Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network

Bridget Ogheneovo Malasowe¹, Deborah Voke Ojie², Arnold Adimabua Ojugo³, Margaret Dumebi Okpor⁴

¹Department Computer Science, Faculty of Computing, University of Delta, Agbor, Nigeria

²Department of Software Engineering, Faculty of Computing, University of Delta, Agbor.

³Department of Computer Science, College of Science, Federal University of Petroleum Resources, Effurun, Nigreria.

⁴Department of Cyber Security, Faculty of Computing, Delta State University of Science and Technology, Ozoro, Nigeria

Email: bridget.malasowe@unidel.edu.ng

Abstract
The adoption of data mining processes is urgently needed due to the everyday generation of large amounts of data at an accelerated rate. The current advancement in the area of data analytics and data science has ushered in a new paradigm shift in the use of machine learning and softcomputing approaches to a new paradigm to render a more beneficial approach in constructing algorithms that can effectively and efficiently assist expert systems to yield new insight to practitioners – to ensure comprehensive decisions on the underlying tuberculosis disease to potential problematic cases. This study explored spatial medical data in disease diagnosis to effectively and efficiently handle problematic cases of Tuberculosis in Nigeria. Bayesian Network algorithm was used to predict potential cases in patients with covid-19 (and other underlying health issues) vis-a-vis its co-prevalence rate with Tuberculosis with data retrieved from the epidemiology laboratory of the Asaba Federal Medical Centre, Delta State. Training and test versions of the data set were separated. Constructed model yields high prediction compared to previous studies in forecast of the prevalence.

Author for Correspondence
Malasowe B. O. et al, DUJOPAS 10 (2a): 80-94, 2024
co-infection rate. Results generated show that the confusion matrix model had sensitivity of 0.81, specificity 0.08, and prediction accuracy of 0.937 for data not originally used to train.

Keywords: Epidemic, Tuberculosis, Algorithm, Pandemic, data set, Network

INTRODUCTION
Infectious diseases often spread easily with a variety of impact according to (Jung et al., 2016; Oyemade & Ojugo, 2020, 2021) – giving society a great deal of concern. Noticeably, such disease range from flu, contagion, epidemics and pandemic (Emordi et al., 2023). Society today – is often poised with the quest to minimize its diffusion, which has led to studies aimed at providing policy-makers/experts with effective measures to help combat the menace (Akazue et al., 2022; Akazue, Yoro et al, 2023; Allenotor et al, 2015). Disease is a disorder that impedes performance, and causes a system’s structure to deviate in its functioning (Laddha et al., 2022). It is a deviation to the norms or function of a system (Allenotor & Ojugo, 2017). A disease is typically related with physical harm and occurs in an exact place. (Peng et al., 2022), and yields specific symptoms that appears (Udeze et al., 2022), to indicate the abnormal conditions in a system (Aghware et al., 2023a, 2023b; Aslan et al., 2022). Diseases spread through the use of aerosols, fomites, oral/injected drugs, direct contact, and vector-borne (Oyewola et al., 2021). Its outbreak across a society (made up of various actors) can morph as epidemic or pandemic (Lawton & Viriri, 2021; Tarawneh et al., 2019; Ucar & Korkmaz, 2020). The current work investigates linkages, organization, and basic properties of interest by modeling propagation events on a social graph of connected players (Aslan et al., 2021; Wemembu et al., 2014). There are four main types of diseases that are known to exist: infectious, deficient, inherited, and non-hereditary (Ojugo & Yoro, 2013; 2021; 2021; Okonta et al., 2013; 2014; Pillai, 2022).

The drive of players have continued to play a critical, important part in the transformational course of modeling the outcome in contagion propagation (Malasowe et al., 2023). Because, if properly harnessed – it will help experts formulate a pathway to de-escalate and manage such local outbreaks via targeted treatments (Muslık et al., 2023; Ojugo & Ekurume, 2021; 2021.). In order to help simulate epidemics and produce a dynamic framework as potential future contagion prevention strategies, this has also made it necessary to incorporate actor migration and their interaction into a model (Soriano-Paños et al., 2019; Zhang et al., 2015). With the unfettered, sporadic movement of actors (in a society modeled as a social graph with probability distribution of each node observed), a plethora of issues arise, which include(s): (a) nodal exposure to illness (i.e directing, (b) recovery rate, or the amount of time it takes for infected nodes or actors to recover based on adoption level, and (c) diffusion, or the propagation of infection (Macias et al., 2009; Ojugo & Otakore, 2018; 2018; 2020; 2020; 2021).

The prevalence rate of the covid-19 illness with underlying tuberculosis on nodal interaction via a cluster exposure was modeled in our study using a Bayesian network in order to forecast the propagation rate with the arrival of seed-node on a susceptible-infect social graph.

MATERIALS / METHODS

Review of Tuberculosis as Underlying Health Complications
Among a plethora of diseases, communicable widespread viruses are the main danger, as they often yield great calamity to our society’s population (Oladele et al., 2024). Through mathematical modeling, these kinds of contagious epidemic diseases can be controlled and
their effects on society diminished (Najm, 2009). Tuberculosis, has been a leading infectious killer to society’s population as it is currently the second most common infectious disease worldwide (Bruinen de Bruin et al., 2020; Hurt, 2019; Said et al., 2023), and the pathogen Mycobacterium tuberculosis is the cause of it (Khan, 2018). When an infected individual or carrier coughs or sneezes while actively ill with tuberculosis, small droplets from their cough or sneeze are inhaled and spread the Mycobacterium tuberculosis infection (Og & Ying, 2021). Its symptoms have been known to affect the lungs as pulmonary tuberculosis (Oyemade et al., 2016), or other bodily regions, such as extra-pulmonary tuberculosis (Mustofa et al., 2023). The Mycobacterium tuberculosis now affects one-third of the world’s population. Additionally, a greater number of tuberculosis patients led the World Health Organization (WHO) to designate the disease a worldwide emergency in 1993.

A scientific model is often a prerequisite to help comprehend the propagation in transmission dynamics of tuberculosis – so as to help experts plan on the means to restrain its spread (Bhavani & Mangla, 2023). With over 34-million persons estimated to have tuberculosis in 2021 – India yields 27% of the population, China with 9%, Indonesia with 8%, Philippines with 6%, Pakistan with 5%, Nigeria with 4%, Bangladesh with 4%, and South Africa with 3% (Vågsholm et al., 2020). The short-course directly observed therapy (DOTS), which consists of four first-line medications, such as isoniazid (INH), rifampin (RIF), pyrazinamide (PZA), and ethambutol (EMB), given daily for two months before INH and RIF treatment is administered for an additional four months, is no longer available to the World Health Organization (WHO) (Voke et al., 2023).

The Corona-Virus as Co-Infection: The Nigerian Frontier

The coronavirus illness, or COVID-19, first appeared in Wuhan, in the Hubei region of China, in December 2019 (Eranga, 2020; Roshan, 2022). About 210 countries have been affected by this epidemic worldwide. The virus has found new epicentres since this incidence. These are Germany, France, Spain, Italy, and the United States of America. Given the potential harm this virus poses to public health, the World Health Organization (WHO) has classified it as a global pandemic and is urging governments and health sectors worldwide to take it seriously (Crawford et al., 2020; Kolawole et al., 2022; Nilam et al., 2020; Osasume, 2021; Ufua et al., 2021). The use of hydrochloroquine and other medications that have not yet undergone clinical testing was authorized (Chen et al., 2022; Telenti et al., 2021).

There were doubts by individuals as to how COVID-19 is communicated. The researchers continued by stating that symptoms could include dyspnea, sore throat, dry cough, shortness of breath, and, in the worst case scenario, fatal pneumonia. Covid-19’s nurturing phase lasts for two to fourteen days(Ampatzidis et al., 2020). Its symptomatic and asymptomatic character has also been linked to its significant global spread, as widespread screening and testing have not been able to accurately identify and manage these symptomatic and asymptomatic patients, continuing to put everyone at risk (Antia & Halloran, 2021; Tsuwa & Yandela, 2021), which will continue to also significantly impact the way of life of the people, and business across the globe (Haryani & Maryono, 2023; Leo et al., 2017; Patrinos et al., 2022; Santura et al., 2021).

No doubt, the world is still stressed with the realities and new norms that were forced on her by covid-19 pandemic – in lieu of the huge economic and financial losses to thousands of businesses globally, shutdown of physical infrastructure, with long, medium and short-term costs and penalties (Khaki & Wang, 2019; Suruliandi et al., 2021). This huge loss in finance can be outrightly attributed to government’s compulsory order to shutdown all business
Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network

operations to avoid human clustering (Ibor et al., 2023). Nigeria was not left out as the lockdown featured high level of restriction on human mobility, good and services as well significantly disrupted outputs and exportations, which in general, hindered business growth to undercut many investments resulting to investors’ huge loss, as well as confidence in the market (Eboka & Ojugo, 2020). Due to the suspension of business operations against this backdrop, there were 318 confirmed cases, 70 patients who had been released, and 10 cases that had been declared in Nigeria by April 11, 2020 (Ojugo & Eboka, 2018; 2020). Government then took radical steps of lockdown with a view to curb the epidemic’s spread (Ihama et al., 2023).

Data Gathering and Collection
We acquired data from the Asaba Federal Medical Centre Epidemiology laboratory located in Asaba, Delta State. In the data collected, there were 4,687 instances with 54 attributes, including the patient’s demographics, results from HIV and other tests, symptoms the patient experiences, their medical history, the diagnostic tools used, the treatment plan that included dosages and regimens specific to the disease type, as well as the drug reaction, follow-up results for the duration of the treatment period, costs, and hospitalizations paid. However, characteristics that could influence a patient's conduct throughout therapy (the treatment outcome could be anything from cured to failed to discontinued to loss of life to transferred out). The data set was organized into groups (Chibuzo & Isiaka, 2020; Durojaye et al., 2015) these are: (a) Patient-related characteristics (age, sex, etc.); (b) regimen-related characteristics; (c) proximity to a health facility-related characteristics; (d) treatment-related side effects; and (e) treatment duration-related characteristics.

A major task in machine learning requires gathering of the right data set with the right format – to effectively, train the model with correctly labeled cases. This helps minimize classification error. However, many observed data sets are rippled with missing data or with unneded data fields. Thus, researchers end up with an imbalanced data set. This implies that cleaning must be done and selected parameters from the data set tuned to yiled an optimal solution. In order to prevent data-type mismatches when end users try to encode the data during the preprocessing stage, the data set must be prepared to suit the model properly. Consequently, the model has received enough training to classify the incoming data into the appropriate classifications (Ojugo, Akazue, et al., 2023; Ojugo, Ejeh, et al., 2023).
Clustering Bayesian Network
A directed acyclic graph that has been successfully implemented in numerous applications is the Bayesian network (BN) (Adishi et al., 2022). This is hinged on Bayes theorem of conditional probabilities and Belief Network. The denotation $P(B, A)$ indicates the probability for which node $A$ is connected to node $B$. The network's ability to tolerate incomplete and noisy data is crucial in this case because it aims to provide a good performance graph even with bulky attributes (Ojugo et al., 2015; 2015; 2015). As such, it is a directed graph that shows a conditional probability table (CPT) for the parent nodes' occurrence. The probability of an event $A$ depending on another event $B$ is represented by the expression $P(A | B)$. This has to do with how much faith it has in the plausibility of an event in light of lacking information. It suggests that: (a) events $P(A | B)$ and $P(B | A)$ are related; (b) $P(A | B)$ is calculated using information about $P(B | A)$; and (c) the outcome updates the conditional probability of an event based on fresh data. And it yields the Equation 1:

$$P(A_k | B) = \frac{P(A_k, P(B | A_k))}{P(A_1, P(B | A_1) + P(A_2, P(B | A_2) + \ldots + P(A_n, P(B | A_n))} \tag{1}$$

Proposed Experimental Cluster Bayesian Network
The Conditional Probability Table was used to train our model (CPT). This suggests that the algorithm builds distribution probability distribution tables for all associated nodes on the network in order to learn the structure and training data-labels (or accomplishments of interest).

This is accomplished using two methods: multinominal estimator-based probability distribution learning and structured learning or accidental discovery utilizing K2, hills climbing, or Tabu-Search. If learned, the model commences its testing. The parameters with probability distribution were chosen in order to apply this model to the prevalence of the underlying illnesses in relation to the classification of the COVID-19treatment. This resulted in the appropriate stochastic results for the relevant underlying achievements. To do this and correctly classify the data-points, the supervised learning approach for Bayesian network building was applied, as shown in Figure 2 below.

![Figure 2: Target Prediction Model Construction](image-url)

Using a matrix that was created from the input data, the model's top eigen vectors allowed it to convert the clustering problems into graph cut problems (Behboud, 2020; Ojugo & Nwankwo, 2021). The properties of the data points clustered around the cutting graph technique so that sparse points were in distinct clusters and densely packed points were in
Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network

the same cluster. Equation 2 provides the least cut formula, where wij represents a vertex's degree from i to j. Ncut and ratio cut are the two modes of balanced cuts. Furthermore, clustering is a two-cut technique relaxing. It is a successful partitioning technique that incorporates the Ng-Jordan-Weiss and normalized cut procedures, which are also spectral partitioning problem relaxations. (Ojugo et al., 2014; Ojugo, Eboka, Yoro, et al., 2015) with the listing described as in algorithm 1.

\[
\text{cut}(x,y) = \sum_{i \in A, j \in B} W_{ij} \quad (1)
\]

Algorithm: Cluster Bayesian Network Algorithm

**Input**: Data set, clusters k, parameter s and number of iterations iter

**Output**: the set of k clusters

1. Calculate the affinity matrix \( A \in \mathbb{R}^{n \times n} \) then define \( A_{ij} = \exp\left(-\frac{\|s_i - s_j\|^2}{2\sigma^2}\right) \) where the original points data are \( s_i \) and \( s_j \) as well as i and j, respectively */

2. If i ≠ j, then the Aii = 0

3. The matrix of diagonal degrees, D, is calculated using the elements:

\[d_i = \sum_j A_{ij}.\]

In light of this, the Laplacian matrix for the graph G with n input nodes or vertices is

\[L_{n \times n} = D - A\]

4. Locate the k-largest eigenvectors of matrix L and \([x_1, x_2, \ldots, x_k] \in \mathbb{R}^{(n \times k)}\)

5. After renormalizing each row of x, create the matrix Y as follows:

\[y_{ij} = \frac{x_{ij}}{\sqrt{\sum_j x_{ij}^2}}\]

6. Reduce the distortion of each row Y such that it can be viewed as a point in the team's clustering.

7. When the rows of yi are part of cluster j, the original point si is allocated to that cluster.

8. Provide back the cluster center and the collection of k clusters.

**RESULTS AND DISCUSSION**

**Performance of Memetic Algorithm**

To compute model accuracy – we evaluated its performance using Equation 3 to yield the confusion matrix as in figure 3 (Akazue, Asuai, et al., 2023; Akazue & Omede, 2023; Obruche et al., 2024; Ojugo, Oyemade, et al., 2015; Okperigho et al., 2024).

\[
F1 \text{ or Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)
\]
Figure 3 yields a performance of 93.7% accuracy, and this agrees with (Barlaud et al., 2019). Results as in Table 2 shows the 5-classes (as confirmed, discharged, absconded, death and treated). With 30-rules generated, 24-of-the-30-rules were correctly classified in the class 0 (Gao et al., 2021; Ojugo, et al., 2015; Ojugo, et al., 2015; Ojugo & Yoro, 2020; Yuan & Wu, 2021; Zareapoor & Shamsoimoali, 2015). Model also correctly identified benign true-positive instance; But in the class 1, the remaining six rules out of thirty were wrongly categorized and flagged as false-positive (Yoro, et al., 2023; Yoro, et al., 2023).

RESEARCH FINDINGS
A Bayesian Network (BN) with 5-fold cross-validation was utilized to effectively evaluate the model based on all correctly classified examples as indicated in the confusion matrix with the 5-classes (a,b,c,d,e) representing the various treatment groups. A confusion matrix shows the rules’ accurate classification and true/false classification. Two cases (from b to e) were correctly categorized as false in class (a), according to Table 2, while a confirmed class of 2568 cases and a discharged class of 1134 cases were reliably classified as true. There are no discernible differences between classes (a) and (b), respectively. As a result, the classification error has little bearing. But for the proposed cluster Bayesian model, the overall software-derived percentage (instances correctly identified) is 93.7563 percent. (Odiakaose et al., 2023).

Table 2. The model's Confusion Matrix

<table>
<thead>
<tr>
<th>class</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>a=confirmed</td>
<td>2568</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>b=discharged</td>
<td>1134</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>c=death</td>
<td>13</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>d=active</td>
<td>409</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>e=absconded</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

CONCLUSION
In order to anticipate the possible spread of the COVID-19 disease among patients with underlying disorders, in particular, we created a target model using the Bayesian Network. The medical data collection was modified from the Federal Medical Centre in Asaba, Delta State, Epidemiology laboratory.

The model's output defied distinct assumptions based on the probability distribution of attributes, yielding better performance and an acceptable degree of accuracy. This suggested model would be useful to medical providers in assessing and calculating the risk of treating COVID-19 instances that test positive for smears. When resources are scarce in a health care system, the model can be a practical and economical tool. Additionally, the Bayesian Network is one of the greatest predictive tools available for building models to handle problems in domains other than component-based system dependability prediction.

REFERENCES


Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network

https://doi.org/10.2139/ssrn.4001431


Laddha, S., Mnasri, S., Alghamdi, M., Kumar, V., Kaur, M., Alrashidi, M., Almuhaimeed, A.,...


Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network


Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease using a Susceptible Infect Clustering Bayes Network

https://doi.org/10.5815/ijeme.2021.02.05


Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network


Co-Infection Prevalence of Covid-19 Underlying Tuberculosis Disease Using a Susceptible Infect Clustering Bayes Network


