation. In CIU layout B, the optimized patient pathways and enhanced workflows contribute to a lower risk of hospital-related infections in contrast to the original design A. When 2 COVID-19 patients were admitted, 10 out of 198 patients met the criteria for nosocomial infections, indicating a reduced infection rate of 5% at a ventilation rate of 1 ACH. These cases comprise 9 infected patients in waiting room A, and 1 patient in waiting room B. With enhanced ventilation at 6 ACH, 8 individuals contracted the disease, demonstrating the effective control of the spread of this airborne-transmitted disease through improved ventilation. This outcome further highlights the combined effect of enhanced building layout

Colour online, B/W in print



(a) SHAP values of Q (ACH)



(b) SHAP values of T (min)

Figure 5. SHAP values of Q (ACH) and T (min). (a) SHAP values of Q (ACH); (b) SHAP values of T (min).

Table 6. Total infections – CIU layout A.

Q(ACH)	Room A	Room B	Room C	Total
1	10	4	0	14
3	7	4	0	11
6	6	4	0	10

Table 7. Total infections – CIU layout B.

Q(ACH)	Room A	Room B	Room C	Total
1	9	1	0	10
3	7	1	0	8
6	7	1	0	8

and ventilation in effectively controlling the spread of airborne diseases.

In addition to the long-range nosocomial infection investigation discussed above, MassMotion (Oasys 2020) also enables short-range congestion analysis, which measures the maximum number of people within a specific area during peak times. This helps to identify areas where the risk of overcrowding is high. In this study, a congestion analysis within a 2 m radius (Section 2.1.1) was conducted to evaluate the short-range infection risk, offering valuable insights into the transmission risk of infectious diseases through droplets. In CIU layout A, Figure 6 (b) shows up to 4 patients within a 2 m radius in the corridor during peak hours, suggesting a violation of the 2 m social distancing requirement during the pandemic and posing a potential risk for nosocomial transmission of the disease. CIU design B incorporates a well-insulated reception area. Figure 6 (c) illustrates that the updated hospital building layout exhibits improved reduction of corridor congestion compared to the initial design in building layout A (Figure 6 (b)). This was achieved through a decrease in the number of patients per square meter.

Colour online, B/W in print

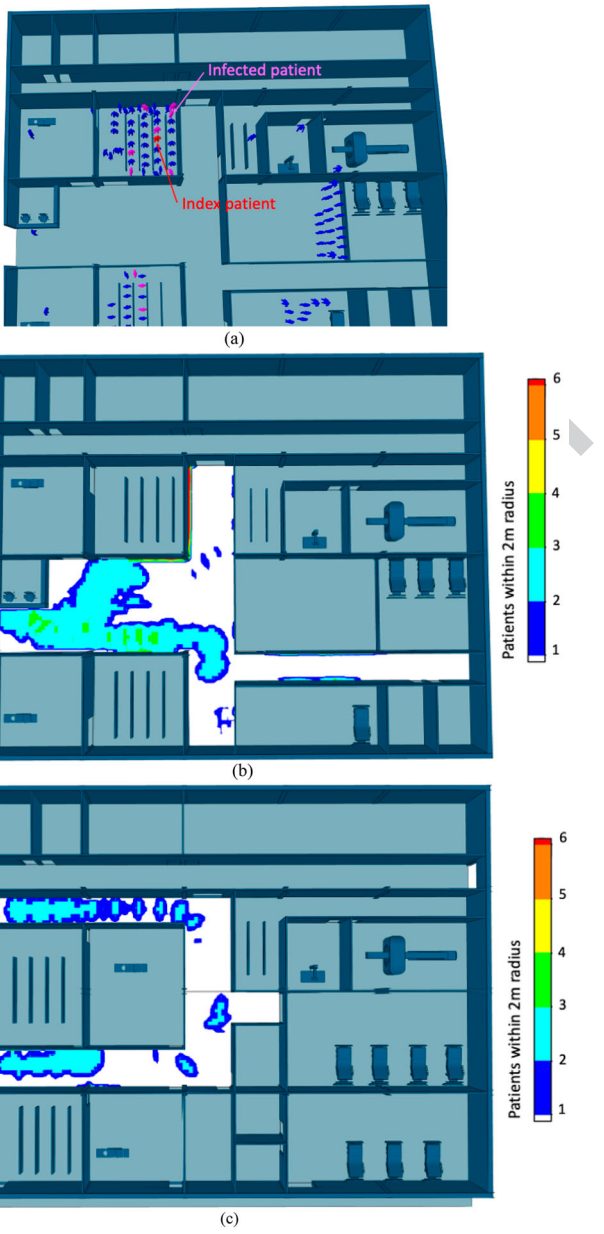


Figure 6. Agent-based modelling results (1 ACH). (a) Visualized result (building layout A); (b) Maximum number of patients within 2 m radius (building layout A); (c) Maximum number of patients within 2 m radius (building layout B).

5.3. Advancing infection control through flexible healthcare facility design using modern construction methods – future-proof hospitals?

The case study result demonstrates the potential for improved control over the transmission of airborne diseases such as COVID-19 through adjustments to hospital building layouts. However, realizing such flexibilities in the real life can be a challenge. Notably, significant alterations to building layouts may face constraints due to existing column gridlines and load-bearing structural elements like columns and walls (Tang and Chen 2021). Integrating modular design and scalable infrastructure into healthcare facilities may enhance their adaptability and flexibility, enabling them to effectively address future challenges like outbreaks and evolving patient demographics. A notable example was the George Eliot Hospital extension project, planned and constructed amid the pandemic (Figure 7). The new development involves a new 30-bed ward comprising over 30 modular units with integrated medical facilities, medical gas piping systems, access controls, fire escape ramps and nurses' stations. The utilization of prefabricated modular units not only reduced the construction timeline from 20 to 14 weeks, but also enabled rapid assembly and reconfiguration of hospital spaces. By adopting these innovative design and construction methodologies, healthcare

Colour online, B/W in print

(a) Modular units in construction

(b) The completed ward

Figure 7. George Eliot Hospital's new 30-bed ward (images used with kind permission from Wernick Buildings). (a) Modular units in construction; (b) The completed ward.

facilities can strengthen their resilience by adjusting layouts as demand changes. This will maximise space efficiency and ensure prompt and efficient care for patients during periods of heightened demand or unexpected events. This also underscores the vision of a future-proof hospital design process.

6. Conclusion

The evaluation of infection risk predictions based on machine learning techniques can provide valuable insights into the complex dynamics of nosocomial infections in healthcare settings. The linear regression model demonstrated a commendable explanatory power with an R_2 value of 91%, offering a robust understanding of infection risk variations based on selected features. The neural network model demonstrated its strength in effectively capturing intricate non-linear patterns within the data. With an R_2value of 86%, the neural network model maintained a high level of accuracy in explaining the variability in infection risk predictions. The SHAP analysis further revealed the model's sensitivity to factors such as air exchange rates and exposure time, providing valuable information regarding the impact of these features on infection risk.

The case study investigation explored the practical implications of these models in assessing the impact of building layouts on airborne disease transmission. The results highlighted the significance of ventilation rates and optimized building designs in controlling nosocomial infections. Notably, the combination of enhanced building layouts and improved ventilation demonstrated a substantial reduction in infection rates, showcasing the synergy between design flexibility and effective computer simulations in healthcare infrastructure planning.

7. Research limitations and future work

This section addresses two major research limitations in the study. Firstly, enhancing the diversity and comprehensiveness of the collected data (e.g. Table 1) is crucial for improving the accuracy of the neural network model. Stressing the significance of obtaining more data is pivotal, as expanding the dataset size often leads to enhanced modelling performance.

Secondly, viral shedding exhibits significant diversity in terms of viral load, measured in RNA/ml. A prior study revealed that approximately 2% of COVID-19 patients could carry nearly 90% of the circulating virions within a community, indicating the presence of supercarriers (Yang et al. 2021). The notable diversity observed in COVID-19 cases, coupled with a substantial proportion of subclinical and asymptomatic individuals, poses significant challenges in effectively managing the disease's spread. This underscores the need for future research in modelling airborne disease transmissions.

Acknowledgements

The author expresses sincere gratitude to Xiangming Zhou, Tony Nichol, Alex Johnson, Bing Chen and Andy King for their valuable support. Additionally, the author extends heartfelt appreciation to his family members, Lulu, Fanyi, and Alice, for their unwavering emotional support throughout the process.

Disclosure statement

No potential conflict of interest was reported by the author(s). Q3

ORCID

Kangkang Tang http://orcid.org/0000-0002-9289-937X

References Q4

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. "TensorFlow: A System for Large-Scale Machine Learning." In *Proceedings*

of the 12th USENIX Conference on Operating Systems Design and Implementation, 265–283. Savannah, GA, USA: USENIX Association.

Aganovic, Amar, Yang Bi, Guangyu Cao, Finn Drangsholt, Jarek Kurnitski, and Pawel Wargocki. 2021. "Estimating the Impact of Indoor Relative Humidity on SARS-CoV-2 Airborne Transmission Risk Using a new Modification of the Wells-Riley Model." *Building and Environment* 205:108278. https://doi.org/10.1016/j.buildenv.2021.108278.

Autodesk. 2024. *Autodesk Revit 2024.* Autodesk.

Buonanno, G., L. Morawska, and L. Stabile. 2020a. "Quantitative Assessment of the Risk of Airborne Transmission of SARS-CoV-2 Infection: Prospective and Retrospective Applications." *Environment International* 145:106112. https://doi.org/10.1016/j.envint.2020.106112.

Buonanno, G., L. Stabile, and L. Morawska. 2020b. "Estimation of Airborne Viral Emission: Quanta Emission Rate of SARS-CoV-2 for Infection Risk Assessment." *Environment International* 141:105794. https://doi.org/10.1016/j.envint.2020.105794.

Campbell, Denis, and Caelainn Barr. 2021. "40,600 People Likely Caught Covid While Hospital Inpatients in England." *The Guardian*. London, UK.

CDC. 2021. "Science Brief: SARS-CoV-2 and Surface (Fomite) Transmission for Indoor Community Environments." Centers for Disease Control and Prevention, Accessed 20/5/2023. https://www.cdc.gov/coronavirus/2019-ncov/more/science-and-research/surface-transmission.html#ref7.

Cournapeau, David, and Matthieu Brucher. 2023. *Scikit-learn 1.3.2*. Scikit-Learn Development Team.

Currie, Christine S. M., John W. Fowler, Kathy Kotiadis, Thomas Monks, Bhakti Stephan Onggo, Duncan A. Robertson, and Antuela A. Tako. 2020. "How Simulation Modelling Can Help Reduce the Impact of COVID-19." *Journal of Simulation* 14 (2): 83–97. https://doi.org/10.1080/17477778.2020.1751570.

Dabisch, Paul, Michael Schuit, Artemas Herzog, Katie Beck, Stewart Wood, Melissa Krause, David Miller, et al. 2021. "The Influence of Temperature, Humidity, and Simulated Sunlight on the Infectivity of SARS-CoV-2 in Aerosols." *Aerosol Science and Technology* 55 (2): 142–153. https://doi.org/10.1080/02786826.2020.1829536.

Daphne, Duval, C. Palmer Jennifer, Tudge Isobel, Pearce-Smith Nicola, O'Connell Emer, Bennett Allan, and Clark Rachel. 2022. "Long Distance Airborne Transmission of SARS-CoV-2: Rapid Systematic Review." *BMJ* 377:e068743. https://doi.org/10.1136/bmj-2021-068743.

de Oliveira, P. M., L. C. C. Mesquita, S. Gkantonas, A. Giusti, and E. Mastorakos. 2021. "Evolution of Spray and Aerosol from Respiratory Releases: Theoretical Estimates for Insight on Viral Transmission." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477 (2245):20200584. https://doi.org/10.1098/rspa.2020.0584.

DHSC. 2021. *HTM 03-01 Specialised ventilation for healthcare buildings: Part A - Design and Validation*: Department of Health and Social Care.

Evans, Stephanie, Emily Agnew, Emilia Vynnycky, and Julie V. Robotham. 2020. "The Impact of Testing and Infection Prevention and Control Strategies on Within-Hospital Transmission Dynamics of COVID-19 in English Hospitals." *medRxiv*. https://doi.org/10.1101/2020.05.12.20095562

Farooq, J., and M. A. Bazaz. 2021. A Deep Learning Algorithm for Modeling and Forecasting of COVID-19 in Five Worst Affected States of India." *Alexandria Engineering Journal* 60 (1), 587–596. https://doi.org/10.1016/j.aej.2020.09.037.

Gagliano, A., M. Puligheddu, N. Ronzano, P. Congiu, M. G. Tanca, I. Cursio, S. Carucci, S. Sotgiu, E. Grossi, and A. Zuddas. 2021. "Artificial Neural Networks Analysis of Polysomnographic and Clinical Features in Pediatric Acute-Onset Neuropsychiatric Syndrome (PANS): From Sleep Alteration to "Brain Fog." *Nature and Science of Sleep* 13:1209–1224. https://doi.org/10.2147/NSS.S300818.

Geng, Yansheng, and Youchun Wang. 2023. "Stability and Transmissibility of SARS-CoV-2 in the Environment." *Journal of Medical Virology* 95 (1): e28103. https://doi.org/10.1002/jmv.28103.

Google. 2022. "Colaboratory." Google, Accessed 3/12/2023. https://colab.research.google.com/.

Greenhalgh, Trisha, Jose L. Jimenez, Kimberly A. Prather, Zeynep Tufekci, David Fisman, and Robert Schooley. 2021. "Ten Scientific Reasons in Support of Airborne Transmission of SARS-CoV-2." *The Lancet* 397 (10285): 1603–1605. https://doi.org/10.1016/S0140-6736(21)00869-2.

Hirose, Ryohei, Hiroshi Ikegaya, Yuji Naito, Naoto Watanabe, Takuma Yoshida, Risa Bandou, Tomo Daidoji, Yoshito Itoh, and Takaaki Nakaya. 2021. "Survival of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and Influenza Virus on Human Skin: Importance of Hand Hygiene in Coronavirus Disease 2019 (COVID-19)." *Clinical Infectious Diseases* 73 (11):e4329-e35. https://doi.org/10.1093/cid/ciaa1517.

Hoertel, Nicolas, Martin Blachier, Carlos Blanco, Mark Olfson, Marc Massetti, Marina Sánchez Rico, Frédéric Limosin, and Henri Leleu. 2020. "A Stochastic Agent-Based Model of the SARS-CoV-2 Epidemic in France." *Nature Medicine* 26 (9): 1417–1421. https://doi.org/10.1038/s41591-020-1001-6.

Iranzo, V., and S. Pérez-González. 2021. "Epidemiological Models and COVID-19: A Comparative View." *History and Philosophy of the Life Sciences* 43 (3): 104. https://doi.org/10.1007/s40656-021-00457-9.

Kinsey, Michael. 2015. *The Verification and Validation of Mass-Motion for Evacuation Modelling*. London, UK: Ove Arup & Partners Ltd.

Kufel, J., K. Bargieł-Łączek, S. Kocot, M. Koźlik, W. Bartnikowska, M. Janik, Ł. Czogalik, et al. 2023. "What Is Machine Learning, Artificial Neural Networks and Deep Learning?-Examples of Practical Applications in Medicine." *Diagnostics (Basel)* 13 (15). https://doi.org/10.3390/diagnostics13152582.

Lundberg, Scott, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." Paper presented at the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach California USA.

Mahmood, Imran, Hamid Arabnejad, Diana Suleimenova, Isabel Sassoon, Alaa Marshan, Alan Serrano-Rico, Panos Louvieris, et al. 2020. "FACS: A Geospatial Agent-Based Simulator for Analysing COVID-19 Spread and Public Health Measures on Local Regions." *Journal of Simulation*, 1–19. https://doi.org/10.1080/17477778.2020.1800422.

Marsden, Andrew G. 2003. "Outbreak of Influenza-Like Illness Related to air Travel." *Medical Journal of Australia* 179 (3):172-173. https://doi.org/10.5694/j.1326-5377.2003.tb05483.x.

Maziarz, Mariusz, and Martin Zach. 2020. "Agent-based Modelling for SARS-CoV-2 Epidemic Prediction and Intervention

Assessment: A Methodological Appraisal." *Journal of Evaluation in Clinical Practice* 26 (5): 1352–1360. https://doi.org/10.1111/jep.13459.

Mohammad, Moman A., Kevin K. W. Olesen, Sasha Koul, Chris P. Gale, Rebecca Rylance, Tomas Jernberg, Tomasz Baron, et al. 2022. "Development and Validation of an Artificial Neural Network Algorithm to Predict Mortality and Admission to Hospital for Heart Failure After Myocardial Infarction: A Nationwide Population-Based Study." *The Lancet Digital Health* 4 (1): e37–e45. https://doi.org/10.1016/S2589-7500(21)00228-4.

Mollalo, A., K. M. Rivera, and B. Vahedi. 2020. "Artificial Neural Network Modeling of Novel Coronavirus (COVID-19) Incidence Rates Across the Continental United States." *International Journal of Environmental Research and Public Health* 17 (12), https://doi.org/10.3390/ijerph17124204.

Morawska, Lidia, and Donald K. Milton. 2020. "It is Time to Address Airborne Transmission of COVID-19." *Clinical Infectious Diseases*, https://doi.org/10.1093/cid/ciaa939.

NICE. 2014. "Developing NICE Guidelines: The Manual." National Institute for Health and Care Excellence, Accessed 1/11/2022. https://www.nice.org.uk/process/pmg20/chapter/introduction.

Oasys. 2020. *MassMotion*. London, UK: Arup.

OCEBM. 2009. "The Oxford Levels of Evidence 2." Oxford Centre for Evidence-Based Medicine, Accessed 1/11/2022. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence.

Pandas development team. 2023. "pandas-dev/pandas: Pandas 2.1.2".

Pantic, Igor, Jovana Paunovic, Jelena Cumic, Svetlana Valjarevic, Georg A. Petroianu, and Peter R. Corridon. 2023. "Artificial Neural Networks in Contemporary Toxicology Research." *Chemico-Biological Interactions* 369:110269. https://doi.org/10.1016/j.cbi.2022.110269.

Python Software Foundation. 2023. Python (3.12.0).".

Qureshi, Zeshan, Nicholas Jones, Robert Temple, Jessica PJ Larwood, Trisha Greenhalgh, and Lydia Bourouiba. 2020. "What is the Evidence to Support the 2-metre Social Distancing rule to Reduce COVID-19 transmission?" The Centre for Evidence-Based Medicine, Accessed 1/5/2023. https://www.cebm.net/covid-19/what-is-the-evidence-to-support-the-2-metre-social-distancing-rule-to-reduce-covid-19-transmission/.

Read, Jonathan M., Chris A. Green, Ewen M. Harrison, Annemarie B. Docherty, Sebastian Funk, Janet Harrison, Michelle Girvan, et al. 2021. "Hospital-acquired SARS-CoV-2 Infection in the UK's First COVID-19 Pandemic Wave." *The Lancet* 398 (10305):1037-1038. https://doi.org/10.1016/S0140-6736(21)01786-4.

Rickman, Hannah M, Tommy Rampling, Karen Shaw, Gema Martinez-Garcia, Leila Hail, Pietro Coen, Maryam Shahmanesh, Gee YenShin, EleniNastouli, and Catherine FHoulihan. 2020. "Nosocomial Transmission of Coronavirus Disease 2019: A Retrospective Study of 66 Hospital-Acquired Cases in a London Teaching Hospital." *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa816.

Riley, E. C., G. Murphy, and R. L. Riley. 1978. "Airborne Spread of Measles in a Suburban Elementary School." *American Journal of Epidemiology* 107 (5): 421–432. https://doi.org/10.1093/oxfordjournals.aje.a112560.

Saran, S., M. Gurjar, A. Baronia, V. Sivapurapu, P. S. Ghosh, G. M. Raju, and I. Maurya. 2020. "Heating, Ventilation and air Conditioning (HVAC) in Intensive Care Unit." *Critical Care* 24 (1): 194. https://doi.org/10.1186/s13054-020-02907-5.

Schijven, Jack, C. Vermeulen Lucie, Arno Swart, Adam Meijer, Erwin Duizer, and Maria de Roda Husman Ana. 2021. "Quantitative Microbial Risk Assessment for Airborne Transmission of SARS-CoV-2 via Breathing, Speaking, Singing, Coughing, and Sneezing." *Environmental Health Perspectives* 129 (4):047002. https://doi.org/10.1289/EHP7886

Shaikhina, Torgyn, and Natalia A. Khovanova. 2017. "Handling Limited Datasets with Neural Networks in Medical Applications: A Small-Data Approach." *Artificial Intelligence in Medicine* 75:51–63. https://doi.org/10.1016/j.artmed.2016.12.003.

Swallow, B., P. Birrell, J. Blake, M. Burgman, P. Challenor, L. E. Coffeng, P. Dawid, et al. 2022. "Challenges in Estimation, Uncertainty Quantification and Elicitation for Pandemic Modelling." *Epidemics* 38:100547. https://doi.org/10.1016/j.epidem.2022.100547.

Tang, Kangkang, and Bing Chen. 2021. "How to Develop More Resilient Hospitals Through Agent-Based Modelling." *Proceedings of the Institution of Civil Engineers - Civil Engineering* 175 (1):27–32. https://doi.org/10.1680/jcien.21.00049.

Tang, Kangkang, and Bing Chen. 2023. "Resilient Hospital Design: From Crimean War to COVID-19." *HERD: Health Environments Research & Design Journal*: 1–20. https://doi.org/10.1177/19375867231174238.

Ulrich, Lorenz, Nico Joel Halwe, Adriano Taddeo, Nadine Ebert, Jacob Schön, Christelle Devisme, Bettina Salome Trüeb, et al. 2022. "Enhanced Fitness of SARS-CoV-2 Variant of Concern Alpha but not Beta." *Nature* 602 (7896): 307–313. https://doi.org/10.1038/s41586-021-04342-0.

V'kovski, Philip, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. 2021. "Coronavirus Biology and Replication: Implications for SARS-CoV-2." *Nature Reviews Microbiology* 19 (3): 155–170. https://doi.org/10.1038/s41579-020-00468-6.

WHO. 2020. "Modes of Transmission of Virus Causing COVID-19: Implications for IPC Precaution Recommendations." World Health Organlization, Accessed 1/5/2020. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

Yang, Qing, Tassa K. Saldi, Patrick K. Gonzales, Erika Lasda, Carolyn J. Decker, Kimngan L. Tat, Morgan R. Fink, et al. 2021. "Just 2% of SARS-CoV-2—Positive Individuals Carry 90% of the Virus Circulating in Communities." *Proceedings of the National Academy of Sciences* 118 (21): e2104547118. https://doi.org/10.1073/pnas.2104547118.