Vehicular Network based Emergency Data Transmission and Classification for Health Care System using Support Vector Machine

Biswa Ranjan Senapati (biswa.mjn@gmail.com)
National Institute of Technology, Rourkela
https://orcid.org/0000-0001-8975-2812

Pabitra Mohan Khilar
National Institute of Technology Rourkela

Tirtharaj Dash
BITS Pilani - K K Birla Goa Campus: Birla Institute of Technology and Science Pilani - K K Birla Goa Campus

Rakesh Ranjan Swain
SOA: Siksha O Anusandhan University

Research Article

Keywords: Delay, NIB, Routeval, Support vector machine, VANET

Posted Date: July 5th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1224943/v1

License: ☝️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Vehicular Network based Emergency Data Transmission and Classification for Health Care System using Support Vector Machine

1Biswa Ranjan Senapati, 1Pabitra Mohan Khilar, 2Tirtharaj Dash, and 3Rakesh Ranjan Swain

1 Department of Computer Science & Engineering, National Institute of Technology, Rourkela, 769008, India.
2 Department of Computer Science and Information Systems Birla Institute of Technology and Science Pilani, K.K. Birla Goa Campus, Goa 403726, India.
3 Department of CSE, ITER Siksha ‘O’ Anusandhan (Deemed to be), Bhubaneswar, India

1biswa.rnjn@gmail.com, pmkhilar@nitrkl.ac.in
2dashtirtharaj@gmail.com
3rakeshswain89@gmail.com

Abstract. COVID-19 pandemic has created an emergency across the globe. The number of corona positive cases and death cases are still rising across the world. The government of all countries is taking various steps to control the infection of COVID-19. One of the steps to control the spreading of the coronavirus is to quarantine. But the number of active cases at the quarantine center is increasing day by day. Also, the doctors, nurses, and paramedical staff providing service to the people at the quarantine center are getting infected. This demands the automatic and regular monitoring of people at the quarantine center. This paper proposed a novel and automated method for the monitoring of people at the quarantine center by two phases. These are the health data transmission phase and health data analysis phase. The health data transmission phase involves component NIBs, RSUs, and vehicles. For the transmission of data from the quarantine center to the observation center, an effective route is determined using route value. The route-value is dependent on factors like density, shortest path, delay, vehicular data carrying delay, and attenuation. The analysis of health data is done by using multi-class SVM. Finally, simulation results are obtained for the two phases. The simulation results obtained for the health data transmission phase is E2E delay, number of network gaps, and packet delivery ratio. The health data analysis phase determines parameters such as precision, recall, accuracy, etc.

Keywords: Delay, NIB, Routeval, Support vector machine, VANET
1 Introduction

COVID-19 is one of the current research topics since its outbreak in December 2019 [1]. In the present scenario, common people are still afraid of COVID-19 as the number of COVID-19 active cases and death cases are still rising across the globe. In a short period, coronavirus affects the health of people across 213 countries [2]. Within a few months, COVID-19 has changed the habit and living style of people. It massively affects every aspect of the individual’s life and propensities. Apart from the health sectors, various other sectors like the global economy, agriculture, transportation, education, sports, and entertainment are severely affected due to the impact of COVID-19. The effect of coronavirus is so dangerous that the World Health Organization declared this outbreak as the global pandemic [3]. To restrict the spread of coronavirus in the lack of medication and vaccine, the government of all countries has taken various steps. Social distancing is the only way to prevent the transmission of coronavirus. For maintaining social distancing, the government has imposed lockdown, shutdown, work from home, etc. Apart from this, states and union territories are divided into different zone like the red zone, orange zone, green zone, containment zone, and buffer zone. Another way to control the spreading of the coronavirus is to quarantine. The objective of quarantine is to prevent transmission of an infectious agent from those potentially incubating the disease [4]. But a recent study found that people in the quarantine center are not safe. In India, most of the corona positive cases are found in the quarantine center. Also, manual periodic health service provision by the doctors, nurses, and paramedical staff at the quarantine center puts the fear of COVID-19 infections on the corona warriors. In Haryana, front-line corona warriors account for 12% COVID-19 infections. In Bihar, 250 doctors and health workers are infected due to COVID-19. All these statistics motivate the automatic monitoring of people at the quarantine center to protect the corona warriors. This paper focuses on the transmission of health parameters from the quarantine center to an observation center for monitoring the health of the people staying at the quarantine center. As some of the quarantine centers are present in some rural remote areas, so transmission of health parameters requires some ad-hoc network. One of the choice for the automatic transmission of health parameter from the quarantine center to the observation center is the vehicular ad-hoc network (VANET). The reason for VANET is due to security provision to data during transmission [5], availability of reliable and efficient data dissemination scheme [6], and availability of the vehicular communication standard for the effective communication among the nodes of VANET [7].

VANET is one of the emerging as well as a challenging research area due to the dynamic, unstructured, random motion, and self-organizing capability of the nodes of the vehicular network [8–10]. Increase in the number of vehicles due to technological advancement, presence of On Board Unit (OBU) in the vehicle for the data communication, availability of vehicular standards like Wireless Access in Vehicular Environment (WAVE) [11] and Dedicated Short Range Communications (DSRC) [12], physical and MAC layer service provision by IEEE 802.1 p [13], availability of 5G for faster data communication in ve-
Vehicle to Vehicle (V2V): Communication between the vehicles with the help of a communication unit (OBU) when the vehicles lie within the communication range of another vehicle is called V2V communication. During traffic congestion due to accidents, route diversion of vehicles is possible due to V2V communication.

2. Vehicle to Infrastructure (V2I or I2V): Communication between the vehicle and RSU or vehicle and NIB when the vehicle and RSU or vehicle and NIB are present within the communication range is called V2I communication. V2I or I2V is generally preferred for broadcasting an advertisement like the offer of the nearer restaurant, petrol pump, etc.

3. Vehicle to Network (V2N): When a vehicle can send and receive data from a network with the help of cloud computing, fog computing, edge computing, etc. is called V2N communication. For the computation of data at a remote location, V2N communication is suitable.

4. Vehicle to Pedestrian (V2P): Communication between a vehicle and a pedestrian with the help of a cell phone present within the communication range is called V2P communication. Broadcasting the warning message to the pedestrian present near a moving vehicle, broadcasting local environmental conditions like rainfall are examples of V2P communication.

The overall types of communication are shown in figure 1.
Since for the proposed work, health data transmission begins at the quarantine center, and communication ends at the observation center, for this, a communication physical network component is required at the quarantine as well as observation center for efficient data communication. Since in the future the whole world will surely overcome this global pandemic, so the physical component at the quarantine and observation center must be flexible, temporary, reliable, and also suitable for the terrestrial region. One such favorable physical component is Network in Box (NIB) which is portable, self-organizing, provides data transmission facilities, network connectivity, and also provides call facilities. Apart from the above-mentioned features of NIB, the latter also provides secure data communication with the network components of VANET like on board unit (OBU) and roadside unit (RSU). Based on the data communication through the nodes of VANET and features of NIB, the major contribution of the proposed work is as follows.

1. Transmission of health parameters from the quarantine center to the observation center through automated routing involving the components RSU, OBU, and NIB.
2. Consideration of forward and backward moving vehicles, NIB at the beginning and at the end of communication to reduce the generic network parameters like number of hops, delay, and network gaps.
3. Analysis of health parameters at the observation center and classification of the person’s health into low risk, medium risk, or high risk due to COVID-19 is done using the One Against All (OAA) method of multi-class classifier of SVM.

The organization of the rest of the paper is shown in Figure 2.

2 Literature Survey

This section broadly discusses the three aspects. The first aspect focuses on various applications of VANET. The second aspect focuses on the applications of NIB. In the last aspect, various techniques for the detection of COVID-19 is discussed.

2.1 Applications of VANET

Availability of OBU in the vehicle, increase in the number of the vehicle creates an ad hoc network called VANET which is used for various applications. Broadly the application of VANET is classified into four categories. These are safety applications, convenience applications, commercial applications, and productive applications. Safety applications of VANET include collision avoidance [16], reliable data dissemination for traffic safety application [17], automatic fire monitoring through VANET [18], wireless access technology for safety application of VANET [19], etc. Convenience applications provide comfort and flexibility to the
driver and passenger of the vehicles and also reduces the waiting time and service time for various applications. Convenience application includes automatic toll tax collection [20, 21], automatic traffic congestion management [22, 23], automatic emergency vehicle detection system [24, 25], automatic parking service through VANET [26], etc. VANET is also suitable to provide various commercial applications due to which VANET is popularly known as the market on wheels. Different commercial application of VANET includes buying and selling of products through vehicles [27, 28] value-added advertisements like the advertisement of the special offers of restaurants, shopping malls, petrol pumps, etc. [29, 30], sharing of Wi-Fi to access the Internet [12, 31], etc. Productive applications of VANET include some applications which are performed efficiently and effectively. Different productive applications provided by VANET are environmental parameter monitoring through VANET [32], secure transaction during automatic toll tax collection [33], etc. Optimization of various vehicular parameters [34] and self fault detection [35] in VANET [36] encourage to use VANET for the wide range of applications. Although VANET is widely used for various applications, due to the random motion of the vehicles, connectivity is a challenging issue.

Fig. 2. Overall organization of the paper
The role of NIB is crucial for connectivity particularly at the source and at the destination. Various applications of NIB are discussed in the next section.

2.2 Application of NIB

Presence of various components in NIB help to use NIB for different applications. Different components present in NIB are Content Server (CS), Information Management System (IMS), Software Gateway (SG), Local Storage System (LSS), evolved Node B(eNB), Sensor devices (SD), etc. [37]. NIB also consists of radio communication device due to which NIB can communicate with the other network like cloud network, access points, RSUs, and OBUs of the vehicles. The most important application of NIB is to provide network connectivity after disasters such as earthquakes, tsunamis, cyclone, super-cyclone, and terrorist attacks to perform the rescue operations ([38]). The second application area of NIB is in the military camp. The establishment of a permanent network for the temporary camp is time-consuming and also costly. In such cases, NIB provides faster, portable, and cheap network deployment for communication ([39]). To meet the demands of the user during an emergency situation, NIB provides the ease and swift communication features due to the above-mentioned components of NIB and interaction with the external environment like mobile devices, cloud network, etc. Another important advantage of NIB is its compact size. Due to its compact size, NIB can be mounted on vehicles, deployed on drones, and can be carried in backpacks which is impossible for the macro base station. Thus, the connectivity issue is not an issue in the presence of NIB [40]. As some of the quarantine centers are present in the terrestrial region having Internet connection issues, so VANET with NIB is the suitable solution for the health data transmission from the quarantine center. The different approach for COVID-19 infection detection is discussed in the next section.

2.3 Approach to detect COVID-19 infection

A lot of research work is going on for the detection of COVID-19 infection due to the absence of vaccines and medicines to cure COVID-19. One of the approach to detect the COVID-19 infection is the nucleic acid test [41]. But to carry out the nucleic acid test is time-consuming and requires extensive human labor. Also, the False Negative Rate (FNR) is high for this approach. Secondly, the oral swab sampling technique is used for the detection of COVID-19 infection [42]. But this approach causes stimulated retching and coughing which increases the exposure risk of sample collecting health workers. Another approach for COVID-19 infection detection is the collection of saliva. This method is more effective than the swab collection approach. Also, the risk of COVID-19 infection for the health worker is less in this approach. Also, this approach is highly consistent. But this diagnosis method has a lower respiratory tract of around 28% which is the limitation of this diagnostic method [43]. Another approach for COVID-19 infection detection is the chest X-ray images using cluster-based principle [44]. But carrying out this diagnosis for many people at the quarantine center is
Health care System using VANET

Time-consuming and also there is a COVID-19 infection risk for health workers carrying out this diagnosis. Looking at the above limitations, a machine learning approach is proposed for the automatic COVID-19 infection detection for the people at the quarantine center and the diagnosis of health data from many quarantine centers is carried out at a single place called as observation center. Also, the transmission of health data from the quarantine center to the base station is done by the nodes of VANET and NIB. The system model for the proposed work is discussed in the next section.

3 System model

This section discusses various assumptions, network model, and communication model. Various notations used for the network model, communication model, and proposed work is given in Table 1.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Meanings</th>
<th>Notations</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle</td>
<td>Junc</td>
<td>Route Value</td>
<td>T</td>
</tr>
<tr>
<td>Route VAL</td>
<td></td>
<td>Distance</td>
<td>V</td>
</tr>
<tr>
<td>QC</td>
<td>QC-ID</td>
<td>OBU</td>
<td>NIB</td>
</tr>
<tr>
<td>OB</td>
<td></td>
<td>RSU</td>
<td>PDR</td>
</tr>
<tr>
<td>NOL</td>
<td></td>
<td>Tr</td>
<td>OC</td>
</tr>
<tr>
<td>QC-ID</td>
<td></td>
<td>OBU</td>
<td>NIB</td>
</tr>
<tr>
<td>BTEMP</td>
<td>HRATE</td>
<td>SBP</td>
<td>DBP</td>
</tr>
</tbody>
</table>

3.1 Assumptions

Various assumptions for the proposed work is mentioned below.

1. Health data transmission starts from a quarantine center using NIB.
2. Health data reception at the observation center is also done using NIB.
3. The communication range of NIB and RSU is the same and is greater than the OBU of the vehicle.
4. The connectivity link between the nodes (OBU, RSU, and NIB) are symmetric.
5. Using Poisson distribution the vehicles are distributed in the network.
6. All the junctions are associated with RSU.
7. Transmission of health data from the source to the destination occurs by multi-hop communication.
3.2 Network model

The vehicular network is modeled as a graph G(V,E).

- V is the set of junctions or RSUs.

\[ V \in \{ \text{Junc}_1, \text{Junc}_2, ..., \text{Junc}_n \} \]  

- E is the set of edges which are the road network connecting the junctions.

\[ E = (\text{SrJunc}, \text{DestJunc}, NOL, V_{\text{max}}) \]  

In the Equation 2, SrJunc indicates the beginning of the road network, DestJunc is the end of the road network, NOL refers to the number of lanes in the road network, and Vmax is the maximum speed of the vehicle in the road network.

Two nodes in the vehicular network can communicate directly if the distance between the two nodes is less than the transmission range. For example, let the coordinate of node1 is \((x_1, y_1)\) and the coordinate of node2 is \((x_2, y_2)\). The condition for direct communication is given in Equation 3.2.

\[ \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \leq T_r \]  

Also, the link between the nodes in the network is undirected. This means if node1 is connected to node2, then node2 is also connected to node1. This is represented by the equation.

\[ (ln_{i,j} == \text{TRUE}) \Leftrightarrow (ln_{j,i} == \text{TRUE}) \]  

3.3 Communication model

Communication in the proposed work is divided into three layers and is shown in figure 3. In the first layer, health data transmission begins from NIB. In the second layer, the data is propagated towards the destination through the nodes of VANET (vehicles and RSUs). In the third layer, health data is transmitted to the NIB present at the observation center from the nodes of VANET. In the second layer, the rules for communication between the nodes of VANET is shown below.

1. The vehicle carrying the data packet transmits to the vehicle moving in the same lane in the same direction towards the destination.
2. The vehicle carrying the data packet transmits to the RSU present in the forward direction of motion towards the center.
3. The vehicle carrying the data packet transmits the data packet to the vehicle moving in the opposite direction present nearer to the destination if no node is present in the same lane and towards the destination.
4. If no neighboring nodes are present either in the forward direction in the same lane or the reverse direction in the second lane, then the vehicle itself carries the message.

The overall communication between the nodes of VANET is shown in Figure 4.

4 Proposed work

The proposed work is divided into two phases.

1. Health data transmission phase: - This involves the transmission of health data from the quarantine center to the base station.

2. Health data analysis phase: - This involves the analysis of health data at the observation center using a machine learning approach called support vector machine (SVM).

Discussion of both the phases is mentioned in the corresponding subsection.

Health data transmission phase

This phase discusses the various health parameters for the transmission and the effective route selection from the quarantine center to the observation center.

Various health parameters for transmission The selection criteria of the health parameters for COVID-19 infection detection is as follows.
1. COVID-19 infection risk for lab technician staff, doctors, nurses, and paramed-ical staff must be less.
2. The selection of health parameters is based on the greater accuracy of de-
tection of the infection of COVID-19.
3. The health parameters must be measured automatically by the sensors and
the parameters data must be easily loaded into the NIB.

Based on the above criteria, for the proposed work four parameters are con-
sidered. These are body temperature, heart rate, systolic blood pressure, and
diastolic blood pressure. Since the health parameters are transmitted from the
quarantine center to the observation center for the people at the quarantine
center, so the packet format for the transmission is shown in figure 5.

**Effective route selection for health data transmission** The selection of
route from source to destination is dynamic in nature as the motion of the ve-
icle is random. From one junction, the next junction is selected based on the
Definition 1: Route\textsubscript{Val}: It is the time required for the transmission of the data packet from one junction to the next junction. Computation of route value is done at the RSU because of the presence of RSU at the junction and its computation capability. The calculation of route value depends upon parameters such as reciprocal of the density of vehicles between the junctions (RDEN), shortest path ratio (SPTH), expected delay (EDEL) which is the sum of transmission delay, propagation delay, queuing delay, and processing delay, vehicle data carrying delay (VDCDEL), and signal loss i.e. attenuation (ATEN).

Definition 2: Reciprocal of density (RDEN): Density refers to the number of vehicles present between two junctions. A greater number of vehicles is suitable for data transmission. But the selection of route value is one of the minimization problems, so RDEN is considered.

Definition 3: Shortest path ratio (SPTH): SPTH is the ratio of the shortest path from the next junction to the destination to that of the current junction to the destination.

Definition 4: Expected delay (EDEL): EDEL is the sum of transmission delay, queuing delay, propagation delay, and processing delay.

Definition 5: Vehicle data carrying delay (VDCDEL): When the vehicle carrying the data packet not able to find the node in the forward direction towards the destination, then the vehicle itself carries the data packet. As the speed of the vehicle is very less as compared to the speed of propagation of the signal, then this led to a delay called vehicle data carrying delay.

Definition 6: Attenuation (ATEN): Attenuation refers to the loss of strength of the signal during its propagation from source to destination. Mathematically, the route value is computed by the Equation 5.

\[ Route\textsubscript{Val} = w_1 \times RDEN + w_2 \times SPTH + w_3 \times EDEL + w_4 \times VDCDEL + w_5 \times ATEN \]  

Algorithm for route selection for health data transmission The algorithm for the route selection for the health data transmission from source to destination is mentioned in Algorithm 1.
Algorithm 1 Route selection for health data transmission

1: **Input:** Number of vehicles, RSUs, NIBs
2: **Output:** Optimized route value from source to destination
3: Transmit health data from NIB of quarantine center to nearest junction (Source junction).
4: for **JuncSource** to **JuncDestination** do
5:  if **Junc** ≠ **JuncDestination** then
6:  for each path connected to the current Junction **JuncCurrent** do
7:  RSU determines EDEL, RDEN, SPTH, VDCDEL, ATEN
8:  RSU computes Routeval by Eq.5
9:  end for
10: Select minimum Routeval as the next forwarding route.
11: Transmit the data packet in the selected forwarding path.
12: for **JuncCurrent** to **JuncNext** do
13:  if **JuncNext** is reached then
14:  Transfer the data to RSU
15:  else
16:  Determine the Vnext vehicle
17:  if Vnext is moving in the same direction and same lane then
18:  Transmit health data
19:  else if Vnext is present in the opposite direction in another lane and
towards the destination then
20:  Transmit health data
21:  else
22:  carry data till it doesn’t found Vnext in the communication range
23:  end if
24:  end if
25: end for
26: end if
27: end for

4.1 Health data analysis phase

For the health data analysis phase, a supervised learning approach i.e. Support Vector Machine (SVM) is considered because of its several advantages like prevention from over-fitting due to good generalization capabilities, efficient handling of nonlinear data, the stability of the hyperplane even in the small change in data [45]. Below we describe three common implementations for multiclass-SVM: One-against-All (OAA), Error Correcting Output Code (ECOC), and One-against-One (OAO). We then provide a justification on why we chose a particular implementation in our work.

1. **One Against All (OAA) method:** It uses the principle of winners takes all strategy. This is also called as WTA-SVM [46]. Let M denote the number of classes. \( W_i \) denotes the \( i_{th} \) class, where \( i = 1, 2, 3, \ldots, M \). It constructs M number of binary classifiers. The output function \( \rho_i \) of the \( i_{th} \) class classifier is trained by taking the positive values of class \( w_i \) and the negative values of all the other classes. Mathematically, \( i_{th} \) class classifier solves the following problem.
The memory requirement for the classification using the OAA method is high which is the limitation of this method.

2. **Error correcting output code (ECOC)**: ECOC is a meta-method which is used for multi-class classification problem [47]. To solve the multi-class problem, ECOC combines many binary-SVM classifiers deal with multiple classes. The working of ECOC is dependent on a matrix called the coding matrix. Design of coding matrix and dependence on the coding matrix is the limitation of the ECOC method and therefore, we do not use this approach in our present work.

3. **One Against One (OAO) method**: This method is also called as max-wins voting (MWV-SVM) method [48]. This is a category of pairwise classification. This method creates one binary classifier for every pair of distinct classes. For N class classifier, MWV-SVM creates $\frac{M(M-1)}{2}$ binary classifiers. For training data from $i_{th}$ and $j_{th}$ class, the following problem is solved.

$$
\min(\omega^{ij}, b^{ij}, \xi^{ij}) \begin{cases}
\frac{1}{2} (\omega^{ij})^T \omega^{ij} + C \sum_{t=1}^{l} \xi^{ij}_t \omega^{ij}_t, & \text{if } y_t = i \\
(\omega^{ij})^T \phi(x_t) + b^{ij} \geq 1 - \xi^{ij}_t, & \text{if } y_t = j, j = 1, ..., l \\
(\omega^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi^{ij}_t, & \text{if } y_t \neq i, j = 1, ..., l \\
\xi^{ij}_t \geq 0, & j = 1, ..., l
\end{cases}
$$

All the above mentioned implementations have their pros and cons for a multiclass classification setting, which is the case discussed in our present paper (in our multiclass classification setting, we are dealing with four different classes). However, there is no conclusive evidence to suggest which of the above three methods is the “best” in a multiclass setting; see for example, [49] compares One-vs-One and One-vs-All classification. Therefore, we adopt the OAA method (also called, One-versus-the-Rest classification) for implementing our multiclass-SVM. This implementation is carried out within the popular Scikit-learn library [50].

5 **Analysis of the proposed work**

This section analyzes the delay and packet delivery ratio of the proposed work.

**Delay Analysis**

The estimated total time $ETT_i$ taken by node $N_i \in N$, transmitting data to its neighbor within its transmission range $t_r$ calculated as in Eq. (8):

$$ETT_i = \chi_{pr} + \chi_{tr} + \chi_q + \chi_{pd},$$


where $\chi_{pr}$ is the propagation delay, $\chi_{tr}$ is the transmission delay, $\chi_q$ is the queuing delay, and $\chi_{pd}$ is the processing delay. The $\chi_q$ and $\chi_{pd}$ values are very less, so we neglected to zero. The value of $\chi_{pr}$ is defined as in Eq. (9):

$$\chi_{pr} = \frac{\text{dist}(N_i, N_j)}{\vartheta},$$

(9)

here $\vartheta$ is the velocity of packet (speed of light) and $N_i$ to $N_j$ distance is calculated as $\text{dist}(N_i, N_j)$, where $N_i, N_j \in N$. The distance is calculated using Friss propagation model[]. The received power $p_r$ is computed as in in Eq. (10):

$$p_r = p_t \times g_t \times g_r \times \frac{\lambda^2}{(4 \times \pi \times d)^2},$$

(10)

where transmitting power is denoted as $p_t$, gain of transmitting antenna and receiving antenna are denoted as $g_t$ and $g_r$ respectively, wavelength of the signal is denoted as $\lambda$, and distance between receiving and transmitting antenna are denoted as $d$. The distance calculated as in Eq. (11):

$$d = \text{dist}(N_i, N_j) = \sqrt{\frac{p_t \times g_t \times g_r \times \lambda^2}{(4 \times \pi)^2} \times \frac{1}{p_r}}$$

(11)

The transmission delay is denoted as $\chi_{tr}$ and computed as in Eq. (12).

$$\chi_{tr} = \frac{P_L}{\lambda},$$

(12)

where length of the packet is denoted as $P_L$ and channel bandwidth is denoted as $\lambda$.

So, the estimated total time is computed as in Eq. (13):

$$ETT_i = \chi_{pr} + \chi_{tr} = \frac{\text{dist}(N_i, N_j)}{\vartheta} + \frac{P_L}{\lambda}.$$

(13)

The estimated total time $ETT_i$ taken by any node for transmitting data to its neighbor node within its transmission range. The communication is performed in 4 ways such as nodes to NIB device, NIB device to vehicular nodes, in between vehicular nodes, and vehicular nodes to NIB device within base station.

### 5.1 PDR Analysis

The packet delivery ratio is dependent upon the average throughput. The average throughput of this network have number of nodes contending for data transfer from source to destination per round trip time (RTT). Here, NIB is used for contention in the whole network for data transmission.

Let us $N$ number of nodes try to contending the mediums in the NIB device. In $N$ number of nodes some nodes are successfully transmits their data packets without collision and some are unsuccessful. We take a random variable $A$ and
its corresponding value is \( \alpha \). Hence, \( \alpha \leq N \). The total period is divided in small slots, i.e. NIB slots. Assume \( \alpha \) nodes choose \( \beta \) slots of the period. Where \( \beta \leq \alpha \), if \( \beta = \alpha \), successful transmission and will be no collisions. Let \( f(\alpha, \beta) \) represents the number of possible ways \( \alpha \) nodes assign to \( \beta \) slots. The problem same as the probability of \( \alpha \) balls into \( \beta \) packets. For this distribution, we need the help of inclusion and exclusion principle.

By inclusion and exclusion principle, for \( t \) finite sets \( x_1, x_2, \ldots, x_t \) represented in Eq. (14)

\[
\bigcup_{i=1}^{t} x_i = \sum_{1 \leq i \leq t} |x_i| - \sum_{1 \leq i < j \leq t} |x_i \cap x_j| + \ldots + (-1)^{n+1} \sum_{1 \leq i < \ldots < i_n \leq t} |x_{i_1} \cap \ldots \cap x_{i_n}| + \ldots
\]

\[+(-1)^{t+1}|x_1 \cap \ldots \cap x_t|,
\]

(14)

where \( |x| \) is the cardinality of set \( x \). If \( \cup \) is the universal set then this equation can also written as in Eq. (15):

\[
\bigcap_{i=1}^{t} x_i^c = |\cup| - \sum_{1 \leq i \leq t} |x_i| + \sum_{1 \leq i < j \leq t} |x_i \cap x_j| + \ldots + (-1)^n \sum_{1 \leq i < \ldots < i_n \leq t} |x_{i_1} \cap \ldots \cap x_{i_n}| + \ldots
\]

\[+(-1)^{t}|x_1 \cap \ldots \cap x_t|.
\]

(15)

So applying these inclusion and exclusion principle, we find out the distribution of nodes into the NIB slots. Let \( \alpha \) nodes and \( \beta \) slots, different cases are their for distribution.

(i) **Case-I:** Let \( \alpha \) distinguishable nodes and \( \beta \) distinguishable NIB slots with exclusion, means no slot contain more than one node, i.e no collision happen and some slot is also empty. In this cases number of different ways of distributing \( \alpha \) nodes into \( \beta \) slots is represented in Eq. (16):

\[
P(\beta, \alpha) = (\beta)_{\alpha} = \beta(\beta-1)(\beta-2)\ldots(\beta-\alpha+1).
\]

(16)

(ii) **Case-II:** Let \( \alpha \) distinguishable nodes and \( \beta \) distinguishable NIB slots with exclusion, i.e no collision happen and no slot is empty. In this cases number of different ways of distributing \( \alpha \) nodes into \( \beta \) slots is represented in Eq. (17):

\[
\begin{cases} 
\beta!, \text{ if } \alpha = \beta \\
0, \text{ if } \alpha \neq \beta
\end{cases}
\]

(17)

(iii) **Case-III:** Let \( \alpha \) distinguishable nodes and \( \beta \) distinguishable NIB slots without exclusion, i.e collision possible and some slot is also empty. In this cases number of different ways of distributing \( \alpha \) nodes into \( \beta \) slots is calculated as \((\beta)^\alpha\).
(iv) **Case-IV**: Let $\alpha$ distinguishable nodes and $\beta$ distinguishable NIB slots without exclusion, means slot may contain more than one node, i.e. collision possible and no slot is empty. In this case number of different ways of distributing $\alpha$ nodes into $\beta$ NIB slots is described by using the inclusion and exclusion principle. Let $x_i$ be the set of distribution contain no nodes in the slot number $i$. So we are consider the distributions which does not lie on the set of $x_1, x_2, ..., x_\alpha$. Here we find, $|x_1 \cap x_2 \cap ... \cap x_\alpha|$ is same as Eq. (15). We find the universal set $\bigcup$ as the distribution of $\alpha$ nodes into $\beta$ slots without any restriction calculated as $\beta^\alpha$. Next for calculating $|x_1 \cap ... \cap x_i|$ for the Eq. (15), this is same as distribution of $\alpha$ nodes in the remaining $\beta - n$ NIB slots, this is calculated as $(\beta - n)^\alpha$ ways. Therefore,

$$
\sum_{1 \leq i_1 < ... < i_n \leq t} |x_{i_1} \cap x_{i_2} \cap ... \cap x_{i_n}| = \left( \frac{\beta}{\beta - n} \right) (\beta - n)^\alpha.
$$

(18)

Substituting all these in Eq. (15), we find the different ways of distributing $\alpha$ nodes into $\beta$ NIB slots is computed in (19):

$$
\beta^\alpha - \beta(\beta - 1)^\alpha + \left( \frac{\beta}{\beta - 2} \right)(\beta - 2)^\alpha + ... + (-1)^n \left( \frac{\beta}{\beta - n} \right)(\beta - n)^\alpha + ... + (-1)^{\beta-1}\beta
$$

$$
= \sum_{n=0}^{\beta-1} (-1)^n \left( \frac{\beta}{\beta - n} \right) (\beta - n)^\alpha
$$

(19)

Let $N_d$ nodes have data to transmits in NIB medium. In the $N_d$ number of nodes $N_c$ nodes contend in the medium and remaining $N_d - N_c$ nodes reserved slots in the NIB medium. The $N_s$ number of nodes of $N_c$, complete the successful transmission in the NIB medium. Therefore the throughput $\sigma$ is calculated as in Eq. (20):

$$
\sigma = \frac{N_s + (N_d - N_c)}{N_d}.
$$

(20)

The steady state PDR equation, we defined as in (21):

$$
\sigma - \frac{N_s + (N_d - N_c)}{N_d} = 0.
$$

(21)

6 Simulation set up and result

This section focuses on the simulation set up and simulation result of the two phases i.e. health data transmission phase and health data analysis phase of the proposed work.

6.1 Simulation set up and simulation result of health data transmission phase

This subsection mentions the performance metric for the health data transmission phase, simulation set up, and simulation result of the above mentioned phase.
Performance metric for the health data transmission phase

1. **End-to-end delay (E2E delay):** E2E delay refers to the time taken for the transmission of data packet from the source node to destination node.

2. **Number of network gaps (NG):** NG is defined as the number of vehicles carrying the data packet due to the unavailability of the nodes in the forward direction towards the destination.

3. **Packet delivery ratio (PDR):** PDR is defined as the ratio of number of packets received at the destination to that of number of packets transmitted from the source.

Simulation set up for the health data transmission phase

The performance of the algorithm for the route selection for the health data transmission is evaluated using SUMO, and OMNet++. The overall area for the network scenario is considered as 1200 × 1200 m². Number of junction considered is 9. The number of vehicles are varied from 20 to 100. The speed of the vehicles are varied between 30 to 50 km/hr. The overall simulation parameters are represented in Table 2.

**Table 2. Simulation parameter set up for data transmission phase**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>1200 × 1200 m²</td>
</tr>
<tr>
<td>Number of Junctions</td>
<td>9</td>
</tr>
<tr>
<td>Number of roads associated with one junction</td>
<td>4</td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>20-100</td>
</tr>
<tr>
<td>Number of quarantine center</td>
<td>4</td>
</tr>
<tr>
<td>Number of observation center</td>
<td>1</td>
</tr>
<tr>
<td>Position of Quarantine center-1 (in meter)</td>
<td>(0,300)</td>
</tr>
<tr>
<td>Position of Quarantine center-2 (in meter)</td>
<td>(0,900)</td>
</tr>
<tr>
<td>Position of Quarantine center-3 (in meter)</td>
<td>(900,1200)</td>
</tr>
<tr>
<td>Position of Quarantine center-4 (in meter)</td>
<td>(900,0)</td>
</tr>
<tr>
<td>Position of observation center (in meter)</td>
<td>(600,600)</td>
</tr>
<tr>
<td>Speed of vehicle (km/hr)</td>
<td>30-50</td>
</tr>
<tr>
<td>Communication range of Vehicle</td>
<td>100 m</td>
</tr>
<tr>
<td>Communication range of RSU</td>
<td>150 m</td>
</tr>
<tr>
<td>communication range of NIB</td>
<td>150 m</td>
</tr>
<tr>
<td>size of packet</td>
<td>512 bytes</td>
</tr>
</tbody>
</table>

The overall network scenario is shown in Figure 6. The vehicular network scenario is created using a tool called Simulation of Urban MOBility (SUMO). The part of the network scenario using SUMO is shown in Figure 7.
The SUMO network is integrated in OMNET++ and the simulation result for the performance parameter is evaluated.

For the networking work E2E delay is one of the important parameter for the performance evaluation. So to determine weights, different value sets for the weights are considered and E2E delay is evaluated. The simulation result for the
E2E evaluation for different sets of weight is shown in Figure 8. The result is the average of 150 simulation result. Also, the weights value are selected in such a way that the sum of all the weights is equal to 1.

![Selection of weights](image)

**Fig. 8.** Selection of weights

From the figure 8, it is found that when all the weights are equal i.e. $w_1 = w_2 = w_3 = w_4 = w_5 = 0.20$, then the E2E delay is minimum. So for the further simulation work, all the weights are selected to be 0.2.

**Simulation result for the health parameter transmission phase** For the proposed network scenario, E2E delay is determined from the simulation result. E2E delay for the proposed work is compared with the existing routing protocol for VANET like GSR, P-GEDIR, A-STAR with respect to density of vehicles and is shown in figure 9.

Figure 9 shows that E2E delay approaches to 0 when the density of vehicles is around 40. Also, the E2E value decreases when the density of vehicles increases. This is because by increasing the density of vehicles network gap reduces. So VDCDEL decreases. Considering the vehicles far in the communication range towards the destination as the next forwarding node reduces the E2E delay for the proposed work as compared to GSR, P-GEDIR, and A-STAR. Since for the proposed work 4 NIBs are considered, so E2E delay for the individual NIB to Observation center is shown in figure 12.

Figure 12 shows the decrease in the E2E delay for the increase in the density of vehicles because of the above mentioned reason. Also, the density of vehicles for the selected path from the individual NIB to observation center is variable.
So E2E delay is variable for the individual NIB. The second performance metric i.e. network gap for the proposed work is compared with GSR, P-GEDIR, and A-STAR and is shown in figure 13.

Figure 12 shows the decrease in the E2E delay for the increase in the density of vehicles because of the above mentioned reason. Also, the density of vehicles for the selected path from the individual NIB to observation center is variable. So E2E delay is variable for the individual NIB. The second performance metric i.e. network gap for the proposed work is compared with GSR, P-GEDIR, and A-STAR and is shown in figure 13.

Figure 13 shows the decrease in the number of network gap for the increase in the density of vehicles. Network gap for the proposed work becomes zero at higher density of vehicles, i.e. at around 60 vehicles. The variation of PDR with respect to density of vehicles for the proposed work is mentioned in Table 3. For the PDR computation, simulation time is assumed to be a constant of 80 sec.

<table>
<thead>
<tr>
<th>Density of vehicles</th>
<th>5</th>
<th>20</th>
<th>35</th>
<th>50</th>
<th>65</th>
<th>80</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet delivery ratio (PDR)</td>
<td>62.26</td>
<td>78.42</td>
<td>87.84</td>
<td>96.46</td>
<td>91.62</td>
<td>85.52</td>
<td>81.58</td>
</tr>
</tbody>
</table>

PDR for the low density of vehicles are less because of more network gaps and greater VDCDEL. But by increasing the density of vehicles, PDR increases
and attend a maximum value. By further increasing the density of vehicles, PDR decreases due to packet collisions in the network.

6.2 Simulation set up and simulation result of health data analysis phase

This section discusses different performance metric, simulation set up and simulation result of the health data analysis phase.

Performance metric for the health data analysis phase As the data analysis phase is based on a machine learning approach SVM, so the performance metric based on machine learning approach is used. The performance metrics are mentioned below.

1. **Precision**: It is defined as the ratio of correctly predicted positive observation to the total predicted positive observation.

   \[
   Precision = \frac{True\text{-}positive}{True\text{-}positive + False\text{-}positive} \quad (22)
   \]

2. **Recall**: It is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

   \[
   Recall = \frac{True\text{-}positive}{True\text{-}positive + False\text{-}negative} \quad (23)
   \]

3. **F1 score**: It is the weighted average of precision and recall.

   \[
   F1\text{-}score = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \quad (24)
   \]
4. **ROC-AUC score**: The Area Under the Curve (AUC) summarizes the receiver-operating characteristic (ROC) curve and measures the ability of the classifier to distinguish between classes. A very high AUC (closer to 1.0) suggests that the model is very good at distinguishing one class against others. We use AUC score in a multiclass setting as we are dealing with four different classes in our dataset.

**Simulation set up for the data analysis phase**: The dataset for the data analysis phase consists of 4 parameters: body temperature, heart rate, diastolic blood pressure, and systolic blood pressure. Based on [?], we believe that high body temperature, more heart rate, and greater blood pressure are the characteristics of symptomatic COVID patient and our interest is in analysis whether a machine learning model is able to accurately diagnose the disease. Thus, the synthesized dataset is prepared based on the following conditions for the four parameters.

1. Body temperature in the range 60F-115F
2. Heart rate in the range of 40 pulse-120 pulse per minute
3. Systolic blood pressure in the range (60 -150) mmHg
4. Diastolic blood pressure in the range of (30-110) mmHg

The synthesized dataset consists of 5000 records of the combination of the four parameters. Based on the feature binning of the above four parameters each record is assigned a category of COVID-19 infection. Each record is assigned a category out of four categories of risk. The four categories of risk for the COVID-19 infected person are normal or no risk, low risk, medium risk, and high risk. The person having high risk must be immediately transferred to COVID hospital. The person having the medium risk must be isolated from the other people in the
quarantine center. Data distribution for various risk labels is shown in Figure 14.

Based on the dataset, the simulation result for the data analysis phase is represented in the next subsection.

The data analysis phase is implemented using Support Vector Machine (SVM) model within Scikit-learn library. The hyperparameters of the SVM is tuned using 5-fold cross-validation. Various hyperparameters are as tabulated below. We
evaluate 100 randomly sampled SVM models based on the above parameters using RandomizedSearchCV. Each model was scored using $f_1$-weighted score. All the simulations are conducted in a Linux machine with 16-core Intel Xeon processor (3.10 GHz), 64-GB main memory.

**Simulation result for the health data analysis phase** The time required for prediction is given in Table 5. The tabulation suggest that the model is very fast at prediction taking almost no time to make a decision on whether the patient (or the person) is normal or at any other risk-level.

**Table 5. Time required for prediction:**

<table>
<thead>
<tr>
<th>For all test instance (s):</th>
<th>0.0933 (1000 instances)</th>
</tr>
</thead>
<tbody>
<tr>
<td>For each test instance (s):</td>
<td>$9.33 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

A fast prediction could allow the system to trigger an alarm to deal with the patient. Although a fast prediction based on the patient’s data is essential in a real-time application, we are also interested to see whether the model is accurate
and confident enough in its prediction. We present the result on various aspect of this in Table 6. The tabulation reports the class-specific scores on various evaluative measures such as precision, recall, f1-score for the 1000 test data-instances available to us. The results suggest that the obtained model is highly accurate in overall while achieving a test accuracy of approximately 97%.

Table 6. Summary of Independent Testing:

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>f1-score</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.95</td>
<td>0.99</td>
<td>0.97</td>
<td>385</td>
</tr>
<tr>
<td>Low-risk</td>
<td>0.98</td>
<td>0.96</td>
<td>0.97</td>
<td>295</td>
</tr>
<tr>
<td>Medium-risk</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>207</td>
</tr>
<tr>
<td>High-risk</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>152</td>
</tr>
<tr>
<td>micro avg</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>1000</td>
</tr>
<tr>
<td>macro avg</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>1000</td>
</tr>
<tr>
<td>weighted avg</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>1000</td>
</tr>
<tr>
<td>samples avg</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>1000</td>
</tr>
<tr>
<td>Testing accuracy</td>
<td>0.968</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We further look closely at the model’s performance based on how confident the model is while predicting one class versus the rest in the multi-class setting. For this, we rely on the ROC curves and the precision-recall curve. The former summarize the trade-off between the true positive rate and the false positive rate for our multiclass-SVM model. The precision-recall curves summarize the trade-off between precision and recall, and are mainly preferred for a problem dealing with imbalanced dataset.

![Class-wise Receiver operating characteristic (ROC)](image)

**Fig. 15.** Class-wise Receiver Operating Characteristic (ROC) Curves. The class labels are 0: Normal, 1: Low-risk, 2: Medium-risk, 3: High-risk.
Fig. 16. Precision-Recall Curves. The class labels are 0: Normal, 1: Low-risk, 2: Medium-risk, 3: High-risk.

These two results for our present model are provided in Fig. 15 and 16 respectively. From these results it is evident that the SVM model is highly confident at predicting one class against the others, that is, the AUC score for individual class is greater than 0.97, with a good trade-off between the precision and recall, as seen from Fig. 16. In summary, these results suggest that the multiclass-SVM could serve as a real-time predictor for health-risk for the problem discussed in our present paper.

7 Conclusion and future scope

In this paper, automatic monitoring of people at the quarantine center of the terrestrial region is done by using NIBs, RSUs, and vehicles. For this automatic monitoring, two phases are discussed which are the health data transmission phase and health data analysis phase. The performance of the health data transmission phase is done with the help of parameters such as E2E delay, number of network gaps, and packet delivery ratio. The performance of the health data analysis phase is done using precision, recall, F1 score, and accuracy. The simulation result shows the proposed work of the health data analysis phase performs better as compared to the existing routing protocol such as GSR, P-GEDIR, and A-STAR. The performance is better because of the evaluation of Routeval at every junction. Simulation result for the synthetic health data analysis phase determines the value of testing accuracy as 95.8%.

The work discussed in this paper have a few limitations and we believe that could be addressed effectively: (a) the work assumes that the vehicles are connected to each other in some fashion; (b) the health data is synthetic. While
we believe that (a) is an ideal case, the work could be extended to incorporate disconnectivities in communication. Limitation (b), however, does not limit the work to be tested for real-world settings. The idea to use synthetic data for our present work is to show that a machine learning model is effective and this could be extended to incorporate any publicly available health data—which, we believe is not the case yet. Further, in the future, the proposed method can be enhanced to use practically in the Quarantine center.

Declaration

Funding: Not Applicable
Conflicts of interest: The authors declare that they have no conflict of interest.
Availability of data and material: Not Applicable
Code availability: Not Applicable

References


