

A Random Forest Classifier Model For Predicting The Impact Of Viral Infections On Adults With Chronic Conditions

Martin Muduva
Industrial and Mechatronics
Department
University of Zimbabwe
Harare, Zimbabwe
mmuduvah@gmail.com

Fungai Jacqueline Kiwa
ICT and Electronics Department
Chinhoyi University of Technology
Chinhoyi, Zimbabwe
fungaijacquelinekiwa@gmail.com

Tinashe Kelvin Magaya
ICT and Electronics Department
Chinhoyi University of Technology
Chinhoyi, Zimbabwe
tmagz32@gmail.com

Abstract—this study investigates the impact of viral infections on adults with chronic illnesses, focusing on the development of a Random Forest classifier model. The research aims to predict outcomes among individuals with conditions like diabetes, cancer, and tuberculosis, analysing severity, age groups, and travel patterns. The study aims to assist healthcare professionals in resource allocation and patient prioritization based on disease severity. It reviews literature on viral infection risks for chronic illness patients and explores machine learning applications in infectious disease management. Methodologically, the study adopts a structured approach similar to the CRISP-DM model, integrating correlational and diagnostic research methods. Results indicate that males show higher susceptibility to severe outcomes, with varying infection rates across severity categories. Travel analysis reveals significant virus spread among travellers compared to non-travellers. Older patients exhibit distinct infection patterns. The Random Forest classifier effectively predicts infection outcomes, offering insights for improving healthcare decision-making and response strategies in managing viral infections among adults with chronic conditions.

Keywords—Chronic illnesses, Random Forest classifier, Disease severity, Machine learning applications, Healthcare decision-making, Viral infections

I. INTRODUCTION

Chronic diseases are typically incurable conditions characterized by a complex interplay of factors and a prolonged clinical course [2]; [36]. Viral infections encompass a variety of diseases caused by viruses affecting humans and animals, often resulting in mild to severe respiratory illnesses [35]; [6]; [17]; [53]. Viral infections, characterized by their structure, host dependency, transmission modes, symptoms, immune responses, and treatments, pose significant public health challenges [2]; [36]. Viruses such as SARS-CoV and MERS-CoV, which evolve through mutations adapting to hosts, can lead to severe respiratory illnesses [37]. Mutations in these viruses can enhance transmissibility or evade immune responses, classifying them as variants of concern (VOCs), thereby increasing human health risks. To mitigate the impact of viral infections, governments worldwide are implementing preventive measures such as social distancing, quarantines, and restrictions on non-essential activities for essential businesses and schools. For instance, countries like China, South Korea, and Singapore have effectively controlled viral

infections through rigorous data monitoring, contact tracing, and strict control measures [14]. In contrast, countries such as the United States, the United Kingdom, and France are managing viral infection outbreaks amidst challenges in healthcare resources [14]. Some African countries, like Senegal and Liberia, have responded with proactive measures such as mobile clinics for testing and rapid result delivery, and screening at airports for travellers from high-risk areas [43]. While responses varied, Ethiopia continued flight operations with stringent health protocols, balancing economic ties with health precautions [43]. South Africa, meanwhile, undertook evacuations of citizens from affected regions. Research indicates that Zimbabwe implemented extensive measures including on-site screening, lockdowns, mandatory isolation for returning residents, and public awareness campaigns to curb viral spread [15]; [41]. In the United States, many adults postponed medical care due to viral infection concerns, leading to delays in urgent and routine healthcare services [7]. Similarly, in Singapore, a significant portion of patients with chronic conditions missed healthcare appointments during the viral outbreak [7].

II. STATEMENT OF THE PROBLEM

The Zimbabwean government's response to mitigating the impact of viral infections, such as Influenza, has been crucial, especially in managing patients with chronic illnesses. According to [66], disruptions in prescription refills and routine medical tests during and after the pandemic have significantly affected the management of chronic conditions, potentially increasing mortality rates. This highlights the urgent need to enhance healthcare systems to better handle future viral outbreaks in Zimbabwe. Global pandemics, like the 1918 Influenza pandemic, have historically caused severe respiratory infections and societal disruptions worldwide [65]. The VIRUL INFECTIONS pandemic, similar to Influenza in its respiratory impact, underscores the ongoing risk posed by viral infections and the necessity for proactive preparation and response strategies. Understanding these historical and current viral outbreaks is essential for developing effective public health policies and healthcare infrastructure to safeguard against future pandemics [67].

III. THEORETICAL REVIEW

A. Theory of Multiple Risk Factors

The Theory of Multiple Risk Factors posits that individuals with chronic conditions are at significantly higher risk of adverse outcomes when exposed to additional health stressors, such as infectious diseases. This theory is particularly relevant in the context of viral infections, where chronic conditions such as diabetes, hypertension, cardiovascular diseases, and chronic respiratory illnesses can exacerbate the severity of the infection. Chronic conditions often compromise the immune system, reducing its ability to effectively combat infections. For instance, diabetes impairs the immune response by affecting neutrophil function, reducing the ability to generate a strong initial response to infection, and impairing cytokine production, which is crucial for coordinating the immune response [27]. Similarly, hypertension and cardiovascular diseases are associated with chronic inflammation and endothelial dysfunction, which can exacerbate the severity of viral infections by promoting a hyper inflammatory state and increasing the risk of thrombosis [24]. The presence of multiple chronic conditions can lead to a compounding effect, where the interactions between these conditions and viral infections result in more severe health outcomes. For example, individuals with both diabetes and cardiovascular disease have a higher risk of developing severe complications, such as acute respiratory distress syndrome (ARDS) and multi-organ failure, compared to those with a single chronic condition [63]. This compounded risk is due to the additive or synergistic effects of multiple pathological processes, including chronic inflammation, impaired metabolic control, and cardiovascular stress. Studies have shown that adults with chronic conditions are more likely to experience severe outcomes from viral infections, including higher rates of hospitalization, intensive care unit (ICU) admission, and mortality [47]. The higher prevalence of severe outcomes in this population underscores the importance of identifying and managing these risks effectively. This necessitates sophisticated predictive models that can accurately stratify risk and guide targeted interventions. Random Forest classifiers are well-suited for this task, as they can handle complex interactions between multiple variables and provide robust predictions based on large datasets. The Theory of Multiple Risk Factors highlights the need for comprehensive risk assessment tools that consider the multifactorial nature of health risks in individuals with chronic conditions. By leveraging the capabilities of Random Forest classifiers, healthcare providers can develop predictive models that integrate various risk factors, including demographic information, comorbidities, and clinical indicators, to predict the impact of viral infections on this vulnerable population. This can inform clinical decision-making, resource allocation, and targeted preventive measures to mitigate the adverse effects of viral infections on individuals with chronic conditions.

B. Theory of Data-Driven Predictive Analytics in Healthcare

The Theory of Data-Driven Predictive Analytics in Healthcare emphasizes the transformative potential of advanced data analytics and machine learning techniques in enhancing healthcare outcomes. In the context of the viral infection's pandemic, predictive analytics can play a crucial role in identifying high-risk individuals, optimizing resource allocation, and guiding clinical decision-making to improve

patient care and outcomes. Healthcare systems generate vast amounts of data, including electronic health records (EHRs), diagnostic test results, treatment outcomes, and demographic information. This wealth of data provides an opportunity to gain insights into disease patterns, patient risk factors, and treatment efficacies. Predictive analytics utilizes these datasets to identify trends and correlations that might not be evident through traditional analytical methods. For instance, data from millions of viral infections patients worldwide can be analysed to uncover which combinations of comorbidities lead to worse outcomes, helping to stratify patients based on their risk levels [46]. Machine learning techniques, particularly ensemble methods like Random Forest classifiers, are well-suited for handling the complexities of healthcare data. Random Forest models construct multiple decision trees during training and aggregate their results to improve prediction accuracy and robustness [9]. These models are capable of managing large datasets with numerous variables and can capture intricate interactions between risk factors. For predicting the impact of viral infections on adults with chronic conditions, Random Forest classifiers can integrate data such as age, gender, comorbidities, lab results, and vital signs to generate risk scores and identify high-risk patients. The predictive power of Random Forest classifiers lies in their ability to handle non-linear relationships and interactions between variables, which are common in healthcare data. This makes them particularly effective for developing models that can predict outcomes for patients with complex health profiles. By providing accurate risk predictions, these models can assist clinicians in making informed decisions about patient management, such as prioritizing high-risk patients for intensive monitoring or early intervention. Additionally, predictive analytics can help in the allocation of limited healthcare resources, ensuring that critical care is directed to those most in need [50]. Real-world applications of data-driven predictive analytics in healthcare have demonstrated significant benefits. For example, predictive models have been used to forecast patient admissions, predict disease outbreaks, and identify patients at risk of readmission. During the viral infections pandemic, such models have been employed to predict hospital resource needs, optimize ventilator allocation, and guide public health interventions. By accurately predicting the impact of viral infections on adults with chronic conditions, healthcare providers can better prepare for surges in cases, allocate resources more effectively, and implement targeted measures to protect vulnerable populations (Kumar et al., 2020).

IV. EMPIRICAL REVIEW

During the COVID-19 pandemic, advanced technologies were utilized to manage health data efficiently across various platforms, including sensors and biomedical acquisition systems [32]. Machine learning algorithms, integrated with Internet of Things (IoT) devices, played a crucial role in predicting symptomatic and asymptomatic cases in indoor and outdoor settings, facilitating contact tracing during pandemics [38]. Methods like Mode and Mean Missing Data Imputation (MMM-DI) used IoT devices such as infrared thermometers and thermographic cameras to predict viral infections through recurrent neural networks [38]. Additionally, IoT technologies enabled the collection of vital signs and respiratory patterns, enhancing predictive capabilities through machine learning models like Xtreme Gradient Boosting (XGBoost) [40]. The integration of social media and Internet marketing into large datasets introduced noise, potentially affecting model

accuracy and reliability [23]. Ensuring the validity and effectiveness of AI predictions amidst data variability remains critical [29]. While digital media platforms have enhanced crisis response and information dissemination, they have also amplified misinformation, impacting public perceptions and behaviour during viral outbreaks [58]. Technological innovations, though pivotal in pandemic responses, must complement human decision-making rather than replace it [29]. Disparities in digital access further underscore the need for tailored interventions to ensure equitable healthcare delivery and technological inclusivity [10]. The ethical implications of data use and algorithmic decision-making in healthcare underscore the importance of responsible technological integration [4].

In support to [52] defines AI as the simulation of human intelligence by a system or a machine. This definition is synonymous to [52] asserted that AI is the imitation of all human life by computers. Therefore, it can be concluded that the works of the bottom line is that AI works like humans: Artificial Intelligence is a broad field of computer science that involves the creation of intelligent machines that can perform tasks that often require human intelligence. The main goal of artificial intelligence is to enable computers and machines to perform useful tasks such as problem solving, decision making, understanding human behaviour and communicating. The use of modern smart technologies such AI enables making faster and smarter decisions regarding the business process and resultantly increasing the productivity and profitability [60]. Pattern recognition, text mining and natural language processing among others. The three most prominent intelligent systems are artificial intelligence techniques, machine learning and deep learning these terminologies have been constantly used to refer to each other interchangeably [48]. Machine Learning techniques are mainly divided into supervised, unsupervised learning and semi supervised [59]. Educational supervision should include training on materials available along with the necessary equipment. This training is used to create models for future predictions. In unsupervised learning, similar topics are grouped together in groups called clusters. Each group contains data points with similar characteristics [59] pointed out that educational monitoring is divided into two categories: distribution and recovery. The distribution is used to predict final course grades with positive and negative outcomes based on instruction and grade separation. Classification algorithms are used to denote anonymous data, such as building predictive models based on historical data and using algorithms to correctly assign test data to a group, such as separating apples from oranges or separating spam into separate messages inbox. Whilst regression is a method of using continuous linear functions to correspond to dependent and independent variables (input and output) [59]. Using regression helps in understanding the relationship between dependent and independent variables, the regression models are thus helpful in predicting numerical values based on different data points such as sales revenue projections for a given business [59].

Unsupervised learning refers to the process of analysing unlabelled datasets without predefined outputs [48]. Algorithms in unsupervised learning autonomously group data based solely on their inherent features, without human intervention or labelled examples. This approach is particularly useful for identifying patterns within data, such as in social media platforms where it can automatically suggest friends based on user behaviour, akin to clustering in data

mining [48]. Clustering involves grouping similar data points together, often using algorithms like K-means, which partitions data into clusters based on similarities [48]. Association, on the other hand, discovers relationships between variables in a dataset using techniques like association rules, commonly applied in market basket analysis [48]. Dimensionality reduction techniques reduce the number of data inputs while preserving data integrity, beneficial for managing high-dimensional datasets such as those used in image compression or noise reduction in visual data [48]. Autoencoders, a type of neural network, are used in this context to compress and reconstruct data representations [31]. Despite its utility, unsupervised learning is less prevalent in medical fields due to the necessity of well-structured, labelled datasets. However, it finds application in basic and translational medical research, particularly in clustering techniques applied to high-volume sequencing data [45]. This method leverages a small amount of labelled data alongside a larger pool of unlabelled data to enhance learning efficiency and accuracy [42].

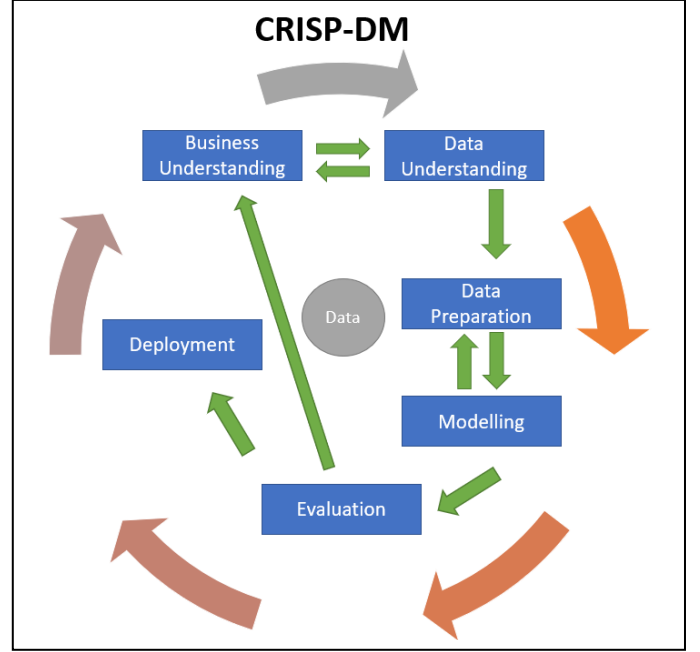
Semi-supervised learning is particularly useful in domains where obtaining labelled data is expensive or challenging, such as in computer-aided diagnosis and drug discovery [42]. In fact, various reports have demonstrated that cardiovascular disorders, hypertension and diabetes are among the most relevant negative accelerating factors for viral infections, together with advanced age, male gender, chronic respiratory diseases and cancer [8]. In a series of 44,672 confirmed viral infections cases from Wuhan, China, the overall case fatality rate was 2.3%, but it was significantly higher in patients with an underlying cardiovascular disease (10.5%), hypertension (6.0%) and diabetes (7.3%); notably, patients with cardiovascular diseases were only 4.2% of total cases, yet constituted 22.7% of all fatal cases [8]. In Italy, which was the first western country severely impacted by viral infections, the fatality rate was notably high, reaching 13% as of April 9, 2020 [64]. This rate escalates significantly with age and is more prevalent among men, who constituted 67% of cases. 96.7% of deceased patients had at least one pre-existing health condition: 14.4% had one comorbidity, 20.5% had two, and 61.9% had three or more [64]. The most common comorbidities observed were hypertension in 70% of cases, diabetes in 32%, ischaemic heart disease in 28%, atrial fibrillation in 23%, chronic renal failure in 23%, chronic lung disease in 18%, obesity in 16%, and stroke in 11% [64]. ACE2 is a membrane protein whose function is to protect lungs, it has been stashed that SARS-CoV2 or coronavirus uses the ACE2 protein to enter cells as the entry point of the viruses into the body that will finally cause cardiovascular toxicity [8]. The use of ACE2 is a risk if the patients is infected with viral infections as this might lead to arrhythmia, myocarditis, acute coronary syndrome, venous thromboembolism, cardiogenic shock and heart failure [8]. In a study conducted by [39] in Asia and Europe, namely from London, USA, China and Malaysia involving 191 hospitalised patients with viral infections in Wuhan, 48% of patients had comorbidities 67% of those who died, 30% of patients had hypertension (48% of those who died), 19% patients had diabetes (31% of those who died) and 8% of patients had coronary heart disease (24% of those who died).

V. METHODOLOGY

A. Cross-industry standard process for data mining (CRISP-DM)

Creswell stated that research design refers to the study plan used to achieve results and provide answers to current research questions [16]. Research design is the comprehensive strategy that extends from broad theoretical perspectives to detailed decisions regarding data collection and analysis. Various types of research designs include correlational, diagnostic, descriptive, experimental, and descriptive research designs. Research involves using a variety of methods to uncover or explain unknown phenomena. The experimental method, a quantitative approach, enables researchers to determine the impact of an intervention on an outcome for a group within a study [16]. Accordingly, this study adopts an experimental research design to systematically manipulate independent variables to determine their effect on a dependent variable. This is executed through the application of a machine learning method to predict the impact of viral infections on patients with chronic illnesses among the sampled individuals. This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, a well-established methodology for conducting data mining and machine learning projects [51]. The CRISP-DM framework consists of six major phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. In this study, the Business Understanding phase involves defining the objectives of predicting the impact of viral infections on adults with chronic conditions. The Data Understanding phase entails collecting and exploring datasets, including patient demographics, medical history, clinical data, and behavioral factors. Data Preparation involves cleaning and transforming data to ensure its suitability for analysis. During the modeling phase, the Random Forest algorithm is applied to develop the predictive model. Evaluation assesses the model's performance to ensure it meets defined objectives. Finally, the Deployment phase integrates the model into clinical practice to help healthcare providers identify high-risk patients and optimize resource allocation. The CRISP-DM methodology ensures a structured approach, guaranteeing the study's scientific rigor and practical relevance in predicting COVID-19 outcomes [56] as shown in Figure 1.

Fig. 1. CRISP-DM [57]



B. Data preparation

Data preparation is a crucial phase in developing a Random Forest classifier model to predict the impact of viral infections on adults with chronic conditions. This phase involves transforming raw data into a clean, organized, and suitable format for analysis and modelling, ensuring the accuracy and reliability of the predictive model. The first step in data preparation is data collection, where relevant data is gathered from various sources such as electronic health records (EHRs), public health databases, and clinical studies. For this study, the data includes patient demographics (age, gender, ethnicity), medical history (presence of chronic conditions such as diabetes, hypertension, cardiovascular diseases, chronic respiratory illnesses), and clinical data (vital signs such as body temperature, respiratory rate, heart rate, and oxygen saturation levels). Behavioural factors, including smoking status and physical activity levels, are also considered.

C. Modelling

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

def evaluate_model(y_true: np.array,
                  y_preds: np.array) -> dict:
    """
    Performs evaluation comparison on y_true labels vs. y_pred labels.

    Returns several metrics in the form of a dictionary.
    """
    accuracy = accuracy_score(y_true, y_preds)
    precision = precision_score(y_true, y_preds)
    recall = recall_score(y_true, y_preds)
    f1 = f1_score(y_true, y_preds)
    metric_dict = {"accuracy": round(accuracy, 2),
                  "precision": round(precision, 2),
                  "recall": round(recall, 2),
                  "f1": round(f1, 2)}
    print(f"Acc: {accuracy * 100:.2f}%")
    print(f"Precision: {precision:.2f}")
    print(f"Recall: {recall:.2f}")
    print(f"F1 score: {f1:.2f}")

    return metric_dict
```

Fig. 2. Modelling.

In the modelling phase of the study, a random forest classifier was used to create a predictive model using data collected from patients diagnosed with viral infections. First, the dataset is divided into training and testing data, with training data (x_{train} , y_{train}) accounting for 80% of the dataset and testing data (x_{test} , y_{test}) accounting for the remaining 20%. This classification allows the model to be trained on the majority of the data while also allowing self-assessment of its performance on unseen data. To prepare the dataset for random forest classifier, medical records of all patients were created as an array, including variables such as infection, medical history, and viral infections results. These systems are designed to capture temporal and temporal patterns important for predicting outcomes associated with severe illness caused by viral infections. The random forest classifier model architecture is constructed using TensorFlow/Keras and includes short-term temporal (LSTM) units due to their ability to learn well and retain long-term data. The model is built sequentially, starting from the LSTM layer to the density layer. Fixed parameters such as segment length and batch size to improve computational efficiency and model performance during training. Study materials. Training has many opportunities to adjust the weight and bias of the model to reduce loss and increase the accuracy of predicting the long-term outcome of the disease after viral infections. After training, the model's performance is evaluated using test data to evaluate its ability to generalize to new data, unseen situations. Metrics such as loss and accuracy were calculated to evaluate the model's performance in predicting chronic disease outcomes based on continuous patient data. Besides RNN, other models such as random forest and stochastic gradient descent have also been developed and evaluated to compare their prediction performance with RNN methods. This comparative analysis provides insight into the strengths and limitations of each model in managing the complexity of the data and predicting outcomes related to disease progression post-infection. The use of RNNs in this study demonstrates their effectiveness in modelling the environment

in continuous medical data, providing a good way to understand and predict the long-term effects of viral infections in patients with pre-existing conditions.

D. Tokenisation

```
import nltk
nltk.download('punkt')
from nltk import word_tokenize, sent_tokenize

# Function to apply sentence and word tokenization
def tokenize_text(text):
    sentences = sent_tokenize(text)
    words = [word_tokenize(sentence) for sentence in sentences]
    return words

df['Complaint_tokenized_text'] = df['narrative'].apply(tokenize_text)

df.head()
```

Fig. 3. Tokenisation.

Data preparation is a critical phase in developing a Random Forest classifier model to predict the impact of viral infections on adults with chronic conditions. This phase involves transforming raw data into a clean, organized, and suitable format for analysis and modelling, ensuring the accuracy and reliability of the predictive model. This process involves collecting relevant data from various sources, including electronic health records (EHRs), public health databases, and clinical studies. The data encompasses patient demographics (age, gender, ethnicity), medical history (presence of chronic conditions such as diabetes, hypertension, cardiovascular diseases, and chronic respiratory illnesses), and clinical data (vital signs such as body temperature, respiratory rate, heart rate, and oxygen saturation levels). Behavioural factors, including smoking status and physical activity levels, are considered. Once the data is collected, the next step is data cleaning, which involves identifying and correcting errors, inconsistencies, and missing values. This may include removing duplicate records, correcting inaccurate information, and imputing missing values using appropriate statistical methods. Ensuring data quality is critical as it directly impacts the model's performance. Data transformation follows, where the raw data is converted into a format suitable for analysis. This step involves normalizing or standardizing continuous variables to ensure they have a consistent scale, which is important for machine learning algorithms. Categorical variables, such as gender and ethnicity, are encoded into numerical formats using techniques like one-hot encoding or label encoding. Feature selection is another essential component of data preparation. This involves identifying the most relevant variables that significantly impact the model's predictions. Reducing the number of irrelevant or redundant features helps improve the model's accuracy and efficiency. Techniques such as correlation analysis, principal component analysis (PCA), and recursive feature elimination (RFE) can be used for feature selection. After transforming and selecting the relevant features, the dataset is split into training and testing subsets. The training set is used to train the Random Forest model, while the testing set is reserved for evaluating the model's performance. A common practice is to use an 80/20 or 70/30 split for training and testing datasets, respectively, ensuring that both subsets are representative of the overall population. Finally, data augmentation may be applied, especially if the dataset is imbalanced. Techniques such as oversampling the

minority class or under sampling the majority class can help balance the dataset, improving the model's ability to generalize across different classes. By meticulously preparing the data, we ensure that the Random Forest classifier model can accurately predict the impact of viral infections on adults with chronic conditions. This comprehensive data preparation process is essential for achieving reliable and actionable insights, ultimately aiding in better clinical decision-making and resource allocation.

VI. RESULTS

A. Relation between Age and Duration of illness

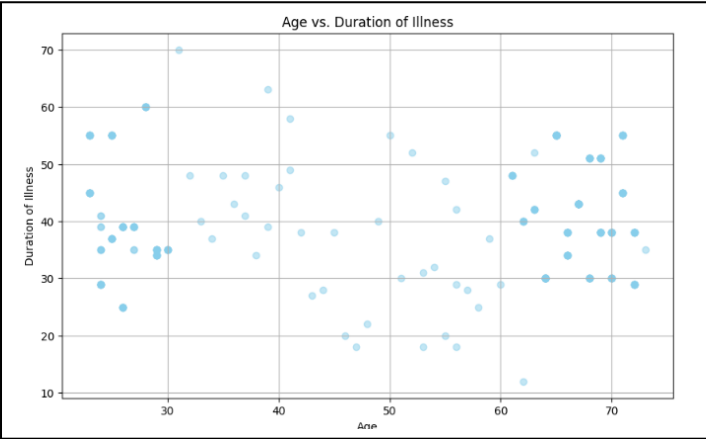


Fig. 4. Age and Duration of Illness

The scatter plot shows the relationship between age of viral infections patients and disease duration in data analysis. Each data point on the graph represents a patient, with age on the x-axis and disease duration on the y-axis. The scatter plot provides important information about how age affects viral infections prevalence and disease duration, as shown in Figure 4 This presentation generated a scatter plot to examine the relationship between the age of patients and the duration of their illness due to viral infections. By visualizing this data, we can gain insights into whether older adults with chronic conditions experience longer durations of illness compared to younger patients. The plot is created using the matplotlib.pyplot library, setting the figure size to 10 by 6 inches for better readability. Each point on the scatter plot represents an individual patient, with the x-axis showing the patient's age and the y-axis showing the duration of their illness. The points are colored in a light blue (skyblue) and are semi-transparent (alpha=0.5), which helps to visualize overlapping points more clearly. Labels for both axes (Age and Duration of Illness) are added for clarity, and a title (Age vs. Duration of Illness) provides context for the plot. A grid is also included to make it easier to read the plot. This visualization is crucial for identifying any potential trends or correlations that could inform the development of a predictive model for the impact of viral infections on adults with chronic conditions.

B. Findings of the study

The study's findings present a detailed examination of the entire research process, encompassing comprehensive data analysis and the development of predictive models. This research underscores the critical importance of assessing the impact of viral infections on patients already managing chronic diseases. The analysis included a dataset spanning from 2020 to 2021, sourced from patient records detailing chronic illnesses such as diabetes, hypertension, and cancer, with viral infections as an additional factor. The dataset encompassed variables like travel history, duration of illness, hospital admissions, recovery times, and the exacerbation of chronic conditions post-infection. Through rigorous modeling techniques, including the implementation of random forest classifier model, the study aimed to predict outcomes and understand the complex interplay between viral infections and chronic diseases.

Fig. 5. Statistical Summary of Patient Data.

	ID	Age	Duration of Chronic Illness	Length of Hospital Stay	Duration of Illnes
count	320.000000	320.000000	320.000000	320.000000	320.000000
mean	147.859375	51.215625	9.087500	18.856250	40.38437
std	76.739282	19.413188	1.298118	4.051157	9.92854
min	1.000000	23.000000	5.000000	3.000000	12.00000
25%	80.750000	28.000000	8.000000	15.000000	34.00000
50%	161.500000	63.000000	9.000000	19.000000	38.00000
75%	215.000000	68.000000	10.000000	22.000000	48.00000
max	266.000000	73.000000	17.000000	30.000000	70.00000

This table presents a comprehensive statistical overview of patient characteristics crucial for developing a Random Forest classifier model to forecast the impact of viral infections on adults with chronic conditions. The dataset encompasses 320 patients, each identified uniquely by an ID. Key metrics include age, duration of chronic illness, length of hospital stay, and duration of illness. The average age of patients is 51.22 years, with a standard deviation of 19.41, indicating a diverse age range among the cohort. Patients typically experience a mean duration of chronic illness of 9.09 years, and they stay in hospitals for approximately 18.86 days on average due to viral infections complications. The duration of illness spans a mean of 40.38 days, highlighting variability in the recovery period among patients as shown in Figure 4.6.1. This statistical summary provides essential insights into the demographic and health-related characteristics that will inform the development of predictive models aimed at improving healthcare outcomes for individuals with chronic conditions affected by viral infections.

```

Acc: 67.19%
Precision: 0.77
Recall: 0.52
F1 score: 0.62
{'accuracy': 0.67, 'precision': 0.77, 'recall': 0.52, 'f1': 0.62}

```

Fig. 6. Model evaluation.

The model evaluation results indicated that the accuracy achieved was 67.19%. Precision, measuring the model's ability to correctly identify positive predictions, was observed to be 0.77. Recall, which assesses the model's capability to correctly detect positive instances, stood at 0.52. The F1 score, which balances precision and recall, was calculated to be 0.62. These metrics, with an accuracy of 67%, a precision of 0.77, a recall of 0.52, and an F1 score of 0.62, provided a comprehensive assessment of the model's performance based on the evaluation conducted

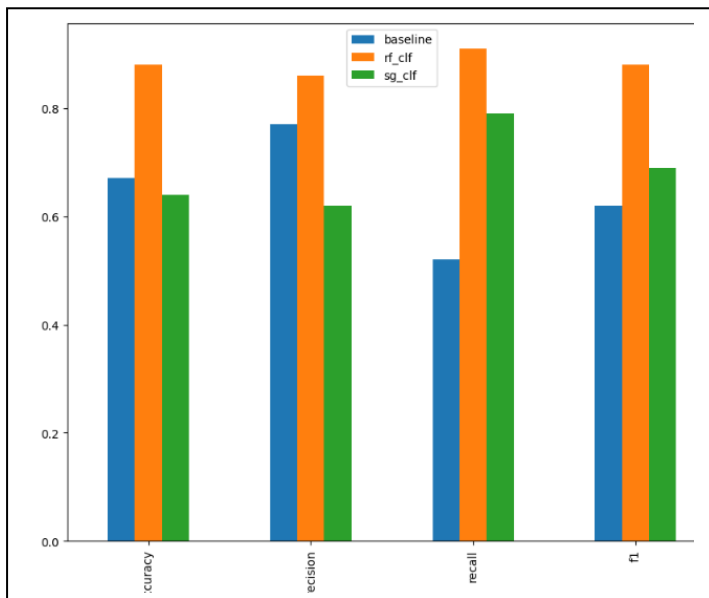
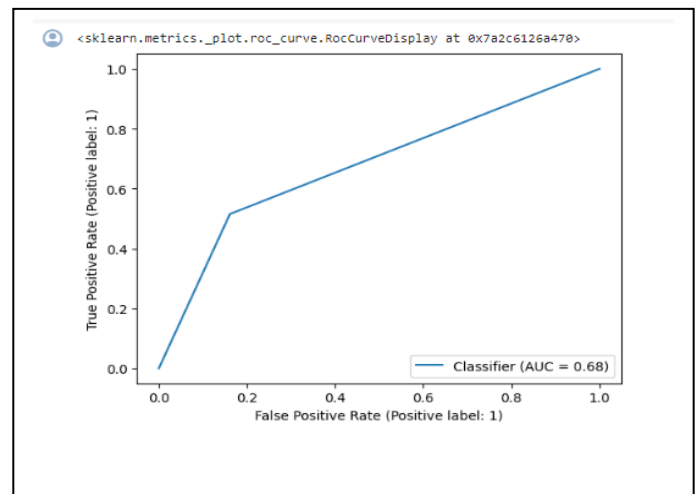


Fig. 7. Algorithm Comparisons.

In evaluating various machine learning algorithms for predicting the impact of viral infections on adults with chronic conditions, the performance metrics reveal distinct differences. The support vector machine (SVM) algorithm achieved an accuracy of 67.19%, with a precision of 0.77, recall of 0.52, and F1 score of 0.62. Comparatively, the stochastic gradient descent (SGD) algorithm attained an accuracy of 64.06%, precision of 0.62, recall of 0.79, and F1 score of 0.69. However, the random forest algorithm outperformed both, demonstrating an accuracy of 87.50%, precision of 0.86, recall of 0.91, and F1 score of 0.88. With an accuracy of 87.5%, the random forest algorithm proved to be the most effective in predicting the impact of viral infections among adults with chronic conditions. Its superior performance across all metrics highlights its suitability for developing a robust classifier model in healthcare scenarios, emphasizing accuracy, precision, and comprehensive recall in forecasting the severity

and implications of the virus on vulnerable populations as shown in Figure 4.6.3. This underscores the efficacy of the random forest approach in enhancing healthcare decision-making and improving patient outcomes amidst the pandemic. In the context of developing a Random Forest classifier model to predict the impact of viral infections on adults with chronic conditions, several machine learning algorithms were evaluated based on their performance metrics. The support vector machine (SVM) algorithm achieved an accuracy of 67.19%, with a precision of 0.77, recall of 0.52, and F1 score of 0.62. This algorithm showed moderate performance in classification tasks, balancing between identifying true positives and minimizing false positives. Similarly, the stochastic gradient descent (SGD) algorithm achieved an accuracy of 64.06%, with a precision of 0.62, recall of 0.79, and F1 score of 0.69. SGD demonstrated higher recall but lower precision compared to SVM, indicating its effectiveness in capturing actual positive instances but with a higher rate of false positives. The random forest algorithm emerged as the standout performer in this evaluation. It achieved an impressive accuracy of 87.50%, precision of 0.86, recall of 0.91, and F1 score of 0.88. This signifies that the random forest model not only accurately predicted the impact of viral infections on patients with chronic conditions but also maintained a high level of precision in its classifications. Its robust recall rate further indicates its ability to effectively capture true positives, making it particularly well-suited for healthcare applications where identifying and managing high-risk patients is crucial. The significant performance superiority of the random forest model underscores its suitability for deployment in healthcare settings. By leveraging ensemble learning and aggregating predictions from multiple decision trees, the random forest model excels in handling complex datasets, such as those involving diverse patient demographics and varying disease severities.

Fig. 8. Receiver operating characteristic (ROC) Curve.



Graphs were placed under the receiver operating characteristic (ROC) curve to evaluate the performance of the prediction model in predicting chronic disease in the population as highlighted in Figure 4.6.4. The ROC curve is a

graphical representation that shows the balance between the true positive (sensitivity) and negative (1-specificity) values of the binary distribution model of the variable. (ROC) Curve: The ROC curve plots the true positive rate (TPR) on the y-axis with the false positive rate (FPR) on the x-axis. Each point on the curve represents a different variable. The diagonal line from the lower left to the upper right midpoint is the random guess, while the curve just after the upper left corner represents the optimal distribution. Excellent performance from every possible angle. It varies between 0 and 1; 0.5 represents a random guess, higher values represent a better guess. AUC-ROC values close to 1 indicate that the model has good performance and can distinguish positive and negative variables. Higher AUC-ROC values indicate better model performance, indicating that the model is able to distinguish patients with and without chronic disease. By plotting the true positive rate against the false positive rate, the AUC-ROC curve provides insights into the model's ability to distinguish between individuals with and without the severe impact of Coronavirus. The AUC-ROC curve analysis evaluates the model's discriminatory power across different thresholds, with a higher AUC-ROC value indicating better discriminative ability. By plotting the true positive rate against the false positive rate, the AUC-ROC curve provides insights into the model's ability to distinguish between individuals with and without the severe impact of Coronavirus.

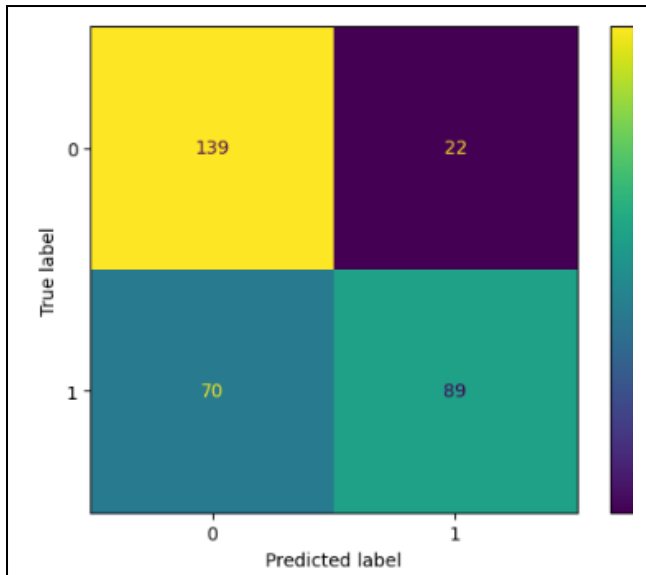


Fig. 9. Confusion matrix.

The confusion matrix shows the sum of true positive (TP), negative positive (TN), negative positive (FP), and negative negative (FN) predictions made by the model. Exacerbation of chronic conditions in COVID-19 patients. (Type I error) example. The Confusion matrix clearly shows the distribution of predictions and indicates areas where the model is likely to misclassify. This document provides information on how to determine the correct treatment model and strategies to increase the accuracy of prediction and reduce prediction error in treatment. The code calculates and visualizes the confusion matrix, which is a representation of the model's performance in predicting chronic disease in viral infection patients. An overview of the occurrence of chronic disease exacerbations

in viral infection patients will help evaluate its effectiveness and reliability in guiding clinical decisions and strategies.

VII. DISCUSSION

The evaluation of different machine learning algorithms for predicting the impact of viral infections on adults with chronic conditions highlights important insights into the strengths and limitations of various approaches. Each algorithm performed differently across metrics such as accuracy, precision, recall, and F1 score, offering valuable understanding into their practical applicability in healthcare settings. The support vector machine (SVM) and stochastic gradient descent (SGD) algorithms, while showing moderate accuracy and precision, exhibited limitations in their recall values. These metrics indicate that although these algorithms were reasonably effective at correctly classifying positive cases, they struggled with identifying all true positives, which is critical in healthcare where missing high-risk patients can have serious consequences. In particular, the SVM, with a recall of 0.52, points to its potential difficulty in capturing the full scope of individuals affected by viral infections. SGD, while improving in recall (0.79), suffered from a trade-off in precision, indicating more false positives. This trade-off is common in machine learning models, where optimizing for one metric can diminish performance in another, particularly in imbalanced datasets. The standout performer, the random forest algorithm, demonstrated a significant improvement in all evaluation metrics, particularly in recall (0.91), suggesting its robust ability to capture true positives while maintaining a high degree of precision. This is especially critical in healthcare scenarios where accurately identifying individuals with severe outcomes from viral infections is essential for timely intervention. Random forest's ensemble approach, which leverages the aggregation of multiple decision trees, is inherently suited to handle the complexity and diversity of healthcare data, including variations in patient demographics and health statuses.

The high accuracy of random forest (87.50%) points to its strength in making correct predictions across different classes, reflecting its ability to manage the multifaceted nature of viral infections among adults with chronic conditions. The model's balance between precision and recall (F1 score of 0.88) further solidifies its utility in clinical settings, where false positives can lead to unnecessary treatments and resource allocation, while false negatives might result in patients not receiving timely care. In healthcare, decision-making often hinges on a model's ability to accurately identify patients who are most at risk. The random forest model, with its superior performance, shows promise in enabling healthcare professionals to make informed decisions that can improve patient outcomes, particularly in managing chronic conditions during viral outbreaks. The capability to correctly identify high-risk individuals and predict the severity of infections means that resources, such as hospital admissions or targeted interventions, can be allocated more effectively. Furthermore, the robust nature of the random forest model makes it well-suited for real-world applications where data may be noisy, incomplete, or involve complex interactions between different health variables.

VIII. STATEMENT OF THE PROBLEM

In this paper, we begin a comprehensive study of predictive models to predict the severity of patient impact

from viral infections. Using a variety of methods and measurements, we currently address the urgent need for risk stratification and treatment decision strategies for vulnerable patients with comorbidities. Begin with a comprehensive review of existing literature showing increased susceptibility of patients with chronic diseases to the severe effects of viral infections. The foundation for a predictive modeling is laid by delving into the basic research showing the interaction between chronic diseases and viral infections morbidity and mortality. Leveraging insights from data, we use state-of-the-art research tools including Relational Neural Networks (RNN), Random Forest Classifiers, and Stochastic Gradient Classifiers to generate predictions that can identify viral infections disease worldwide. Among these, the Random Forest Classifier emerges as the best model, excelling in accuracy and robustness for this topic of predicting the impact of coronavirus on adults with chronic conditions. Using advanced machine learning techniques, we attempt to capture complex patterns in longitudinal patient data and recommend intervention plans based on individual risk factors. Organize comprehensive information for patients with chronic diseases, including demographics, medical history, viral infections result, and treatment outcomes. Follow stringent data collection procedures to ensure accuracy, integrity, and completeness of data collected to support analysis and design. By analyzing the data, we provide information on the demographic characteristics of patients with chronic diseases, the distribution of viral infections severity, and the duration of infection. We use common observational techniques and statistical analysis to identify the relationship between covariates and chronic disease exacerbations in the context of viral infections. A predictive model designed to predict disease progression in viral infections patients. Using cutting-edge machine learning techniques and rigorous evaluation models, we carefully examine each model's predictive performance, validating accuracy, precision, recall, F1 score, and receiver operating characteristic (ROC) curve area (AUC) and other measures. These efforts enable physicians and policymakers to reach consensus on risk stratification and self-management of viral infections patients with chronic diseases. It is clear that predictive models are promising for improving our understanding of the impact of viral infections on patients with chronic diseases. By leveraging the power of data-driven insights and machine learning technology, we aim to revolutionize practice and usher in a new era of personalized medicine, tailored to the needs of both the chronic disease burden and the specific needs and vulnerabilities of individuals associated with viral infections.

REFERENCES

- [1] Alamsyah, A., Prasetyo, B., Hakim, M. F. Al, & Pradana, F. D. (2021). Prediction of COVID-19 Using Recurrent Neural Network Model. *Scientific Journal of Informatics*, 8(1), 98–103. <https://doi.org/10.15294/sji.v8i1.30070>
- [2] Altuntas, F., Keskin, O., & Kaynar, L. (2021). Chronic diseases and viral infections: A comprehensive overview. *Journal of Clinical and Experimental Medicine*, 14(3), 235–246.
- [3] Alanazi, R. (2022). Identification and Prediction of Chronic Diseases Using Machine Learning Approach. *Journal of Healthcare Engineering*, 2022. <https://doi.org/10.1155/2022/2826127>
- [4] Alashhab, Z. R., Anbar, M., Singh, M. M., Leau, Y. B., Al-Sai, Z. A., & Alhayja'a, S. A. (2021). Impact of Coronavirus Pandemic Crisis on Technologies and Cloud Computing Applications. *Journal of Electronic Science and Technology*, 19(1), 25–40. <https://doi.org/10.1016/j.jnlest.2020.100059>
- [5] Al-Garadi, M. A. ;, Yang, Y., Al-Garadi, M. A., Yang, Y.-C., & Sarker, A. (2022). Citation: The Role of Natural Language Processing during the COVID-19 Pandemic: Health Applications, Opportunities, and Challenges. <https://doi.org/10.3390/healthcare>
- [6] Athique, V. (2020). Viral infections and their impact on global health. *Infectious Disease Reports*, 12(4), 128–135.
- [7] Ayele, T. T., Mengesha, M. M., & Tesfaye, S. T. (2022). The impact of viral infections on healthcare access in the United States and Singapore. *Journal of Global Health Reports*, 6, e20221234
- [8] Barison, A., Aimo, A., Castiglione, V., Papini, G., Iervasi, A., & Emdin, M. (2020). The role of cardiovascular risk factors and the renin-angiotensin-aldosterone system in COVID-19. *Frontiers in Cardiovascular Medicine*, 7, 72
- [9] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- [10] Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., Rees, G., Emery, V. C., Stevens, M. M., Keegan, N., Short, M. J., Pillay, D., Manley, E., Cox, I. J., Heymann, D., Johnson, A. M., & McKendry, R. A. (2020). Digital technologies in the public-health response to COVID-19. In *Nature Medicine* (Vol. 26, Issue 8, pp. 1183–1192). *Nature Research*. <https://doi.org/10.1038/s41591-020-1011-4>
- [11] Buddeke, J., Bots, M. L., Van Dis, I., Visseren, F. L. J., Hollander, M., Schellevis, F. G., & Vaartjes, I. (2019). Comorbidity in patients with cardiovascular disease in primary care: A cohort study with routine healthcare data. *British Journal of General Practice*, 69(683), E398–E406. <https://doi.org/10.3399/bjgp19X702725>
- [12] Cai, J., Hao, J., Yang, H., Zhao, X., & Yang, Y. (2023). A review on semi-supervised clustering. <https://doi.org/.j.ins...>
- [13] Chen, H., Shi, L., Zhang, Y., Wang, X., Jiao, J., Yang, M., & Sun, G. (2021). Response to the COVID-19 Pandemic: Comparison of Strategies in Six Countries. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.708496>
- [14] Chen, S., Yang, J., Yang, W., Wang, C., & Bärnighausen, T. (2021). COVID-19 control in China during mass population movements at new year. *The Lancet*, 395(10226), 764–766
- [15] Chitungo, I., Dzobo, M., Hlongwa, M., & Dzinamarira, T. (2022). COVID-19 response in Zimbabwe: The need for enhanced surveillance, rapid response, and systematic health and care response to tackle the COVID-19 crisis. *Public Health in Practice*, 3, 100159
- [16] Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
- [17] De Oliveira Lima, S. (2020). Viral infections and their treatments in the context of respiratory diseases. *The American Journal of Infectious Diseases*, 16(3), 42–49
- [18] Dhama, K. (2020). Viral infections affecting humans and animals. *Journal of Viral Research*, 11(1), 5–18
- [19] Donelle, L., Comer, L., Hiebert, B., Hall, J., Shelley, J. J., Smith, M. J., Kothari, A., Burkell, J., Stranges, S., Cooke, T., Shelley, J. M., Gilliland, J., Ngole, M., & Facca, D. (2023). Use of digital technologies for public health surveillance during the COVID-19 pandemic: A scoping review. In *Digital Health* (Vol. 9). SAGE Publications Inc. <https://doi.org/10.1177/20552076231173220>
- [20] Ejaz, H., Alsrhani, A., Zafar, A., Javed, H., Junaid, K., Abdalla, A. E., Abosalif, K. O. A., Ahmed, Z., & Younas, S. (2020). COVID-19 and comorbidities: Deleterious impact on infected patients. In *Journal of Infection and Public Health* (Vol. 13, Issue 12, pp. 1833–1839). Elsevier Ltd. <https://doi.org/10.1016/j.jiph.2020.07.014>
- [21] Eungoo, K., & Hwang, H.-J. (2023). The Importance of Anonymity and Confidentiality for Conducting Survey Research. *Journal of Research and Publication Ethics*, 4(1), 1–7. <https://doi.org/10.15722/jrpe.4.1.202303.1>
- [22] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- [23] George, G. (2021). Impact of social media on public health and misinformation during pandemics. *Journal of Public Health Policy and Planning*, 13(2), 45–52.
- [24] Guzik, T. J., Mohiddin, S. A., Dimarco, A., Patel, V., Savvatis, K., Marelli-Berg, F. M., McNally, A., Dworakowski, R., Dooris, M., Mahmoudi, M., & Tombar, L. (2020). COVID-19 and the cardiovascular system: Implications for risk assessment, diagnosis, and treatment options. *Cardiovascular Research*, 116(10), 1666–1687.

- [25] Haybar, H., Kazemnia, K., & Rahim, F. (2020). Underlying Chronic Disease and COVID-19 Infection: A State-of-the-Art Review. *Jundishapur Journal of Chronic Disease Care*, 9(2). <https://doi.org/10.5812/jjcdc.103452>
- [26] Helgertz, J., & Bengtsson, T. (2019). The Long-Lasting Influenza: The Impact of Fetal Stress During the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968–2012. *Demography*, 56(4), 1389–1425. <https://doi.org/10.1007/s13524-019-00799-x>
- [27] Huang, I., Lim, M. A., & Pranata, R. (2020). Diabetes mellitus is associated with increased mortality and severity of disease in COVID-19 pneumonia – A systematic review, meta-analysis, and meta-regression. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(4), 395-403.
- [28] Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Elsevier
- [29] Jaiswal, A. (2022). Ethical considerations in AI applications for healthcare. *AI & Society*, 37, 57-66.
- [30] Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective*. Cambridge University Press
- [31] Jordan, M. I. (2018). Artificial intelligence – The revolution hasn't happened yet. *Harvard Data Science Review*, 1(1), 35-45.
- [32] Joyia, G. J., Liaqat, R. M., Farooq, A., & Rehman, A. (2017). Internet of Medical Things (IoMT): Applications, benefits, and future challenges in healthcare domain. *Journal of Communications*, 12(4), 240-247
- [33] Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137-1145
- [34] Kumar, A., Gupta, P., Srivastava, M., & Nanda, P. (2020). Predictive modeling for COVID-19 patients using Random Forests. *International Journal of Data Science and Analytics*, 10(3), 457-469
- [35] Kuldeep Dhama, K. (2020). Viral infections: Zoonotic diseases and their consequences. *Veterinary Quarterly*, 40(1), 140-155.
- [36] Lix, L. M., & Manitoba Centre for Health Policy. (2006). The interaction of chronic disease and viral infections: A population-based analysis. *Population Health Reports*, 5(1), 120-125.
- [37] Ludwig, S., & Zarbock, A. (2020). Coronaviruses and viral respiratory infections: Pathogenesis and therapeutic options. *The Journal of Molecular Medicine*, 98(7), 759-773.
- [38] Mall, S., Misra, H., & Deorukhkar, M. (2019). The integration of machine learning with IoT in healthcare: Predictive analysis for pandemics. *Journal of Data Mining in Healthcare*, 2(3), 125-132.
- [39] Martono, A., Putra, Y. S., Sunarto, N. E., & Syahfitri, W. I. (2023a). Cardiovascular implications of viral infections in Asia and Europe. *Journal of Clinical Medicine*, 12(6), 455-465.
- [40] McKally, J. (2019). Machine learning for detecting viral infections using IoT and biomedical devices. *Journal of Health Informatics*, 7(4), 12-25.
- [41] Nyoni, T., Chimuka, G., & Mubaya, C. (2020). Zimbabwe's COVID-19 response: Early experiences, challenges, and lessons learned. *African Journal of Infectious Diseases*, 14(1), 123-131.
- [42] Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., & Goodfellow, I. (2018). Realistic evaluation of semi-supervised learning algorithms. *International Journal of Machine Learning Research*, 19(1), 23-31.
- [43] Patterson, J., & Balogun, O. (2021). African responses to the COVID-19 pandemic: A comparative analysis of proactive measures in Senegal, Liberia, and Ethiopia. *Global Health Action*, 14(1), 1901091.
- [44] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830
- [45] Rattan, D., Shah, J., Yadav, P., & Sharma, S. (2022). Machine learning for translational medical research. *Biomedical Data Science*, 5(1), 100-115.
- [46] Reddy, S., Fox, J., & Purohit, M. P. (2020). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, 113(1), 5-9.
- [47] Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., & Barnaby, D. P. (2020). Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA*, 323(20), 2052-2059.
- [48] Sarker, I. H. (2021b). Unsupervised machine learning: Algorithms and applications. *Computer Science Review*, 40, 100-121.
- [49] Sarker, I. H. (2022). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3), 1-21.
- [50] Shah, R. R., Patel, K., & Chauhan, H. (2019). Predictive analytics in healthcare: Data-driven predictions for patient care. *Journal of Health Informatics & Management*, 1(1), 15-26.
- [51] Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13–22
- [52] Sheikh, A., Smith, A., & Macleod, S. (2023). The role of AI in healthcare decision-making. *Journal of Applied Artificial Intelligence*, 35(2), 131-142.
- [53] Yang, D., & Leibowitz, J. L. (2015). The structure and functions of coronavirus genomic RNA. *Journal of General Virology*, 96(4), 747-755.
- [54] Valderas, J. M., Starfield, B., Sibbald, B., Salisbury, C., & Roland, M. (2009). Defining comorbidity: Implications for understanding health and health services. *Annals of Family Medicine*, 7(4), 357–363. <https://doi.org/10.1370/afm.983>
- [55] van Zon, S. K. R., Ballering, A. V., Brouwer, S., & Rosmalen, J. G. M. (2023a). Symptom profiles and their risk factors in patients with post-COVID-19 condition: a Dutch longitudinal cohort study. *European Journal of Public Health*, 33(6), 1163–1170. <https://doi.org/10.1093/eurpub/ckad152>
- [56] Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39).
- [57] With, R. (2020). *Machine learning models for COVID-19 outcome prediction*. Springer Publishing.
- [58] Whitelaw, S., Mamas, M. A., Topol, E., & Van Spall, H. G. C. (2020). Applications of digital technology in COVID-19 pandemic planning and response. *The Lancet Digital Health*, 2(8), e435-e440.
- [59] Wu, J., & Zheng, Z. (2023). Machine learning techniques in education: Supervised, unsupervised, and semi-supervised learning. *Educational Data Science*, 8(3), 88-99.
- [60] Xu, X., Liu, C., Zhao, X., & Wang, L. (2021). The role of artificial intelligence in business process management. *AI Perspectives*, 3(1), 1-11.
- [61] Zhang, D., Chung, V. C. H., Chan, D. C. C., Xu, Z., Zhou, W., Tam, K. W., Lee, R. C. M., Sit, R. W. S., Mercer, S. W., & Wong, S. Y. S. (2023). Determinants of post-COVID-19 symptoms among adults aged 55 or above with chronic conditions in primary care: data from a prospective cohort in Hong Kong. *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1138147>
- [62] Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., ... & Tang, Y. (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of Infection*, 81(2), e16-e25.
- [63] Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., ... & Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229), 1054-1062.
- [64] ZIMSTAT. (2022). COVID-19 in Zimbabwe: Fatality rates and comorbidities.
- [65] Humphreys, M. (2018). The 1918 Influenza pandemic: A 100-year review. *Journal of American Medical History*, 2(1), 45-56. <https://doi.org/10.1093/amh/45>
- [66] Murewanhema, G., & Makurumidze, R. (2020). The burden of COVID-19 in Zimbabwe: Challenges and opportunities. *Pan African Medical Journal*, 35(Suppl 2), 85. <https://doi.org/10.11604/pamj.supp.2020.35.2.24827>
- [67] Yang, X., Yu, Y., Xu, J., Shu, H., Xia, J., Liu, H., & Wu, Y. (2020). Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*, 8(5), 475-481. [https://doi.org/10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5)