

## Agent Based Decision Support System for Identifying the Spread of Nosocomial Infections in a Rural Hospital

Nisha Wilvicta.J\*, E.Grace Mary Kanaga\*\*

\*(Department of Computer Science, Karunya University, Coimbatore-114)

\*\* (Department of Computer Science, Karunya University, Coimbatore-114)

### ABSTRACT

Nosocomial infections are prevalent throughout the hospitals across the world. Only proper identification of the disease spread vectors would help to take proper mitigation strategies to curb the disease in the correct time. These nosocomial infections are more prevalent in the rural areas due to the improper sanitation and improper hygiene. Studying the various transmission vectors would help to prevent the spread of nosocomial infections in a rural hospital. Agent Based Modeling has been used to examine the spread of nosocomial infections in a rural hospital so that proper mitigation strategies can be undertaken. The aim of this project is to identify the various transmission vectors for the spread of nosocomial infections and classify them using Iterative Dichotomizer 3 (ID3). From the decision tree the infected patients within close contact has been identified as the major transmission vector.

*Keywords* – Agent Based Modeling, Decision Support, Decision Tree, ID3 Algorithm, Nosocomial Infections

### I. Introduction

This paper presents an Agent Based Modeling (ABM) framework that is used to simulate the spread of nosocomial infections in a rural hospital. Nosocomial infections are considered as a major reason for the increase in patient mortality especially in rural hospitals. Proper identification of the transmission vectors would help to curb the spread of the infections at an early stage itself.

Various modeling techniques are available to simulate the spread of nosocomial infections [1]. Healthcare models are primarily used to model the healthcare systems. The healthcare models can be used to identify disease transmission vectors at an early stage and thereby reduce the mortality rate. The primary objective of healthcare models is to reduce the patient mortality and morbidity rate by providing quality health care to the patients. The modeling techniques that are available are mathematical models[2], cellular-automata based models [3], network-based models[3], agent-based models[4] and individual-based models[5]. Of all the modeling techniques, ABM has been used as the modeling technique due to the dynamic and

flexible nature of the agents [6]. The results that are obtained using the ABM has been validated using the decision tree classification algorithm.

The transmission vectors are identified and classified using the ID3 decision tree classification algorithm in a more accurate manner. The different attributes are specified along with the different values that they may take. Based on the various attribute values the decision tree is constructed. Decision Tree algorithms work based on the processing and deciding on the various attribute values [7]. The decision tree acts as a decision support system that enables the healthcare practitioners to take the necessary steps to prevent the spread of nosocomial infections.

The following sections describe the agent based modeling framework that has been used to simulate the spread of nosocomial infections. The decision tree classification algorithm that has been used to identify the various transmission vectors for the spread of nosocomial infections have been also described in the following sections.

### II. ABM Framework for Rural Hospital

An ABM framework was developed to simulate the spread of nosocomial infections in Karunya Rural Community Healthcare Centre. The ABM framework was based on the framework described in [8]. The framework that was developed has been depicted in Fig.1. The various agents who are involved in the framework include the patient agent, nurse agents and the doctor agents. The patient agents can be of high priority or low priority. The high priority patients are given treatment before the low priority patients. The agents can be either infected or uninfected.

The low priority patients will wait in the waiting room for the triage process. During the triage process the triage nurses will note down the patient details such as the temperature, height, weight, blood pressure, etc. After the triage process, the patient again waits in the waiting room for registration. Once the registration is complete, the patient will be assigned to a doctor. The patient again waits for his time slot to visit the doctor. The doctor evaluates the patient and prescribes medicine. The patient purchases medicine and departs the hospital.

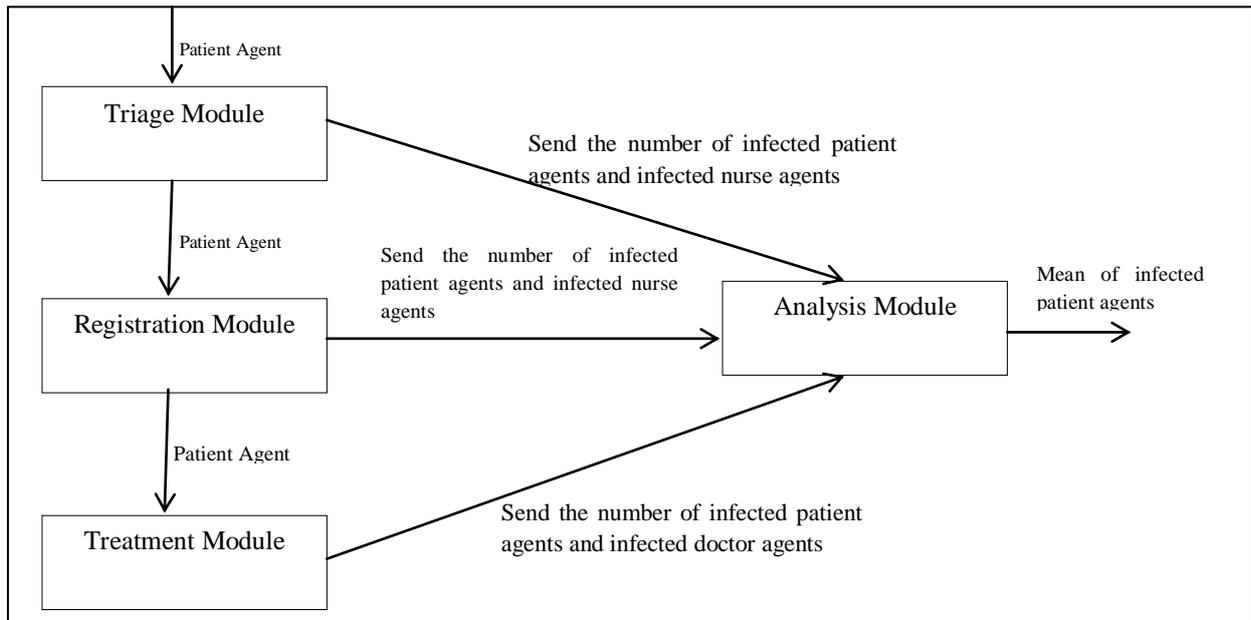


Fig.1 ABM framework

During each of the phase the agents would come in contact with each other. The transmission of infection from one agent to another is primarily driven by the distance between the agents. The distance between the agents would determine the type of contact between the agents. Three types of agent contacts have been identified namely close contact, casual contact and no contact [9]. The agent contact scenarios have been depicted in Fig.2 where X represents the distance between the agents.

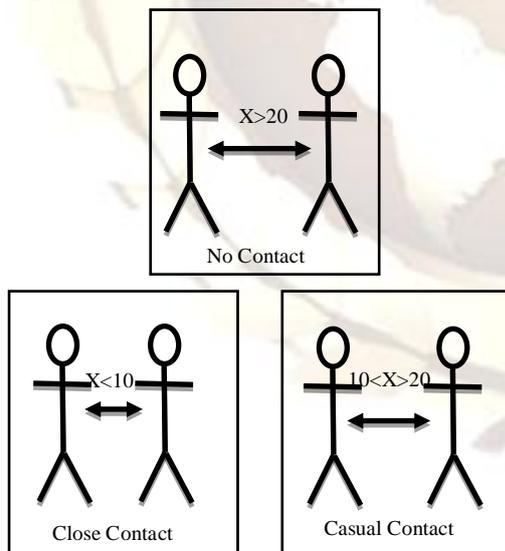


Fig.2 Agent contact scenarios

Each agent is given a time period time during which it can communicate with and react to the other agents. The proximity of the uninfected patient from the infected patient determines the transmission of infection.

If agent1 is located in position (x1, y1) and agent2 is located in position (x2, y2), then the distance between the agents is given by (1):

$$\text{Distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2} \quad (1)$$

If the distance between the agents is less than 30cm (10 pixels) then the contact is considered as close contact. If the distance is between 30 and 60 cm (10 and 20 pixels) then the contact is considered as casual contact. If the distance between the agents is greater than 60cm (20 pixels) then the contact is considered as no contact. Agents who are within close contact have greater chances of getting infected when compared to the agents having casual contact.

The size of the waiting room would also affect the transmission of infections. The smaller the size of the waiting room the probability that the infection might get spread is greater. The smaller size of the waiting room would bring larger number of agents into close contact thereby increasing the chances of infection spread [9].

The number of triage nurses and registration nurses would also affect the number of patients who can be serviced concurrently. This would also reduce the number of patients who would have to wait in the waiting room for their turn. Increase in the number of doctors would also have a similar effect on the transmission probability [9].

Using the agent contact the transmission of nosocomial infections in a rural hospital is thoroughly studied so that proper control strategies can be taken to control the spread of nosocomial infections.

### III. ID3 Decision Tree Algorithm

ABM is not sensitive enough to quantitatively and exactly validate the full range of infection control strategies. The number of patients getting infected would help to determine the mean of the patients getting infected. The current approach determines the transmission dynamics and the transmission vectors manually. These results have been further validated using the decision tree classification algorithms.

The decision tree classification algorithm is used to classify the various attributes based on which the decision would be taken as to how the infection spreads from uninfected patients to infected patients [10].

The ID3 algorithm is used to construct the decision tree by performing a top-down, greedy search through the given sets to test each attribute value at every tree node. In order to select the most useful attribute for classifying a given set, a metric known as information gain is used. To find an optimal way to classify a learning set, we need to minimize the depth of the tree [11, 12]. The information gain would serve the purpose of measuring which questions provide the most balanced splitting. It specifies the attribute based on which we can split the decision tree to create the next level of the decision tree. Using the ID3 the various attributes can be classified that would lead to the spread of nosocomial infections in a rural hospital.

The various attributes that have been used to identify the spread of nosocomial infections in a rural hospital is shown in Table 1.

**Table 1** Decision tree attributes and values

ATTRIBUTES	VALUES
Agent Type	patient, doctor, nurse
Priority	High, Low
Contact type	Close, Casual, No
Infectious	Y (Yes), N(No)
Decision	N(Negative), P(Positive)

Learning set was generated using the various attributes and values. The Learning set would contain the values for the different attributes at any particular point of time. The entire learning set is initially considered as a single tree rooted at a root node. Then the information gain and entropy for each of the attribute has to be determined. The attribute that is having the largest information gain is selected as the node based on which the decision tree can be split. The process is recursively applied until the Entropy is obtained as zero.

The ID3 algorithm is as follows:

ID3(S,A,V):

Step 1: Begin

Step 2: Load learning set and create decision tree root node 'decNode'. Add learning set S into the root node as its subset.

Step 3: For decNode, compute Entropy (decNode.subset) using (2).

If Entropy (decNode.subset) ==0, then goto Step 4 else goto Step 5.

Step 4: decNode.subset consists of records all with the same value for the categorical attribute, return a leaf node with decision attribute: attribute value;

Step 5: Compute information gain for each attribute left (have not been used in splitting), find attribute A with Maximum (Gain(S, A)). Create child nodes of this decNode and add to decNode in the decision tree.

Step 6: For each child of the decNode, apply ID3(S,A,V) recursively until reach node that has entropy=0 or reach the leaf node.

Step 7: End ID3.

The Entropy of the learning set(S) is calculated using (2):

$$\text{Entropy}(S) = \sum -p(K) \log_2 p(K) \quad (2)$$

where

$p(K)$  is the proportion of S belonging to class K.  $\sum$  is over the total number of outcomes.

Information gain is defined as the expected decrease of entropy related to a definite attribute when splitting a decision tree node. The Information gain for an attribute A in a Learning Set S is calculated using (3):

Gain(S,A)=Entropy(S)-

$$\sum_{v=1}^n (|S_v| / |S|) \times \text{Entropy}(S_v) \quad (3)$$

Where

$\sum$  is each value v of all possible values of attribute A

$S_v$  = subset of S for which attribute A has value v

$|S_v|$  = number of elements in  $S_v$

$|S|$  = number of elements in S

The ID3 generates a decision tree for the identification of the spread of nosocomial infections in an accurate manner. This can be used as a decision support tool for identifying the proper transmission vectors for the spread of nosocomial infections.

### IV. Results and Discussion

The ABM framework was applied to study the spread of nosocomial infections in Karunya Rural Community Healthcare Centre. The layout of the rural hospital was simulated with a triage area, registration room, waiting room and the treatment room. The layout of the rural hospital is depicted in Fig.3.

The doctor agents, nurse agents and patient agents are generated based on the simulation parameters that are specified at the start of the scenario. The triage nurse agents occupy the triage room while the registration nurse agents occupy the registration room. The doctor agents occupy the treatment room. The low priority patient agents would have to go through the normal flow of the rural hospital whereas the high

priority patient agents would move directly to the treatment room after the triage process. During each of the position of the patient agent, the distance between the agents is calculated based on which the type of contact is determined. Based on the

type of agent contact it can be determined whether the patient agent gets infected or not. In each of the scenario, the number of patient agents who get infected are determined.



Fig.3 Rural Hospital Layout

Scenarios were simulated with varying number of infected patient agents but keeping the nurse agents and doctor agents to be uninfected. The number of patient agents who got infected due to the infected patient agents are depicted in Table 2.

**Table 2** Number of Infected Patient Agents due to Already Infected Patient Agents

TOTAL NUMBER OF AGENTS	INITIAL NUMBER OF INFECTED PATIENT AGENTS	FINAL NUMBER OF INFECTED PATIENT AGENTS
22	8	9
30	10	12
50	25	30
60	25	27
75	50	60

Scenarios were simulated by varying the number of nurse agents who are infected. The nurse agents are capable of transmitting the nosocomial infection to the uninfected patient agents. The number of patient agents who got infected due to the infected nurse agents is depicted in Table 3.

Scenarios were simulated with the doctor agents as transmission vectors and the transmission of nosocomial infections due to the infected doctor agents were also studied. The number of infected doctor agents was varied and the transmission of nosocomial infection was studied. The number of

patient agents who got infected due to the infected doctor agents is represented in Table 4.

**Table 3** Number of Infected Patient Agents due to Already Infected Nurse Agents

TOTAL NUMBER OF AGENTS	INITIAL NUMBER OF INFECTED NURSE AGENTS	FINAL NUMBER OF INFECTED PATIENT AGENTS
22	1	8
30	1	12
50	2	20
60	2	22
75	3	42

**Table 4** Number of Infected Patient Agents due to Already Infected Doctor Agents

TOTAL NUMBER OF AGENTS	INITIAL NUMBER OF INFECTED DOCTOR AGENTS	FINAL NUMBER OF INFECTED PATIENT AGENTS
22	1	9
30	1	11
50	2	15
60	3	15
75	3	40

The variation in the number of infected patient agents due to the infected nurse agents, patient agents and doctor agents is depicted in Fig. 4.

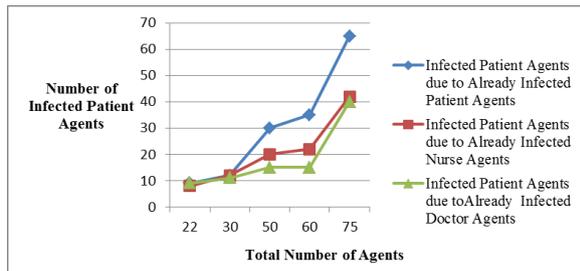


Fig.4 Number of Infected Patient Agents

The graph shows a linear increase in the number of patient agents who get infected due to the infected patient agents when compared to the number of infected patient agents due to the infected nurse agents and doctor agents.

Hence ABM provides a realistic visualization of the hospital layout and it assists in identifying the transmission vectors for the spread of nosocomial infections in a rural hospital. The ABM acts a decision support tool that enables the healthcare practitioners to take the necessary steps to prevent the spread of nosocomial infections in a rural hospital.

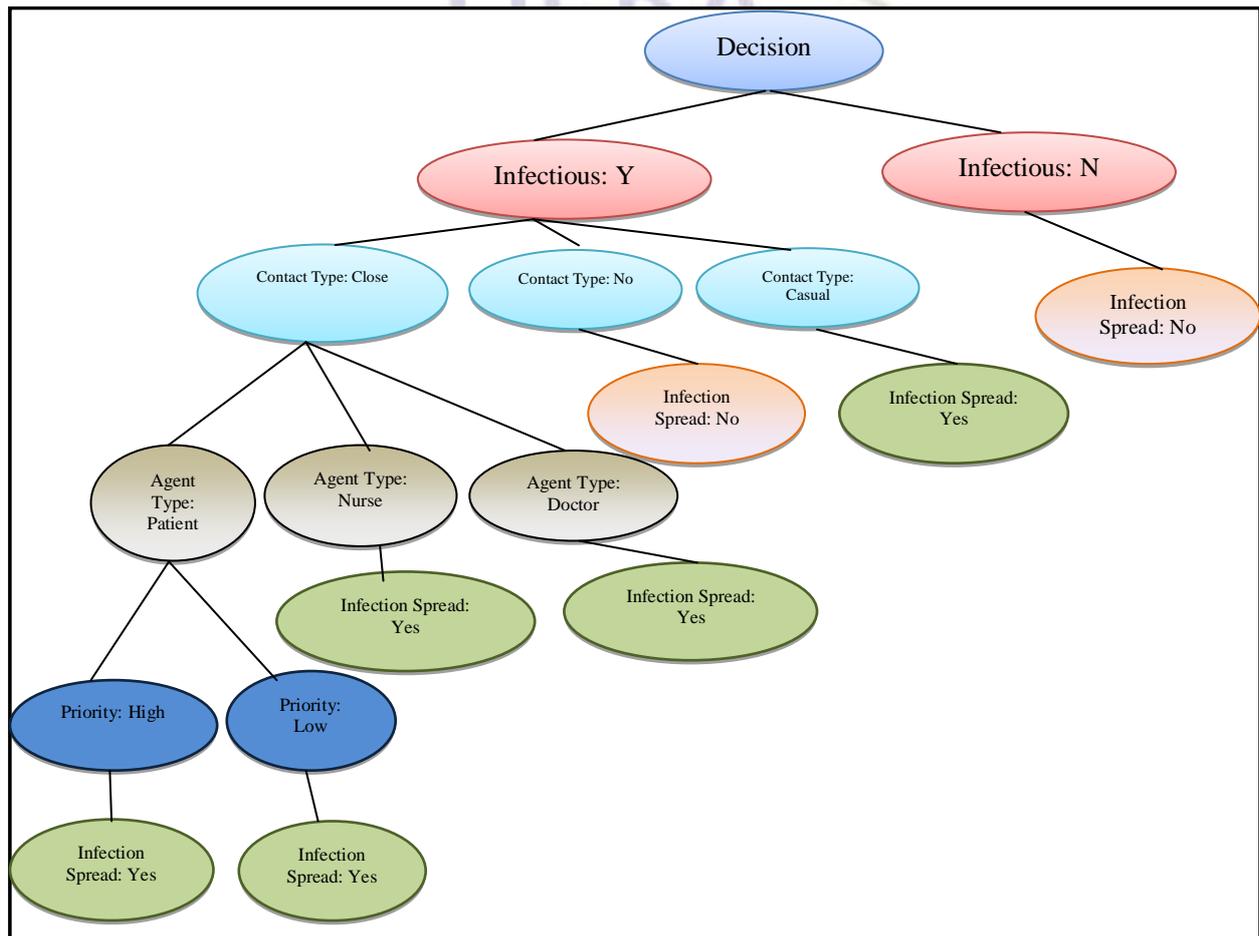


Fig.5 ID3 Decision tree structure for the spread of nosocomial infections.

The results that were obtained using the ABM has been further validated using the ID3 decision tree algorithm. The ID3 decision tree algorithm has been used to generate the decision tree for the spread of nosocomial infections. The decision tree that was generated also yields the same result as of ABM which validates the results.

The generated ID3 decision tree structure is depicted in Fig.5. From the generated ID3 decision tree it can be concluded that the infected patients having close contact are the major transmission vectors. These

results enable the healthcare practitioners to take the necessary steps to prevent the spread of nosocomial infections.

Hence the ABM and the ID3 decision tree structure acts as a decision support tool. It enables the proper control strategies to be taken so that the spread of nosocomial infections can be controlled to an extent. The control of the spread of nosocomial infections would lead to a reduced patient mortality rate.

## V. Conclusion

The study provides insight into the spread of nosocomial infections in a rural hospital. Nosocomial infections are a major cause of patient mortality and also cause the rise in healthcare cost. Trying to curb the disease after the outbreak would prove to be a futile effort. As always prevention is better than cure ABM has been applied to study the spread of the nosocomial infections so that proper control strategies can be incorporated at the appropriate time to prevent a major patient fatality.

ABM has been used to generate a realistic model of the scenario which enables to study the spread of nosocomial infections in a rural hospital in a very effective manner. The ABM acts as a decision support tool for the healthcare workers and the policy makers. From the simulated scenarios the infected patients have been identified as the major reason for the spread of nosocomial infections in a rural hospital. The results that have been obtained from the ABM have been further validated using the ID3 decision tree algorithm. The ID3 also identifies the infected patients within close contact as the major transmission vector for the spread of nosocomial infections.

Proper identification of the transmission vectors of the spread of nosocomial infections in a rural hospital would help to reduce the patient mortality rate. This would lead to providing better healthcare environment to the patients in a rural hospital.

## REFERENCES

- [1] L. Temime, G. Hejblum, M. Setbon and A. J. Valleron, The rising impact of mathematical modeling in epidemiology: Antibiotic resistance research as a case study, *Epidemiological Infections*, 136, 2007, 289-298
- [2] G. V. Bobashev, D. M. Goedecke, F. Yu and J. M. Epstein, A hybrid epidemic model: Combining the advantages of agent-based and equation-based approaches, *Proceedings of the Winter Simulation Conference*, 2007, 1532-1537
- [3] S. Hoya, A. Martin and G. Rodriguez, Using Cellular Automata to simulate epidemic diseases, *Applied Mathematical Sciences*, 3, 2009, 959-968
- [4] E. Bonabeau, Agent-based modeling: Methods and techniques for simulating human systems, *Proceedings of the National Academy of Sciences [Online]*, 2002, 99(suppl. 3):7280-7287
- [5] B. S. Ong, M. Chen, V. Lee and J. C. Tay, An individual-based model of influenza in nosocomial environments, *Proceedings of the International Conferences on Computational Sciences-Part I*, 2008, LNCS 5101, 590-599
- [6] A. K. Kanagarajah, P. A. Lindsay, A. M. Miller and D. W. Parker, An exploration into the uses of agent-based modeling to improve quality of health care, *International Conference for Complex Systems*, Boston, M.A., 2006
- [7] P.M.Barnaghi, V.A.Sahzabi and A.A. Bakar, A comparative study for various methods of classification, *International Conference on Information and Computer Networks*, Singapore, 2012, 62-66
- [8] M.Laskowski, R.D.McLeod, M.R. Friesen, B. W. Podaima and A.S. Alfa, Models of emergency departments for reducing patient waiting times, *PLoS ONE*, 4(7), 2009
- [9] Marek Laskowski, Bryan C.P. Demianyk, Julia Witt, Shamir N.Mukhi, Marcia R. Friesen and Robert McLeod, "Agent-based Modeling of the spread of influenza like illness in an emergency department", *IEEE Transactions on Information Technology in Biomedicine*, Vol.15, No.6, pp 877-889, 2011
- [10] Matthew N. Anyanwu & Sajjan G. Shiva, Comparative analysis of serial decision tree classification algorithms, *International Journal of Computer Science and Security*, Vol.3, pp. 230-240, 2009
- [11] Smitha T, Sundaram V, Classification rules by decision tree for disease prediction, *International Journal of Computer Applications*, Vol. 43, 2012, pp.35-37
- [12] Venkata Naga Sudheer T, A. Rama Mohan Reddy, Decision Tree Construction Based on Degree of Rough Sets, *International Journal of Engineering Research and Applications*, 1(3), 677-682