Big data analytics on the impact of OMICRON and its influence on unvaccinated community through advanced machine learning concepts

Amalraj Irudayasamy1 · D. Ganesh2 · M. Natesh3 · N. Rajesh4 · Umi Salma5

Received: 25 February 2022 / Revised: 13 June 2022 / Accepted: 27 June 2022
© The Author(s) under exclusive licence to The Society for Reliability Engineering, Quality and Operations Management (SREQOM), India and The Division of Operation and Maintenance, Lulea University of Technology, Sweden 2022

Abstract New SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus Type-2) variant termed to be “B.1.1.529” subtype mutation, which is a primary concern, might heavily influence further transmission, virulence and even affect the functioning of test methods and efficacy medications (vaccines). It is still not clear on the timeline for the Omicron (B.1.1.529) subtype to develop protective immunity or even when normal activities will rebound in our everyday lives. Computational analysis on the available big dataset of the Omicron variants’ and their effects on the unvaccinated population indicate that the concerned variant seemed to have a stronger propensity for the vulnerable group (unvaccinated community). In consequence of the terrible COVID-19 epidemic, scientific research on vaccine development and their future enhancement throughout the world have been stepped up significantly. We assessed approved vaccines’ effect on morbidity, hospital stays, and fatalities worldwide. Through available big datasets, an Ensemble learning strategy was used to estimate the likelihood of an unvaccinated person contracting a virus. Overall incidence rates dropped from 18.56 per cent to 2.8 per cent for the vaccinated community during the observation period. People ≥ 60 years had the most remarkable percentage drop (~ 15 per cent). In addition, about 70.4 per cent, immunization through vaccines lowered the number of hospitalizations (both ICU and non-ICUs) and fatalities. Through our research observations, the facts clear that immunization through vaccination has a significant influence on decreasing COVID-19 rapid transmission, even if it provides only a modest level of protection. However, to accomplish this effect, non-pharmaceutical therapies must be maintained indefinitely.

Keywords Machine learning · Ensemble learning · Omicron · Vaccination · Fatality · COVID-19 · Accuracy

1 Introduction

On November 26th, in the year 2021, the World Health Organization (WHO) identified the variation B.1.1.529 called Omicron, a lineage of COVID-19 (World Health Organization, 2021). Within South Africa, the Omicron strain has caused an increase in the number of individuals who have got diagnosed with COVID-19 (SARS-CoV-2). As a result, provincial medical researchers worldwide are presently investigating possible complications.

There is a dearth of information on the viral factors that may accelerate its spread. And over thirty mutations in this
variant of severe acute respiratory type result in amino-acid alterations in the spiked structure, fifteen of which are situated in the RBD (Receptor Binding Domain) that is critical for virion contact driven by the ACE-2 (Angiotensin Converting Enzyme-2) ligand. From the sequencing of the Omicron spiking protein, it has been envisaged to estimate the dissemination frequency. These observations revealed a set of mutations that seems to have the potential to increase viral pathogenicity.

Professionals cannot adequately predict the diagnosis of Omicron sufferers until late in the course of the medical condition. A patient’s health may suddenly worsen to a vulnerable predicament throughout the period of Omicron prevalence, which might surprise even the finest knowledgeable physicians. AI-based models may still be helpful because ML algorithms can find complicated trends in big datasets, which the average intellect cannot do (Shilo et al. 2020; Yu et al. 2018). When it comes to fighting Omicron consequences, different AI methods were used, including epidemiological research (Bird et al. 2020; Wang et al. 2020) customized prognosis (Oh et al. 2020; Mei et al. 2020), mortality assessment (Liang et al. 2020; Barish et al. 2020), and more. Although numerous Omicron diagnostic approaches have already been suggested (Wynants et al. 2020), no systematic investigation has analyzed and equated the predictive value of suspected and unsuspected characteristics.

Designing a fatality predictive algorithm based on the regular clinical evidence from the admission/diagnosed date, laboratory evaluations, and demographic information was one of the three critical goals of this research. In the next step, it is hoped that vaccinated, unvaccinated, and partially vaccinated suspected patient characteristics can be used to determine conditions and mortality outcomes in the future. Finally, to give a meaningful evaluation among vaccinated, unvaccinated, and partially vaccinated suspected characteristics in terms of fatality, severity, and mild symptoms (asymptotic) predictive capacities. Clinical evidence from the admission/diagnosed date, laboratory evaluations and demographical information were analyzed for each suspect. An investigation of the predictive capability of such characteristic sets was conducted employing multiple base learners, SVM (Imtiaz Khan et al. 2021), K-NN (Peterson 2009), and Random Forest (Ong et al. 2022). The predictive outcome of each base learner is then again operated via the stacking ensemble technique, which refines the process and produces fine tuned predictions with expected accuracy. After being admitted to the hospital (within 48 h), the majority of Omicron sufferers had their primary exacerbation phase (Wu & McGoogan 2020). Because of this, we used information from day one of a sufferer’s hospitalization to build an optimal model which could be used in reality.

The Ensemble learning method is analyzed empirically in this study. Along with predefined datasets, data from previous research is used to predict the number of Omicron sufferers for the upcoming years. So, our research aim is to.

- To better plan for the worst-case situations, this research provides a methodology for quickly estimating the incidence rate and proportion of new cases of Omicron.
- Through available big datasets, an effective learning strategy needed to be developed and used to estimate the likelihood of an unvaccinated person contracting a virus.
- Finally, to assess the approved vaccines’ effect on morbidity, hospital stays, and fatalities worldwide.

### 2 Emersion of omicron

The most significant emerging SARS-CoV-2 mutant was provisionally revealed from a sample acquired on the 9th of November 2021, but subsequently, on the 26th of November, it was first termed as B.1.1.529 variant as a VOC, namely “Omicron” (World Health Organization 2021). Among the COVID-19 variants, Omicron has the maximum mutations (almost 50 till now), making it of particular interest as well as risky because of the infectious spike-S peptide area’s which is modified around 26–32 times (World Health Organization 2021), fifteen of which occur in the RBD sector (European Centre for Disease Prevention and Control (ECDC) 2021a). As soon as the Omicron was declared a variant of concern in late 2021, countries worldwide began to report instances, with the majority of those incidences being linked to immigration to regions like Europe and Australia (Maxmen 2021).

Omicron has previously been diagnosed in numerous foreign tourists who visited all across the world (Maruki et al. 2022). Omicron has already been detected in 150 nations and territories involving approximately 5 K hundred confirmed incidences and more than 100 fatalities as of 22nd January of 2022, after the speedy dissemination of this novel form (European Centre for Disease Prevention and Control (ECDC), 2021a; Mohapatra et al. 2021). It has been discovered that the active (instantaneous) dissemination rate of the Omicron variation is around 3.19-fold higher than other variants. As a result, a fast rise in Omicron instances may be detected in the near future due to its significant characteristic of increased infection rates and prevalent risk analysis (Kumar et al. 2021). The evolution of COVID-19 is diagrammatically depicted in Fig. 1.
2.1 Omicron characteristics

There were four significant epidemics of COVID-19 from an earlier stage of 2020, and their proportion of confirmed instances after the outbreak. According to epidemiological research, roughly three months after the Beta variant’s breakout, the incidence of infection had surged to 50 per cent of the overall of all diagnoses. Over the same time span, meanwhile, the risk of transmission of the Delta variation increased to almost 80%, reflecting faster infectiousness across the world than the Beta mutant. On the other hand, in South Africa, the rate of Omicron spread exceeded 85 per cent after around 25 days. A fast growth curve of Beta is estimated to be 1.7, whereas for Delta, it is determined as 1.5, and for Omicron, 1.2 days (Karim & Karim 2021). According to these findings, the Omicron strain is potentially more contagious than other variants. According to a recent systematic review of epidemiological investigation (Pulliam et al. 2021), Omicron has been consociated with a higher incidence of severe acute respiratory reinfection in South Africa. Therefore, a fresh Omicron outbreak in South Africa, or perhaps even globally, must not be underestimated.

2.2 Vaccine and its effectiveness

The far more viable method of preventing and controlling Omicron has been shown to be Vaccination. The human immune response may be trained via all immunization platforms. Whole Virus and Component Viral Vaccines seems to be two types of COVID-19 Vaccination. Different kinds of Vaccination are depicted diagrammatically in Fig. 2.
2.2.1 Component viral vaccines

- **Protein Subunit**: Protein complexes comprise a singular protein chain and additional biological macromolecules that combine concurrently (“co-assemble”).
- **DNA-based** and **RNA-based**: The virus includes genomic elements (like mRNA) that serve as a blueprint for highly contagious protein production.
- **Virus-like Particles (VLP)**: Proteins that imitate the virus’s architecture but lack genomic elements.
- **Replicating Viral Vector**: Includes genetic elements out of a virus that is not harmful and can replicate indefinitely.
- **Non-Replicated Viral Vector**: It encompasses genomic elements from a virus that isn’t harmful but can’t reproduce itself.

2.2.2 Whole virus vaccines

- **Live-Attenuated**: Includes weaker versions of the viral infection (attenuated).
- **Inactivated**: Includes no live viruses; just dead ones (inactivated).

The ‘S’ protein immunotherapy is the mainstay of the present COVID-19 vaccinations that are currently in usage. It’s possible that Omicron is produced by immune response sensitivity, based on the molecule’s 32 peptide alterations, which include three modest reductions and one minor inclusion. Since the variation has a greater potential to resist existing vaccinations, these changes might have an enormous impact. Omicron’s ‘S’ gene mutations significantly influence existing vaccinations throughout the global communities, but it is hard to evaluate the real impact of such transitions precisely. In the first place, a person’s immune system may respond differently to various vaccinations. Antibodies induced through vaccines might vary depending on the respondent’s ethnicity, sex, ageing and other health conditions that they have. In addition, the capability to completely regulate diverse trial circumstances may restrict the trustworthiness of data assessment over demographics.

In order to determine whether a formulated vaccine is safe and efficacious, it must first be tested on a large number of individuals via medical studies. A drug’s clinical testing efficacy is determined by how effectively it functions. In contrast, a vaccine’s effectiveness measures how effectively it works in reality (actual environment). Individuals who have been inoculated (vaccinated) and those who have not are both considered. A vaccine’s performance, in fact, is influenced by a variety of circumstances. Omicron is a new variant that has the potential to alter health-oriented scenarios in the future. Participant numbers and ages are significant factors in clinical studies. The wellness of vaccination recipients is also a concern. New information regarding omicron and its vaccine efficacy is gathered at breakneck speed. According to current optimal estimations, vaccines are thought to be 30–40% successful in reducing outbreaks and 70% successful in reducing serious illness.

The vaccines mentioned in the Table 1 have already been recognized, authorized, approved, certified, awarded emergency usage authorization, or allowed access beyond clinical studies by a governing agency, a domestic agency, or perhaps a similar organization through standard procedure.

An early randomized controlled trial in South Africa found modest indications that the Omicron variation might evade immunogenicity and have a faster rate of transmission, both of which could have serious repercussions. Figure 3 exhibits the Observed cases of COVID-19 variants days since the breakout (Walls et al. 2020). Furthermore, the Omicron variation was shown to be ten times extremely contagious as well as twofold virulent than Delta strain using an AI strategy that was developed and verified using a large variety of empirical data sets.

3 Methodology

The health records of 5000 people (comprises both gender) who underwent treatment during the high incidence of the omicron variant all around the globe were primarily obtained through Kaggle online resource (Omicron Daily Cases by Country (COVID-19 Variant), n.d.). The parameters described in Table 2 were used to diagnose and classify patients based on their severity. The proposed model is evaluated using Tensorflow, which is an ML platform that is publicly available and covers the whole learning process for experimental purposes. The entire dataset is segregated into two divisions for testing and training the model in the proportion of 25:75, respectively, for 500 epochs.

It is crucial to establish which features may operate better for training the model during the maximum likelihood estimation. The complexity of the prediction algorithm is reduced as a result of streamlining it. This research drives to investigate and select essential attributes for predicting mortality and other impacts due to Omicron. To study the most critical aspects (feature extraction) for fatality estimation, Fast Neighborhood Component Analysis (FNCA) (Yang et al. 2012) was used. In order to optimize the efficiency of classification or regression techniques, this non-parametric methodology is used to choose essential features. This method circumvents NCA’s (Neighborhood Component Analysis) drawbacks by using a Frobenius regularization term to minimize the estimated leave-one-out prediction inaccuracy (Lin et al. 2012).

FNCA utilizes a unique likelihood function with lower computing complexity than NCA, which speeds up training.
considerably. To accommodate for small training datasets and high-dimensional datasets, the eventual optimal solution can be expressed as,

$$\delta(l) = \sum_i \sum_{j \in \mathcal{E}_i} \left[ \frac{\varphi}{2} \| l \|_2^2 + \mu_{ij} \right]$$

(3.1)

Normalization parameters, $\varphi$ can be modified via cross-validation in order to get the best results. It is significant to mention that such a parametric component is eliminated since $\varphi$ indicates the respective changes, and the eventual outcome remains unchanged. The preceding gradient is obtained by dividing $\varphi(l)$ by the conversion matrices $l$. Such gradient is used in the future learning process and is expressed as,

$$\frac{\partial F}{\partial l} = 2 \sum_i \left( \sum_{j \neq i \in \mathcal{E}_i} \mu_{ij} \cdot (a_i - a_j)(a_i - a_j)^T + (\mu_i - 1) \right) \times \mu_i$$

(3.2)

Table 1 WHO approved emergency use vaccines

| Name of the approved vaccines | Commercial name | Type | Number of countries approved | Age limitations |
|------------------------------|-----------------|------|-----------------------------|----------------|
| Novavax                      | Nuvaxovid/Covovax/ NVX-CoV2373 | Protein Subunit (Recombinant, adjuvanted) | 35 | 18+ |
| Covovax                      | Covovax | Protein Subunit (SARS-CoV-2 rS Protein Nanoparticle [Recombinant]) | 03 | 18+ |
| Moderna Biotech              | Spikevax | mRNA (nucleoside modified) | 85 | 12+ |
| Pfizer/BioNTech              | Comirnaty | mRNA (nucleoside modified) | 137 | 5+ |
| Janssen Ad26.COV2-S/Ad26COVS1/INJ-78436735 | Johnson & Johnson | Non Replicating Viral Vector (Ad26.COV2-S [recombinant]) | 106 | 18+ |
| Vaxzeveria from Oxford/AstraZeneca | Vaxzeveria | Non Replicating Viral Vector (ChAdOx1-S [recombinant]) | 137 | 18+ |
| Covishield from Serum Institute of India | Covishield | Non Replicating Viral Vector (ChAdOx1-S [recombinant]) | 47 | 18+ |
| Covaxin                      | Covaxin | Inactivated (Whole Virion Inactivated Corona Virus vaccine) | 13 | 18+ |
| Covilol                       | Sinopharm | Inactivated (Vero Cell) | 89 | 18+ |
| Sinovac                      | CoronaVac | Inactivated (Vero Cell) | 53 | 18+ |

Ensemble prediction is a technique that employs a variety of training and modeling algorithms (base model) to train massive diversified datasets, which further predicts the desired outcome. Finally, a unique overall prediction is generated by aggregating the predictions of every baseline method. The intention of employing the ensemble classifier technique is to decrease the prediction overfitting (generalization error). In addition, this may reduce biased or inconsistency (variance) issues based on the aggregating learners used. In most cases, ensemble approaches provide more precise outcomes than a solitary predictor. For ensembling the baseline classifier, a stacking approach is utilized.
Stacking is the second level of ensemble approach, sometimes referred to as layered generalization, is a strategy for combining various approaches via an additional classification approach. First, ML algorithms are trained using the available training dataset before being used to create a newer dataset. Then, the ensemble ML algorithm uses this unique dataset as an input for training purposes. Figure 4 depicts the proposed framework.

Non-overlapping training data from a huge dataset $M$ can be represented using an efficient data division approach (Salloum et al. 2019). This is an instance of a multidimensional data collection $M = \{d_1, d_2, \ldots, d_n\}$ that includes both vaccinated and unvaccinated ($P$) records, which is usually large. All the identified features are fed as input parameters to the first level of ensemble approach, which comprises base-level classifiers that includes SVM, K-NN, and Random Forest (RF) algorithms. The following subsections provide a short description of all the three base classifiers.

### 3.1 Support vector machine (SVM)

One of the most widely utilized regularized and supervised learning approaches is SVM. A hyper-plane based upon that maximum range among two categories of datasets is frequently used to solve classification difficulties. This is termed "determination border", which segregates two distinct data categories on each side of the feature space (e.g. severe or moderate).

### 3.2 K-Nearest neighbor (K-NN)

According to the K-NN approach, data sets are categorized primarily on their closeness and relationship to similar accessible data points. This approach is non-parametric. This methodology is built on the presumption that comparable data points may be identified nearby. Consequently, it aims to determine the relationship among data points, generally using any distance measure, and after that allocates a class/label depending upon the majority class/label or average.

As the testing sample increases, it becomes least desirable for classifying problems though it is simple and possess features like short computation time.

### 3.3 Random forest (RF)

For regression and categorization applications, random forest is a valuable ML approach. Specifically, the term “forest” refers to a group of statistically independent decision trees that are subsequently integrated in order to minimize variation and provide better precise factual predictions.

The diagrammatical representation of the target class is depicted in Fig. 5, which illustrates the three categories that include fatalities, severe conditions, and mild symptoms due to the contract of the Omicron virus. All the three classes are comparatively analyzed for both vaccinated as well as unvaccinated.

| Features                      | Vaccinated | Unvaccinated |
|-------------------------------|------------|--------------|
| **Gender (%)**                |            |              |
| Male (n = 2500)               | 67.32 (1683)| 32.68 (817)  |
| Female (n = 2500)             | 70.16 (1754)| 29.84 (746)  |
| **Age**                       |            |              |
| 18–25                         | 640        | 643          |
| 25–45                         | 648        | 618          |
| 45–60                         | 612        | 628          |
| 60+                           | 600        | 611          |
| **Clinical symptoms**         |            |              |
| Cough                         | 982        | 1518         |
| Myalgia                       | 487        | 2013         |
| Dyspnea                       | 689        | 1811         |
| Fever                         | 1134       | 1366         |
| Fatigue                       | 567        | 1933         |
| **Comorbiditie**              |            |              |
| Cardiovascular Disease        | 433        | 2067         |
| Hypertension                  | 387        | 2113         |
| Diabetes Mellitus             | 456        | 2044         |
| **Vital signs**               |            |              |
| Blood pressure                |            |              |
| Min, 70                       | 732        | 1768         |
| Max, 120                      | 1567       | 933          |
| Pulse Rate (60 – 100 beats)   | 1894       | 606          |
| Temperature (36.1–37.2)°C     | 2347       | 153          |
| SpO2                          | > 90: 1983 | > 90: 1757   |
|                               | ≤ 90: 517  | ≤ 90: 757    |
| Respiratory rate              |            |              |
|                               | < 30: 1950 | < 30: 2150   |
|                               | ≥ 30: 550  | ≥ 30: 350    |

| Laboratory record             |            |              |
| Blood features                |            |              |
| WBC                           | RBC        |              |
| HB                            | HCT        |              |
| PLT                           | MCV        |              |
| MCH                           | MCHC       |              |
| RDW                           | ESR        |              |
| **Coagulation features**      |            |              |
| PT                            | INR        | PTT          |
| Prominent chemical features   |            |              |
| BUN                           | ALT        |              |
| Cr                            | ALKp       |              |
| AST                           | LDH        |              |
an unvaccinated dataset to analyze the effect of the vaccine on the Omicron variant.

4 Performance analysis

The entire dataset from level-1 are fed into meta-classifier which is split into two dataset (75:25) for training and testing purpose respectively. Accuracy () is determined by calculating the proportion disparity among the predicted (estimated) and real (actual) targets. This is referenced with a deviation cutoff, over which the obtained experimental values is considered accurate. The crucial performance evaluation parameter of the proposed EL approach is utilized to measure if the objective has been achieved.

\[
\alpha = \frac{100}{n} \sum_{i=1}^{n} Z_i, \quad Z_i = \begin{cases} 
1, & \text{act} \_\text{diff} (\mathbb{P}_T - \mathbb{R}_T) \leq 2; \\
0, & \text{Otherwise}
\end{cases}
\]

(4.1)

As seen in expression (4.1), \( \mathbb{P}_T \) denotes the prediction mark, whereas \( \mathbb{R}_T \) indicates the actual impact and \( Z \) symbolizes the hypothetical component that holds the overall difference between the estimated and actual values.

The model was validated using testing datasets of both vaccinated and unvaccinated records. Table 3 reveals that the unvaccinated proportions are more prone to the infection, which states that the vaccination played a crucial role in controlling the contagious disease.

At varying parameters, the Area Under Curve (AUC)-Receiver Operating Characteristics (ROC) serves as an indicator of categorization accuracy. ROC is a confidence graph, and AUC is an indicator of how well the data can be divided into distinct groups. It indicates how well the proposed model can differentiate between various aspects of subjects. The ROC plots are generated by graphing the True Positive Rate (TPR) versus the False Positive Rate. Having an AUC close to 1 indicates that the proposed model has a high degree of separation. An imperfect system with an AUC close to zero indicates that it has the poorest capacity to segregate the desired classes. Thus, the stronger the AUC, the coefficient of determination can identify subjects with and without the virus impact.

Omicron vaccinations lowered the high prevalence of SARS-CoV-2 strains and the related infection. Due to recording delays, we could not assess the effect of the

---

Fig. 4 Proposed frameworks

![Diagram of proposed frameworks](image)

Fig. 5 Targeted outcomes

![Diagram of targeted outcomes](image)
Omicron impact upon COVID-19 colligated mortality with vaccination participation in January'. However, since Omicron became the dominating variant until December', massive gains in incidence rates were documented among immunized and unvaccinated people, resulting in reliable predictions.

Figure 6 reveals the estimated ranges among the targeted classes with the utilization of Clinical Symptoms (CS), Comorbidities (C), and Vital Signs (VS) on unvaccinated datasets. It’s been observed that the unvaccinated aged proportions from the dataset are more likely prone to the infection, and the fatality rate is also higher than the other age groups. As per the observed discrepancies that we determined when testing the concepts among various features, vital signs has significant weightage in deciding the final prediction. In addition, we intended to see if there might be a difference in the range of fatality estimates across models.

In the early stages of observational computation, rudimentary indications of deaths without immunization remained consistent with previous trials protecting against Omicron infection. The Omicron variant and accompanying COVID-19 symptoms are clearly visible in the vaccinated test set analysis. Figure 7 show that vaccine intake decreases the severity of illness effect and death rate.

There has never yet been a large-scale study comparing the probability of Omicron hospitalization after immunization with that of the general population of all age-group. There was a low mortality rate of 1.06 per cent among those 60 and older in this research, compared to a death rate of 3.6% among those in the 25 to 60-year-old age range. Clinical research and real-world investigations have shown that vaccinations are very effective in older adults. It will be possible to determine the vaccine’ long-term influence on the chance of developing severe types of Omicron with a more extended period of follow-up.

Receiver operating characteristic (ROC) graphs had been utilized to properly assess the prediction methods’ effectiveness, regardless of the method’ evaluation criterion. Figure 8 showed that the proposed EL model was able to accurately predict the experiment’ outcome (which is close to 1). No empirical difference was found for baseline classifiers, despite the EL model’ marginally improved performance owing to the accurate resultants. The findings demonstrated the model’ resilience to the study population and a statistically negligible variance among the vaccinated and unvaccinated datasets.

The classifiers were tested and validated with predetermined predictor variables chosen at random out of the whole population of vaccinated and unvaccinated samples in order to examine the dependence of mortality risk factors provided by each predictor. Figure 9 show that the vaccinated group has a minor advantage over the unvaccinated group when it comes to feature descriptor levels.

### Table 3 Omicron impact on both vaccinated and unvaccinated datasets

| Features | Vaccinated | Unvaccinated |
|----------|------------|--------------|
|          | Mild symptoms | Severe condition | Fatalities | Mild symptoms | Severe condition | Fatalities |
| **Gender (%)** | | | | | | |
| Male \((n = 2500)\) | 71.23 | 25.18 | 3.59 | 43.14 | 38.19 | 18.67 |
| Female \((n = 2500)\) | 69.56 | 28.43 | 2.01 | 59.25 | 22.27 | 18.48 |
| **Age** | | | | | | |
| 18–25 | 92.58 | 5.62 | 1.8 | 72.13 | 21.34 | 6.53 |
| 25–45 | 87.37 | 11.12 | 1.51 | 64.15 | 23.92 | 11.93 |
| 45–60 | 85.24 | 12.67 | 2.09 | 79.42 | 13.17 | 7.41 |
| 60+ | 86.58 | 12.36 | 1.06 | 45.18 | 36.12 | 18.7 |

![Fig. 6 Prediction analysis using the features of clinical symptoms, Comorbiditie, and Vital signs on unvaccinated datasets](image-url)
Vaccinated people had a lower death rate, despite a bit of increase in overall severity compared to the unvaccinated.

5 Conclusion

Ultimately, the resulting dataset shows a 97.2 per cent decrease in mortality risk (vaccinated). These findings indicate that immunization has a significant influence on lowering the incidence of Omicron’s serious effects (hospitalization) and fatality rate. When it came to Comorbidities and daily routines, both vaccinated, and unvaccinated patients were quite comparable. During the studies, it was found that both vaccinated and unvaccinated people became infected. People who have been vaccinated are far less susceptible to getting Omicron, which lowers healthcare-associated illnesses and protects the wider community. Overall the proposed EL approach has performed optimally for large scale datasets (big datasets). According to various researches, the resilience induced by vaccination was expected to persist for more than 12 months. Researchers may benefit from the analysis of each vaccination provided in Table 1 for different age groups, which will help them identify the gaps and weaknesses.

![Fig. 8 ROC curve of EL model for three different subsets of features](image8)

Fig. 8 ROC curve of EL model for three different subsets of features

![Fig. 9 Omicron impact analysis on three target class](image9)

Fig. 9 Omicron impact analysis on three target class

**Funding** There was no outside funding or grants received that assisted in this study.

**Declarations**

**Conflict of interest** The authors assure don’t have any conflict of interest.

**Research involving human participants and/or animals** The authors assure no animals and humans involved.

**Informed consent** Not Applicable to this research.
