

Original Article

Drug utilisation patterns & clinical outcomes in hospitalised COVID-19 patients: A geospatial & machine learning approach

Dhruva Kumar Sharma¹, Madhab Nirola³, Mousumi Gupta⁴, Arpan Sharma⁵, Prasanna Dhungel³ & Barun Kumar Sharma²

Departments of ¹Pharmacology, & ²Radiodiagnosis, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok, ³Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Majitar, Sikkim, ⁴Centre for Data and Learning (Research Lead), & ⁵Center for Health Policy, The Asian Development Research Institute, Patna, India

Received March 14, 2024; Accepted March 11, 2025; Ahead of print April 17, 2025; Published *** *, 2025

Background & objectives: Coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has posed challenges in clinical management due to a lack of established treatment guidelines. This study aimed to analyse drug utilisation patterns and identify factors influencing clinical outcomes in COVID-19 patients.

Methods: A retrospective study was conducted on 380 confirmed COVID-19 patients admitted between April and June 2021 at a tertiary hospital in Sikkim, India. Study participants demographics, medications, comorbidities, outcomes, and geospatial data were collected with due approval from the Institutional Ethics Committee. Machine learning classification and regression models were used for analysis.

Results: The Random Forest classification model achieved the highest accuracy of 90.7 per cent and an AUROC score of 0.86. Methylprednisolone use was associated with an 11.4 per cent mortality rate. Geospatial analysis identified significant mortality clustering in the East district for female study participants and in the East and North districts for male study participants, with a Moran's I index of 0.125080 and a z-score of 8.642819, indicating statistically significant spatial clustering.

Interpretation & conclusions: The study provides insights into COVID-19 management practices and outcomes. Machine learning identified relationships between factors associated with mortality, which could be due to advanced disease state, associated co-morbidities or post-treatment issues. Further prospective studies are needed to validate findings and address limitations.

Key words Clinical outcomes - COVID-19 - drug utilisation - geospatial analysis - machine learning

The most infectious coronavirus, SARS-CoV-2, was recently identified as the cause of the coronavirus disease 2019, or COVID-19. It first came to light in

Wuhan City, Hubei Province, China, in late December 2019¹. It rapidly spread across the globe, and on January 30, 2020, a Public Health Emergency of

International Concern (PHEIC) was declared¹. Due to its rapid spread around the world, WHO finally proclaimed the SARS-CoV-2 epidemic a worldwide pandemic on March 11, 2020². The lack of specific antiviral agents for treating COVID-19 led to attempts to use various medication strategies³. However, it was challenging for the health authorities worldwide to curb this pandemic with specific, effective, and safe antiviral drugs. Development of new antiviral drugs would take many years to reach the beneficiaries as new drug development is a tedious, time-consuming process with lurking uncertainty of it being appropriate in all aspects. This led to considering the available antivirals to be tested for novel coronavirus for their efficacy against it and the process of doing so was referred to as drug repurposing or drug reprofiling. For instance, some of the drugs repurposed for the treatment of COVID-19 were ivermectin, remdesivir, and azithromycin⁴. Though chloroquine, an antiprotozoal agent, showed some activity against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), due to its potential for drug-drug interactions, it is currently not recommended for treating COVID-19 patients except for during clinical trials⁵. According to the experts, ivermectin was repurposed against COVID-19 infection after demonstrating a significant reduction in viral load within 48 h of administration. However, ivermectin's antiviral concentration was only reached after a high dosage, and was associated with many negative side effects, such as confusion, depression, ataxia, psychosis, and seizures. This underscored the fact that ivermectin's safety for use in human therapy can only be achieved at the standard dose ($\leq 200 \mu\text{g}/\text{kg}$)⁵. Remdesivir, an investigational nucleoside analogue, is a broad-spectrum antiviral drug with *in vitro* action against RNA viruses. The use of remdesivir for the treatment of COVID-19 infection also had to undergo various phases of approval and restricted use advisories by respective health authorities. Remdesivir was frequently added to and removed from the management guidelines even though numerous trials are still in progress due to the ongoing variability in study results^{6,7}. Such frequently changing recommendations have been a challenge for physicians managing COVID-19 patients with concerns, including medication choices, drug combinations, and safety issues. Hence, one of the objectives of this study was to understand the drug utilisation patterns and their impact on clinical outcomes in hospitalised COVID-19 patients through a best-fit machine learning model. According to the ICMR-COVID-19 National Task

Force/ Joint Monitoring Group, Ministry of Health & Family Welfare, Government of India, patients with comorbidities like diabetes mellitus, liver, kidney, or chronic lung disease, obesity, active tuberculosis, chronic lung, kidney, or liver disease, diabetes mellitus, and other immunocompromised states (like HIV) are considered to be at high-risk patients of severe COVID disease or mortality⁸. These individuals have the worst prognosis and frequently experience worsening conditions, including pneumonia and ARDS. Since SARS-CoV-2 is still a relatively new virus, little information is available. Nonetheless, compared to people without comorbidities, patients with comorbidities experience worsening results^{8,9}.

Recent studies have examined drug utilisation patterns and clinical outcomes in hospitalised COVID-19 patients across different regions. Mustafa *et al*¹⁰ observed that corticosteroids and antibiotics were widely used without strong evidence-based support. Similarly, in Colombia systemic corticosteroids were the most prescribed drugs, with usage patterns varying by sex, age, and geographical region¹¹. A study in California tracked medication use trends throughout 2020¹². In another study, researchers developed a machine learning model based on patient characteristics and clinical states, achieving high accuracy in forecasting critical care needs and mortality¹³. Similarly¹⁴, the authors used machine learning and deep learning algorithms for early prediction of COVID-19 severity using clinical and laboratory data from two Manipal hospitals. Nature-inspired feature selection identified key markers such as C-reactive protein, basophils, lymphocytes, albumin, D-dimer, and neutrophils, achieving 95% accuracy. Explainable artificial intelligence (AI) techniques were used to identify the model's predictions and its potential for deployment in healthcare facilities to provide timely interventions. These studies highlight the diverse approaches to COVID-19 treatment across different healthcare systems and emphasise the importance of evidence-based practices and resource management in addressing the pandemic's challenges.

Materials & Methods

This retrospective cross-sectional study was conducted by the department of Pharmacology, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok, Sikkim. Prior to the initiation of the study, due ethical approval from the Institutional -Research Committee and the medical superintendent

of the concerned hospital (Central Referral Hospital) were obtained. This study was conducted as per the guidelines laid down by the Declaration of Helsinki 1975 and its further amendments. The data of the SARS-CoV-2 positive affected individuals was retrieved retrospectively from the inpatient hospital records, which were available in the medical records department (MRD) of the hospital using a coded and anonymised data collection proforma. Hence, the data collected did not include any individual identifiers in any form.

Study design and population: Despite extensive global research on COVID-19, gaps persist in understanding between drug utilisation, clinical outcomes, and spatial mortality patterns, especially in under-represented regions like Sikkim. Most studies focus on isolated aspects, such as drug efficacy or mortality risk factors, without integrating machine learning and geospatial techniques. This retrospective cross-sectional chart review was conducted in a tertiary care teaching hospital in Sikkim to analyse the drug utilisation practices and adherence to standard treatment guidelines in the management of hospitalised, laboratory-confirmed SARS-CoV-2 positive cases. The study also aimed to identify models that could offer insights into the associations between specific medications, age, gender, comorbidity, and patient outcomes as clinical endpoints (death or discharge).

Inclusion/exclusion criteria: All SARS-Cov-2-positive patients of all genders aged ≥ 18 yr belonging to all ethnic groups admitted to the designated COVID ward of the hospital were included in the study. However, patients under 18 yr and admitted under obstetric case were excluded from the study. The exclusion of individuals under 18 yr was done primarily based on significant differences in the clinical presentation, disease progression, and the differences in treatment protocols of COVID-19 in children and younger age groups. All patients admitted in-between April 26, 2021 to June 26, 2021 (complete enumeration for the period of 2 months) were included in the study. This helped ensure comprehensive data collection, given the small population and limited case numbers and to analyse the drug utilisation patterns without the need for sampling, reducing the risk of selection bias and enhancing the representativeness of our findings.

Study timeline and methodological phases: The timeline for the study was systematically structured

to ensure all the analysis, from data preprocessing to model validation and geospatial visualisation. Data acquisition, preprocessing, and cleaning were completed within four wk, including handling null values and encoding categorical variables. Data analyses were conducted over two wk to identify patterns and relationships. The machine learning workflow, model training, hyperparameter tuning, and testing spanned five wk. This process involved iterative adjustments to identify the best-performing models for regression and classification tasks. Geospatial analyses required two weeks, including the development of heatmaps and computation of Moran's I index. The study was completed within a carefully planned 15-wk period, ensuring comprehensive analysis and validation of findings.

Data cleaning and bias mitigation: The data obtained included demographics (age, gender, ethnicity), length of hospital stay (admission date/discharge date), medications used, and clinical endpoints (death/discharge). The hospital data did not contain the latitude and longitude data, which was then added through the Google API using the address provided in the hospital data. The paper-based records were manually typed and converted into a digital format and were stored in an MS Excel file. The data was then pre-processed and prepared for regression and classification analysis. Initially, null values were identified and removed from the dataset. A total of 380 patients' data qualified to be included in the study and were used for analysis. Comma-separated values occurred when multiple values were entered in a single field, such as medicine used or symptoms, and were separated into individual columns. Categorical variables were encoded using label encoding. In label encoding each unique category or label in a categorical variable was assigned a unique integer value.

Dataset bias posed a potential threat to validation in this study, as the data were exclusively sourced from a single tertiary care hospital in Sikkim, limiting demographic and clinical diversity. To address this, we performed rigorous preprocessing to handle missing values and ensure uniform encoding of categorical variables. Balanced sampling was applied during model training to mitigate class imbalance, particularly for clinical endpoints like mortality. K-fold cross-validation validated model performance across data splits, reducing overfitting risks. Feature importance analysis *via* the Random Forest model minimised the

Table I. The table shows the anatomical therapeutic classification (I, II, and V)

Medicines used	Generic name	Repurposed classification	ATC level 1	ATC level 2	ATC level 5
Afogatran	Dabigatran	Bat4	Dabigatran - B	Dabigatran - B01	Dabigatran - B01ac07
Ivermectin	Ivermectin	A1	Ivermectin - P	Ivermectin - P02	Ivermectin - P02cf01
Dexamethasone/Dexa/ Dexona	Dexamethasone	Bc1	Dexamethasone - H	Dexamethasone - H02	Dexamethasone - H02ab02
Methylpred/Ivepred/ Medrol/Solumedrol/ Predmet/Wysolone	Methylprednisolone	Bc4	Methylprednisolone - H	Methylprednisolone - H02	Methylprednisolone - H02ab04
Remdesevir/Remdes	Remdesivir	A2	Remdesivir - J	Remdesivir-J05	Remdesivir - J05ax21
Faviflu/Flavipiravir/ Fabiflu/Fluguard	Favipiravir	A4	Favipiravir - J	Favipiravir - J05	Favipiravir - J05ax21

ATC, anatomical therapeutic chemical

impact of disproportionately weighted variables. These measures improved the reliability and generalisability of the findings, though future multi-centre studies are needed for broader validation.

Drug utilisation and clinical endpoints: All medications given to the SARS-CoV-2 patients during their hospitalisation were collected and recorded. Medications received by the patients were further classified according to the Anatomical Therapeutic Classification (ATC-I, II, and V) system shown in table I. In this table, we presented a sample of medicine classification data, illustrating the categorisation of a portion of the total medications utilised. The complete list of remaining medicines has been provided in the supplementary file.

Three categories were used, as follows, regarding medication details⁴:

- (i) Medicines repurposed to treat COVID-19 (these included pharmacological agents, including remdesivir, azithromycin, and ivermectin that were being studied or reported to have possible effects against COVID-19).
- (ii) Supportive drugs (these included supportive treatment recommended for the treatment of COVID-19 patients, such as blood thinners, corticosteroids, and antibiotics).
- (iii) Other drugs (these included drugs, which might be useful for COVID-19 patients, such as statins, omeprazole, montelukast, cetirizine, *etc.*).

The evaluation of drug use started on the day of admission and went on until the clinical endpoints of interest were achieved – inpatient mortality or hospital discharge – occurred. The data were analysed to determine the relationship between age, gender, and medication use and the achievement of the clinical endpoint (discharge/death) utilising various mathematical models. Medication used, age, and gender were considered as independent variables, whereas attainment of the clinical endpoint was taken as dependent variable. The dataset was split into two subsets: a training set used to train the models, and a testing set to evaluate their performance. We used 80 per cent of the data for training purposes and 20 per cent data for testing purposes. We used an 80/20 split to maintain consistency with standard practices in machine learning while ensuring enough data for model training to learn the underlying patterns effectively. random forest (RF), multi-layer perceptron (MLP), ridge regression and logistic regression algorithms were used to analyse the data.

Regression and classification analysis: Regression and classification analysis are two important supervised machine-learning techniques used to analyse data and make predictions. The regression model was used to examine the relationship between a dependent variable and one or more independent variables. This model predicted the value of the dependent variable based on the values of the independent variables. Classification analysis was used when the predictive variable was categorical or discrete, with a limited number of classes. In the current study, we performed comparison analyses on regression and classification algorithms. Comparison analysis of machine learning models

provides a best-fit algorithm with the data. These steps ensured the reliability and accuracy of our predictions regarding the relationship between drug utilisation and clinical outcomes. This comparison helped us validate the robustness of the analyses. For regression analysis, we used RF, MLP, and Ridge Regression. In classification analysis, we used logistic regression, RF, and MLP. Mean Squared Error, R^2 value is treated as a quantitative parameter to evaluate and compare the effectiveness of regression algorithms. Accuracy and Area Under the Receiver Operating Characteristic Curve (AUROC) scores validated the classification model. A random forest model was used to identify feature importance values. Feature importance values represent the contribution of each feature to the predictive power of the model for the binary classification task. Feature importance values are usually relative and can be interpreted as the impact of each feature on the model's ability to discriminate between two classes. The feature importance values generated by the model provide insights into the relative contribution of each variable to the model's predictive accuracy. Feature importance considers the combined and interactive effects of all variables to generate the values.

This study employed machine learning models rather than traditional statistical hypothesis testing to analyse drug utilisation patterns and predict clinical outcomes. The analysis was conducted using Python (Scikit-learn, NumPy, SciPy) for machine learning and ArcGIS Pro for geospatial analysis. Model performance was assessed using accuracy, AUROC, Mean Squared Error (MSE), and R^2 values, allowing for a comprehensive evaluation of classification and regression tasks. Feature importance was determined using the RF model to identify key predictors of clinical outcomes. The Moran's I index and z-scores were used to analyse geospatial clustering of mortality. These methods provided an objective, data-driven approach for evaluating drug utilisation and patient outcomes without relying on conventional statistical significance tests. The software used in this study was Anaconda (2023.09-0) for machine analysis and ArcGIS Pro (3.1.4) for Geospatial analysis.

In regression models, the average squared difference between the predicted and actual values in a regression model is quantified by the MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

The lower values of the MSE represent a closer fit between expected and actual outcomes and thus serve as a measure of each model's prediction accuracy¹⁵. R-squared (R^2) quantifies the percentage of the dependent variable's variation that can be accounted for by the independent variables.

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

An R^2 value equal to 1 indicates a perfect fit, whereas 0 indicates that the model is unable to explain any variation¹⁶. Negative values imply that the model is worse than the simple mean.

In classification models, accuracy determines the proportion of correctly classified instances out of the total instances in a classification model. Higher accuracy suggests better overall performance.

$$Accuracy = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (3)$$

AUROC evaluates the area under the ROC curve, representing the trade-off between true positive rate (sensitivity) and false positive rate (1 - specificity)¹⁷.

$$AUROC = \int_0^1 TPR(FPR^{-1}(t)) dt \quad (4)$$

AUROC = 0.5 implies no discrimination, while higher values signify better performance. An AUROC of 1 indicates a perfect classifier.

Moran's I index: Moran's I index is a measure used in spatial statistics to assess the spatial autocorrelation of a dataset, indicating the degree of similarity between nearby observations. It quantifies whether similar values tend to be clustered together or dispersed across a geographic area. The formula for Moran's I involves the summation of products of the differences between each pair of values, the spatial weights matrix, and the mean value of the entire dataset. We have also used spatial autocorrelation in our dataset using Moran's I index¹⁸ to identify if the data is clustered or dispersed.

The equation for Moran's I is:

$$I = \frac{n}{W} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

Table II. Classification and regression models performance

Model classification models	Measure	Values
Random forest	Accuracy	90.7%
	AUROC	0.86
Logistic regression	Accuracy	88.1%
	AUROC	0.7
Multilayer perceptron	Accuracy	81.5%
	AUROC	0.74
Regression models		
Random forest	MSE	0.080
	R-Squared (R^2)	0.144
Multilayer perceptron	MSE	0.089
	R-Squared (R^2)	0.048
Ridge regression	MSE	0.087
	R-Squared (R^2)	0.068

AUROC, area under the receiver operating characteristic curve;
MSE, mean squared error

Here, n represents the number of spatial units or observations x_i and x_j are the values of the variable of interest at locations i and j , \bar{x} is the mean of all observations, and w_{ij} denotes the spatial weight between locations i and j . The index ranges from -1 (perfect dispersion) to 1 (perfect clustering), with 0 indicating a random spatial pattern.

Results

Algorithm performance: We performed comparison analysis on three regression and classification algorithms. Our objective was to establish correlations between the medications administered to patients, their age, and gender, in relation to the endpoint, which signifies whether a patient's outcome resulted in 'Death' or 'Discharge'. This study compared the performance of several machine learning models for classification and regression tasks as shown in table II. For classification, RF, Logistic Regression, and MLP models were evaluated on their ability to predict categorical outcome variables. RF performed the best with an accuracy of 90.7 per cent and AUROC of 0.86, outperforming the other models. For regression problems, RF, MLP, and Ridge Regression models were tested. RF yielded the lowest MSE of 0.080 and highest R-squared of 0.144, indicating it best captured the variability in continuous target variables compared to the other regression models.

Overall, RF emerged as the top performer across classification and regression modelling. It achieved the

highest accuracy and AUROC for classification, as well as the lowest MSE and highest R-squared in regression problems. These results suggest RF to be a robust machine learning approach that effectively handles both categorical and continuous target variables. It outperformed other commonly used algorithms like Logistic Regression, MLP, and Ridge Regression on this dataset.

Visualisations: As per the ATC medicine classification, we classified our data as shown in table I. Top 20 medicines used by the patients during their treatment were identified and demonstrated in figure 1. Similarly, the total patient's clinical severity was determined, and gender-wise categorisation was done, as shown in table III. The total value in table III indicates the total number of patients. Gender-wise numbering was done based on severity category. As per our visualisation in table III, male patients had the highest percentage of death. Also, in the case of comorbidity, stroke, lung diseases, and sepsis had the highest death percentage.

The feature importance values identified using the RF model are shown in table IV. The higher feature importance values indicate that a particular feature has a stronger influence on the model's prediction. In table IV, we have shown the feature importance values from our best-performing model (RF classification model). Table IV shows that methylprednisolone had the highest feature importance value of 0.018907, indicating it was the most influential predictor of outcome. Visualising table IV helped us to understand which medicines were found to be most impactful for predicting whether a patient was discharged or died. Methylprednisolone emerged as the strongest individual predictor of increased mortality risk. The age-wise distribution of death patients shown in figure 2 reveals that mortality was highest among older age groups. The highest number of deaths (7 patients) occurred in the 61-70 yr age group, followed by six deaths in the age bracket 71-80 yr. Only three deaths were reported below the age of 30.

GIS analysis: GIS stands for Geographic Information System and the technology combines computer hardware, software, data, and analytical methods to capture, manage, analyse, and present spatial and geographic data. We used our data to visualise the heatmap for expired male and female patients. The heatmaps shown in supplementary figure 1 and 2 present the area with the highest number of expired female and male patients. The higher colour intensity in the map

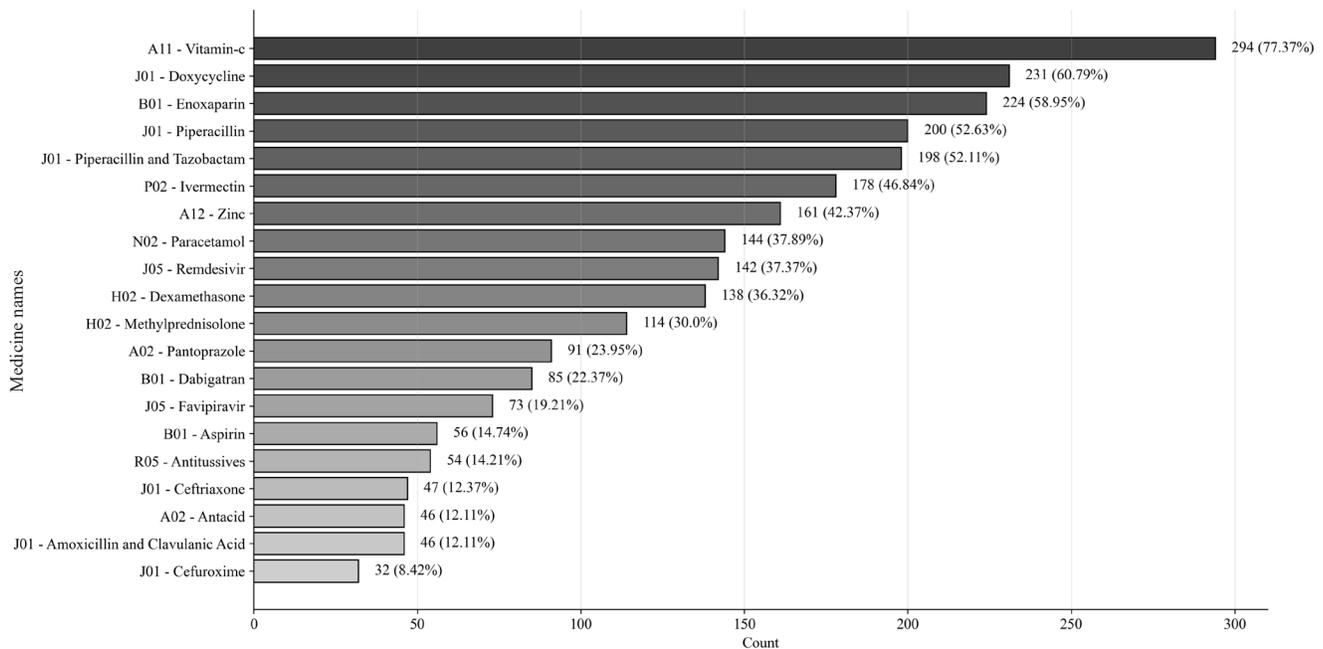


Fig. 1. This figure shows the top 20 highest-used medicines to the study participants during the treatment.

Table III. Study participants' clinical severity characteristics

Item	Mild; n (%)	Moderate; n (%)	Severe; n (%)	Death; n (%)
Total (n=380)	167 (43.9)	166 (43.6)	47 (12.3)	
Gender				
Male (n=227)	86 (37.8)	106 (46.6)	35 (15.4)	21 (56.7)
Female (n=153)	81 (52.9)	60 (39.2)	12 (7.8)	16 (43.2)
Comorbidity in death patients				
Hypertension (n=85)	43 (50.5)	25 (29.4)	17 (20)	9 (10.5)
Diabetes Mellitus (n=73)	39 (53.4)	21 (28.7)	13 (17.8)	6 (8.2)
Lung Diseases (n=10)	3 (30)	2 (20)	5 (50)	4 (40)
Stroke (n=2)	0 (0)	0 (0)	2 (100)	2 (100)
Sepsis (n=1)	0 (0)	0 (0)	1(100)	1 (100)
Kidney injury (n=11)	9 (81.8)	0 (0)	2 (18.1)	1 (9)

Table IV. The above table shows the feature importance values contributed by each feature, resulting in the best algorithm performance *i.e.* values generated by random forest classification model

Encoded value	Medicine names	Feature importance	Number of deaths	Number of discharges	Total study participants took medicine	Death to total study participants ratio
8	Vitamin C	0.012059	24	270	294	0.081633
91	doxycycline	0.012337	20	211	231	0.08658
61	enoxaparin	0.012857	24	200	224	0.107143
18	piperacillin	0.015894	20	180	200	0.1
84	tazobactam	0.015894	18	180	198	0.090909
68	remdesevir	0.0178	15	127	142	0.105634
18	methylprednisolone	0.018907	13	101	114	0.114035

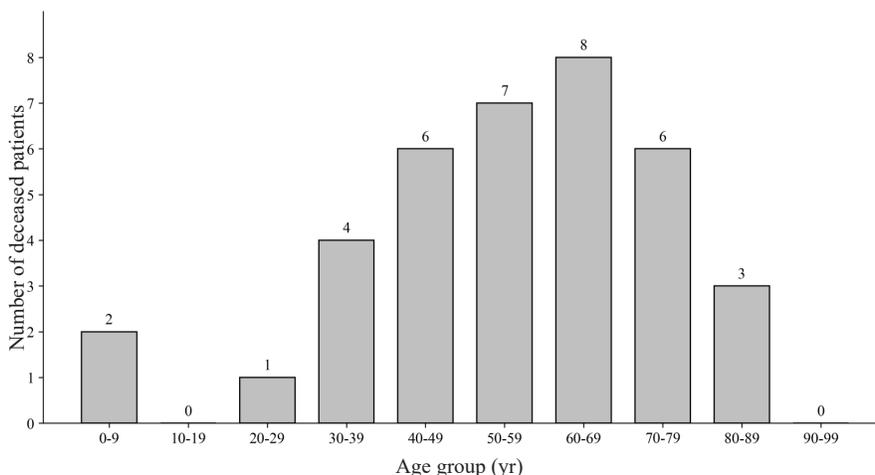


Fig. 2. Age-wise distribution of deceased patients.

means increased density of patients and lower intensity colour means indicates sparse patient density. We can visualise in supplementary figure 1 that female patients from the East district of Sikkim had the highest number of death cases. In the case of male patients, as shown in supplementary figure 2, we can visualise that East district patients, and a lower number of North district study participants had expired. The observed clustering may reflect gender-specific vulnerabilities influenced by healthcare access, comorbidities, or differences in disease progression. These insights suggest the need for targeted interventions, such as improving healthcare infrastructure and outreach programmes, particularly in the East district for females and in the North district for males. The clustering underscores the importance of investigating underlying socio-economic and healthcare disparities that may disproportionately affect certain gender groups in these regions. We also visualised a Spatial Autocorrelation in our data using Moran's Index through ArcGIS Pro. We achieved a Moran's Index value of 0.125080 and a z-score value of 8.642819. These values indicate a clustered pattern in our data with a higher significance level.

Discussion

This retrospective study provides valuable real-world evidence on drug utilisation patterns and their relationships with clinical outcomes among 380 COVID-19 inpatients at a tertiary care hospital in Sikkim, India. Our findings revealed a high utilisation of vitamins, antibiotics, anticoagulants, and steroids, which aligns with other recent studies on COVID-19 hospitalisations¹⁹⁻²¹. The frequent prescription of vitamin C likely stems from its proposed immunomodulatory

effects, although evidence supporting its benefits in COVID-19 remains limited²². The widespread use of steroids raised concerns due to their connections with delayed viral clearance and increased mortality risk, as indicated by certain analyses^{9,23-25}.

The present study provided valuable insights into drug utilisation patterns and their associations with outcomes, highlighting the importance of interpreting these findings within the clinical and logistical context to guide management strategies. A key finding was the link between the use of methylprednisolone, remdesivir, and enoxaparin with increased mortality odds 0.114035, 0.105634, and 0.107143, respectively. However, the associations observed do not indicate causation and must be interpreted within the broader clinical context. The study contributes crucial insights for optimising clinical management, emphasising the need to carefully use corticosteroid, remdesivir, and enoxaparin. A meta-analysis of observational data similarly found corticosteroid therapy to be associated with elevated COVID-19 mortality, especially with long-term use²⁶. Potential mechanisms include exacerbation of comorbidities like diabetes and immunosuppression leading to secondary infections^{27,28}. However, timing, dosage, and patient factors may modify corticosteroid effects²⁹. The highest number of deaths occurred in the 61-70 yr age group, followed by six deaths in the age bracket 71-80 yr. This finding verifies existing evidence that the risk of severe death from COVID-19 increases steadily with age³⁰. Higher age is a major risk factor due to age-related weakening of immune power and increased likelihood of underlying chronic conditions³¹. A systematic review of 45 observational COVID-19 studies revealed significantly elevated

hospitalisation, ICU admission, and deaths above the age of 60³².

In this study, machine-learning models provided insights into predicting COVID-19 outcomes based on drug utilisation and patient characteristics. The random forest classification model performed best, highlighting the importance of selecting appropriate algorithms for different datasets and problems. Prior studies have also effectively applied machine learning to analyse and forecast COVID-19 prognosis^{33,34}. The spatial autocorrelation analysis conducted using Moran's I value of 0.125080 and z-score of 8.642819, provided valuable insights into the geographical clustering of COVID-19 deaths in our region. Such geographic clusters likely arise due to variations in population density, mobility patterns, implementation of preventive measures, healthcare infrastructure, and socioeconomic conditions across districts³⁵. Areas with poor health access may experience elevated morbidity and mortality clustering³⁶. Moreover, densely populated urban settlements can accelerate spread through immediate contact networks³⁷.

Many studies used machine learning and deep learning models to enhance treatment decisions and resource allocation during the COVID-19 pandemic. A study developed machine learning models to identify COVID-19 patients who benefit most from corticosteroid or remdesivir treatment, using data from 10 U.S. hospitals³⁸. Another study³⁹, developed machine learning algorithms to predict ICU admission and mortality in COVID-19 patients using data from 635 individuals. Key predictors for mortality included age, procalcitonin, C-reactive protein, lactate dehydrogenase, D-dimer, and lymphocytes, while ICU admission predictors included procalcitonin, lactate dehydrogenase, C-reactive protein, oxygen saturation, temperature, and ferritin. Like the above studies, other researchers⁴⁰ applied machine learning to hospital data to assist in managing ICU admissions, mortality, and length of stay for COVID-19 patients. Five algorithms, including XGBoost, RF, and LogitBoost, were used, with ensemble stacking boosting predictive performance. This approach demonstrates the potential of ML in optimising patient management during COVID-19 and future health crises.

The current study presents critical clinical insights into the management of hospitalised COVID-19 patients, particularly regarding drug utilisation and patient outcomes. A key finding is the association of methylprednisolone, remdesivir, and enoxaparin use

with increased mortality odds (0.114035, 0.105634, and 0.107143, respectively), underscoring the need for cautious administration of the above drugs, especially in patients with comorbidities. The geospatial clustering of mortality highlights the importance of region-specific healthcare resource allocation and prioritising high-risk area identification through spatial analysis. Integrating machine learning models into clinical workflows can enhance early risk stratification, enabling clinicians to make data-driven treatment decisions. This study identifies key prognostic factors, as table III mentioned, and contributes to the growing evidence supporting individualised approaches in COVID-19 management. However, future research should focus on prospective validation across multi-centre cohorts to improve generalisability and applicability in diverse healthcare settings.

This study had few limitations. Its retrospective design limited causal inferences, and the single-centre dataset might have reduced the study's generalisability. Key variables like socioeconomic status and prior health status were not captured, and the machine learning models relied on a limited feature set. The geospatial analysis identified mortality clusters but did not account for confounding factors such as healthcare disparities. Despite these limitations, the use of machine learning approach in analysing drug utilisation patterns and trends offered valuable insights with wide implications for researchers, clinicians, and policymakers.

In conclusion, this retrospective study provided insights into real-world drug utilisation practices and clinical outcomes in hospitalised COVID-19 patients. The high usage of supportive medications aligned with literature; however, methylprednisolone therapy was linked to higher mortality, warranting careful risk-benefit assessments. Machine learning techniques effectively identified relationships between prognostic factors like medications, comorbidities, age, and gender with clinical endpoints. The RF model demonstrated good predictive performance, highlighting the potential value of such computational approaches. Geospatial analysis found mortality clustering in certain districts with females more impacted in the East district and males additionally in North district. This adds a locational dimension to consider for public health planning. External testing on datasets from other settings can assess model's generalisability. Including variables like socioeconomic status, healthcare access, and vaccination status would enhance robustness, while

longitudinal studies could provide deeper insights into long-term outcomes. Further research building on these methodologies can help enhance prognosis and optimise resource allocation and care.

Acknowledgment: Authors acknowledge Dr(s) Nongmaithem Shivarjit, Shomik Bhattacharya, and Dr. Sagarika Sharma, Intern, Central Referral Hospital for their contribution in collection of the data. Authors would also acknowledge Dr. Amlan Gupta, Associate Director, Histopathology & Transfusion Medicine, Jay Prabha Medanta Super Specialty Hospital for his support and guidance.

Financial support & sponsorship: Authors thank the Sikkim Manipal University for providing the financial support in the form of an Intramural Grant under Dr. TMA PAI Seed Grant ((SMIMS/IEC/2021-45), awarded to first author (DKS).

Conflicts of Interest: None.

Use of Artificial Intelligence (AI)-Assisted Technology for manuscript preparation: The authors confirm that there was no use of AI-assisted technology for assisting in the writing of the manuscript and no images were manipulated using AI.

References

- World Health Organization. *COVID 19 Public Health Emergency of International Concern (PHEIC). Global research and innovation forum: Towards a research roadmap.* Available from: [https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-\(pheic\)-global-research-and-innovation-forum](https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum), accessed on September 14, 2022.
- World Health Organization. *WHO Director-General's opening remarks at the media briefing on COVID-19 11 March 2020.* Available from: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020#:~:text=Pandemic%20is%20not%20a%20word,threat%20posed%20by%20this%20virus>, accessed on September 14, 2022.
- Singh M, de Wit E. Antiviral agents for the treatment of COVID-19: Progress and challenges. *Cell Rep Med* 2022; 3 : 100549.
- Punekar M, Kshirsagar M, Tellapragada C, Patil K. Repurposing of antiviral drugs for COVID-19 and impact of repurposed drugs on the nervous system. *Microb Pathog* 2022; 168 : 105608.
- Hossen MS, Barek MA, Jahan N, Safiqul Islam M. A review on current repurposing drugs for the treatment of COVID-19: Reality and challenges. *SN Compr Clin Med* 2020; 2 : 1777-89.
- Samim MM, Dhar D, Singh V, Nagaraja BS, Shahyan MS, Dey T, et al. CoRe study: COVID-19 and remdesivir: An insight into the current health planning and policy. *J Family Med Prim Care* 2022; 11 : 4671-87.
- World Health Organization and others. WHO recommends against the use of remdesivir in COVID-19 patients. Geneva: WHO; 2020.
- Gupta A, Marzook H, Ahmad F. Comorbidities and clinical complications associated with SARS-coV-2 infection: An overview. *Clin Exp Med* 2023; 23 : 313-31.
- Sanyaolu A, Okorie C, Marinkovic A, Patidar R, Younis K, Desai P, et al. Comorbidity and its impact on patients with COVID-19. *SN Compr Clin Med* 2020; 2 : 1069-76.
- Mustafa ZU, Know CS, Salman M, Kanwal M, Riaz MB, Parveen S, et al. Pattern of medication utilisation in hospitalised patients with COVID-19 in three district headquarters hospitals in the Punjab province of Pakistan. *Explor Res Clin Soc Pharm* 2022; 5 : 100101.
- Valladales-Restrepo LF, Giraldo-Correa JA, Aristizábal-Carmona BS, Constain-Mosquera CA, Sabogal-Ortiz A, Machado-Alba JE. Prescription patterns of drugs given to hospitalised COVID-19 patients: A cross-sectional study in Colombia. *Antibiotics (Basel)* 2022; 11 : 333.
- Watanabe JH, Kwon J, Nan B, Abeles SR, Jia S, Mehta SR. Medication use patterns in hospitalised patients with COVID-19 in California during the pandemic. *JAMA Netw Open* 2021; 4 : e2110775.
- Roimi M, Gutman R, Somer J, Ben Arie A, Calman I, Bar-Lavie Y, et al. Development and validation of a machine learning model predicting illness trajectory and hospital utilisation of COVID-19 patients: A nationwide study. *J Am Med Inform Assoc* 2021; 28 : 1188-96.
- Chadaga K, Prabhu S, Sampathila N, Chadaga R, Umakanth S, Bhat D, et al. Explainable artificial intelligence approaches for COVID-19 prognosis prediction using clinical markers. *Sci Rep* 2024; 14 : 1783.
- Dalton LA, Dougherty ER. Bayesian minimum mean-square error estimation for classification error – Part I: Definition and the Bayesian MMSE error estimator for discrete classification. *IEEE Trans Signal Process* 2011; 59 : 115-29.
- Plonsky L, Hessameddin G. Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *Modern Language J* 2018; 102 : 713-31.
- Hoo ZH, Candlish J, Teare D. What is an ROC curve? *Emerg Med J* 2017; 34 : 357-9.
- Chen Y. New approaches for calculating Moran's index of spatial autocorrelation. *PLoS One* 2013; 8 : e68336.
- Rosenberg ES, Dufort EM, Udo T, Wilberschied LA, Kumar J, Tesoriero J, et al. Association of treatment with hydroxychloroquine or azithromycin with in-hospital mortality in patients with COVID-19 in New York state. *JAMA* 2020; 323 : 2493-502.
- Rawson TM, Moore LSP, Zhu N, Ranganathan N, Skolimowska K, Gilchrist M, et al. Bacterial and fungal coinfection in individuals with coronavirus: A rapid review to support COVID-19 antimicrobial prescribing. *Clinical Infectious Diseases* 2020; 71 : 2459-68.
- Ssentongo P, Heilbrunn ES, Ssentongo AE, Advani S, Chinchilli VM, Nunez JJ, et al. Epidemiology and outcomes

- of COVID-19 in HIV-infected individuals: a systematic review and meta-analysis. *Sci Rep* 2021; *11* : 6283.
22. Hemilä H, Chalker E. Vitamin C can shorten the length of stay in the ICU: a meta-analysis. *Nutrients* 2019; *11* : 708.
 23. Li Q, Li W, Jin Y, Xu W, Huang C, Li L, *et al*. Efficacy evaluation of early, low-dose, short-term corticosteroids in adults hospitalised with non-severe COVID-19 pneumonia: A retrospective cohort study. *Infect Dis Ther* 2020; *9* : 823-36.
 24. Yang Z, Liu J, Zhou Y, Zhao X, Zhao Q, Liu J. The effect of corticosteroid treatment on patients with coronavirus infection: A systematic review and meta-analysis. *J Infect* 2020; *81* : e13-20.
 25. Fernández-Cruz A, Ruiz-Antorán B, Muñoz-Gómez A, Sancho-López A, Mills-Sánchez P, Centeno-Soto GA, *et al*. A retrospective controlled cohort study of the impact of glucocorticoid treatment in SARS-coV-2 infection mortality. *Antimicrob Agents Chemother* 2020; *64* : e01168-20.
 26. Cheng W, Li Y, Cui L, Chen Y, Shan S, Xiao D, *et al*. Efficacy and safety of corticosteroid treatment in patients with COVID-19: A systematic review and meta-analysis. *Front Pharmacol* 2020; *11* : 571156.
 27. Li H, Chen C, Hu F, Wang J, Zhao Q, Gale RP, *et al*. Impact of corticosteroid therapy on outcomes of persons with SARS-coV-2, SARS-coV, or MERS-coV infection: A systematic review and meta-analysis. *Leukemia* 2020; *34* : 1503-11.
 28. Russell CD, Millar JE, Baillie JK. Clinical evidence does not support corticosteroid treatment for 2019-nCoV lung injury. *Lancet* 2020; *395* : 473-5.
 29. WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, *et al*. Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: A meta-analysis. *JAMA* 2020; *324* : 1330-41.
 30. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: Immunity, inflammation and intervention. *Nat Rev Immunol* 2020; *20* : 363-74.
 31. Zhang JJ, Dong X, Cao YY, Yuan YD, Yang YB, Yan YQ, *et al*. Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy* 2020; *75* : 1730-41.
 32. Buikema AR, Buzinec P, Paudel ML, Andrade K, Johnson JC, Edmonds YM, *et al*. Racial and ethnic disparity in clinical outcomes among patients with confirmed COVID-19 infection in a large US electronic health record database. *E Clinical Medicine* 2021; *39* : 101075.
 33. Vaishya R, Javaid M, Khan IH, Haleem A. Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes Metab Syndr* 2020; *14* : 337-9.
 34. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, *et al*. Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020; *369* : m1328.
 35. Nsoesie EO, Brownstein JS, Ramakrishnan N, Marathe MV. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza Other Respir Viruses* 2014; *8* : 309-16.
 36. Kwan M-P. The uncertain geographic context problem. *Ann Am Assoc Geogr* 2012; *102* : 958-68.
 37. Khavarian-Garmsir AR, Sharifi A, Moradpour N. Are high-density districts more vulnerable to the COVID-19 pandemic? *Sustain Cities Soc* 2021; *70* : 102911.
 38. Lam C, Siefkas A, Zelin NS, Barnes G, Dellinger RP, Vincent JL, *et al*. Machine learning as a precision-medicine approach to prescribing COVID-19 pharmacotherapy with remdesivir or corticosteroids. *Clin Ther* 2021; *43* : 871-85.
 39. Hou W, Zhao Z, Chen A, Li H, Duong TQ. Machine learning predicts the need for escalated care and mortality in COVID-19 patients from clinical variables. *Int J Med Sci* 2021; *18* : 1739-45.
 40. Saadatmand S, Salimifard K, Mohammadi R, Kuiper A, Marzban M, Farhadi A. Using machine learning in prediction of ICU admission, mortality, and length of stay in the early stage of admission of COVID-19 patients. *Ann Oper Res* 2022 Sep : 1-29.

For correspondence: Dr Dhruva Kumar Sharma, Department of Pharmacology, Sikkim Manipal Institute of Medical Sciences, Sikkim Manipal University, Gangtok 737 102, Sikkim, India
e-mail: dhruvadoc@gmail.com