

structure. Among the decision tree models, the one with 500 samples per internal node and a maximum tree depth of 15 achieved the highest average AUROC accuracy score of 96.057%. Logistic regression, being quick to train and less prone to overfitting, presented an easily manipulable and interpretable model. The best average AUROC accuracy obtained with logistic regression was 95.999%.

A research study conducted by Hong et al. [11] in 2020 focused on predicting mortality rates in Malaysia using three hybrid Lee-Carter (LC) models: LC-ARIMA, LC-Artificial Neural Network (LC-ANN), and LC-Random Forest (LC-RF). The study compared these models based on error metrics such as mean absolute percentage error (MAPE), root mean square error (RMSE), and absolute forecast error (AFE). Historical mortality data from 2000 to 2016 for Malaysia were obtained from the World Health Organization (WHO). The LC-ANN model performed the best in predicting mortality rates for males, while the LC-ARIMA model was the most effective for females in Malaysia. These findings provided valuable insights for predicting mortality rates in Malaysia and could contribute to the development of more efficient risk prevention and management systems in the life insurance industry. The study also identified the LC-ARIMA model as the best forecasting model for countries with longer life expectancy and good healthcare systems, such as Sweden, Ireland, Japan, Hong Kong, Norway, Switzerland, and Czechia. On the other hand, the LC-ANN model showed better performance in countries with less efficient healthcare systems, limited accessibility, lower coverage, and higher prevalence of risky behaviors like alcohol and tobacco consumption. The study made significant contributions by developing and analyzing the LC-RF model, applying and comparing the three hybrid models across multiple countries, and examining factors influencing the trend of change in mortality rates, including life expectancy, healthcare system quality, environmental cleanliness, income levels, personal behaviors, and dietary preferences. Additionally, the study analyzed factors affecting mortality rates in Malaysia, highlighting a concerning increase in the 15-24 age group due to high suicide ideation among male youths. These findings can be valuable for government agencies and NGOs responsible for youth well-being and empowerment, enabling them to take necessary actions to reduce suicide rates and increase life expectancy among Malaysian youths.

Another study conducted by Adegbosin et al. [12] aimed to explore the effectiveness of machine learning (ML) techniques in predicting under-five mortality (U5M) in low-income and middle-income countries (LMICs) and identify significant predictors of U5M. The researchers analyzed data from the Demographic and Health Survey, specifically focusing on 34 LMICs, which included a total of 1,520,018 children from 956,995 unique households. The primary outcome measure was U5M, and the study also compared the efficacy of different deep learning algorithms: deep neural network (DNN), convolutional neural network (CNN), and hybrid CNN-DNN with logistic regression (LR) for predicting child survival. The results revealed several significant predictors of U5M, including the duration of breastfeeding, the number of antenatal visits, household wealth index, postnatal care, and maternal education level. These variables were found to have a substantial impact on under-five mortality rates in LMICs. Furthermore, the study demonstrated that deep learning techniques outperformed logistic regression in predicting child survival. A comparative analysis of different models was conducted to assess their performance before feature selection. The results revealed that the hybrid CNN-DNN model outperformed all other models in terms of all metrics, while logistic regression (LR) performed the worst. Interestingly, feature selection did not lead to an improvement in LR's performance. However, for all deep learning models, feature selection resulted in a performance gain. The CNN-DNN model showed the most significant performance gain with a sensitivity of 0.71 and specificity of 0.83. Both before and after feature selection, the CNN-DNN model outperformed other classifiers. In contrast, LR had a sensitivity of 0.47 and specificity of 0.53, DNN had a sensitivity of 0.69 and specificity of 0.83, CNN had a sensitivity of 0.68 and specificity of 0.83, and the hybrid CNN-DNN model had a sensitivity of 0.71 and specificity of 0.83. Overall, this study provides valuable insights into the factors influencing under-five mortality in low- and middle-income countries (LMICs). It highlights the superior predictive capabilities of deep learning models compared to traditional approaches like logistic regression, offering significant potential for improving child health outcomes in LMICs.

Finally, a study by Elfassihi and Boumezoued [13] aimed to forecast future mortality by comparing the output forecasts of traditional stochastic mortality models with those of machine learning techniques, specifically neural networks and random forests. The researchers obtained data from the Human Mortality Database (HMD), a collaboration between the Max Planck Institute for Demographic Research and the University of California at Berkeley. The dataset covered the years 1950 to 2016 and was split into training and test sets. The first split included data from 1950 to 1990 for training and data from 1991 to 1999 for testing. The second split used data from 1950 to 1999 for training and data from 2000 to 2016 for testing. Stochastic models, such as the Lee-Carter model and the Age-Period-Cohort (APC) model, were employed in this study. The mean squared error (MSE) was used as the error metric for assessing the estimated mortality force. Results from this study demonstrated that the machine learning approaches yielded significantly lower errors compared to the stochastic models for both split scenarios. In contrast to the two baseline mortality models, the random forest method demonstrates a significant improvement in error reduction. Therefore, despite its greater complexity in implementation and interpretation, these algorithms possess the ability to acquire knowledge from data and provide more precise predictions of future death rates for HMD countries compared to conventional mortality models.

METHODOLOGY

This study started with collected relevant data, then move on to data pre-processing, a theoretical discussion on the proposed model structures, and finally statistical analysis for performance evaluation. Figure 1 depicts the overall flow of the research process.

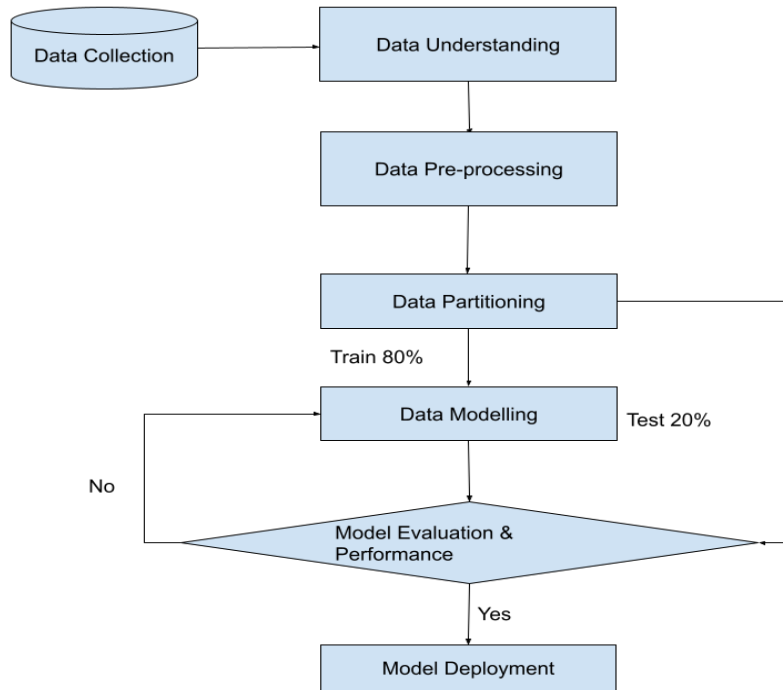


Figure 1. Research framework

Data source and collection for this study were retrieved from Population Division of the United Nations Secretariat's Department of Economic and Social Affairs. The data includes population estimates for 237 nations or territories from 1950 to the present, backed up by historical demographic patterns evaluations. Because the focus of this study is on Malaysian mortality trends, only Malaysian cases were extracted and used. The dataset has two columns and 74 rows. The two columns represent the years and the number of deaths per 1000 people in that year. The year column ranges from 1950 to 2022.

It is essential to do data pre-processing before feeding it into the model. In data pre-processing phase, several processes were performed including handling missing values, managing categorical variables, dataset splitting, and feature scaling. Next, for modelling, recurrent neural networks (RNN) and Long-Short Term Memory (LSTM) were used in this study. RNN is a form of artificial neural network that is used to analyse time series or sequential data. RNNs, unlike feedforward neural networks, feature connections that allow information to be transmitted from prior steps of the sequence to the present step, allowing them to exhibit temporal dynamics and capture dependencies across time. The capacity of an RNN to maintain an internal state or memory that captures information from earlier inputs is its defining feature. As the network processes each new input in the sequence, this memory is updated and passed forward. Each step's output is influenced not just by the current input but also by the network's prior states, allowing it to learn and recognise patterns in sequential data.

The recurrent unit is the fundamental building piece of an RNN, and it typically comprises of a hidden state and an activation function. The hidden state reflects the network's memory or context at a particular time step, and the activation function governs how the input and previous state are combined to form the output. Figure 2 shows an unrolled RNN.

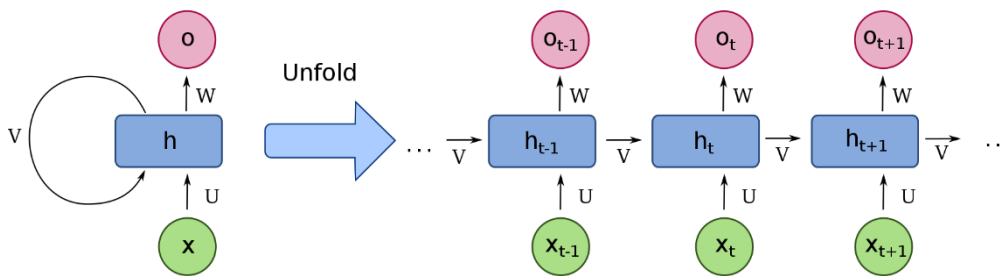


Figure 2. Unrolled Recurrent Neural Network [14]

In this scenario, a specific section of a neural network, referred to as A , examines an input x_t and generates an output value h_t . The inclusion of a loop facilitates the transfer of information from one network step to the following step. The

sequential structure of RNN indicates their strong connection to sequences and lists, making them the ideal neural network architecture for handling such data.

Classic RNNs often encounter difficulties in capturing long-term dependencies due to the vanishing or exploding gradient problem. To address this issue, several improved RNN architectures have been developed. One popular variation is the LSTM network, which incorporates gated memory cells and additional connections to control the information flow. The LSTM cell incorporates "gates" that combine weight multiplication and nonlinearity to carry out mathematical operations. These gates control the manipulation of information within the cell state, which acts as the primary pathway for information in an LSTM. An LSTM consists of three types of gates: the forget gate determines the extent to which the previous cell state should be ignored, the input gate determines the amount by which the new input should modify the cell state, and the output gate determines how much of the current cell state should be added to the current hidden state. A visual representation of an unrolled LSTM, showcasing the computation of the hidden state, can be seen in Figure 3.

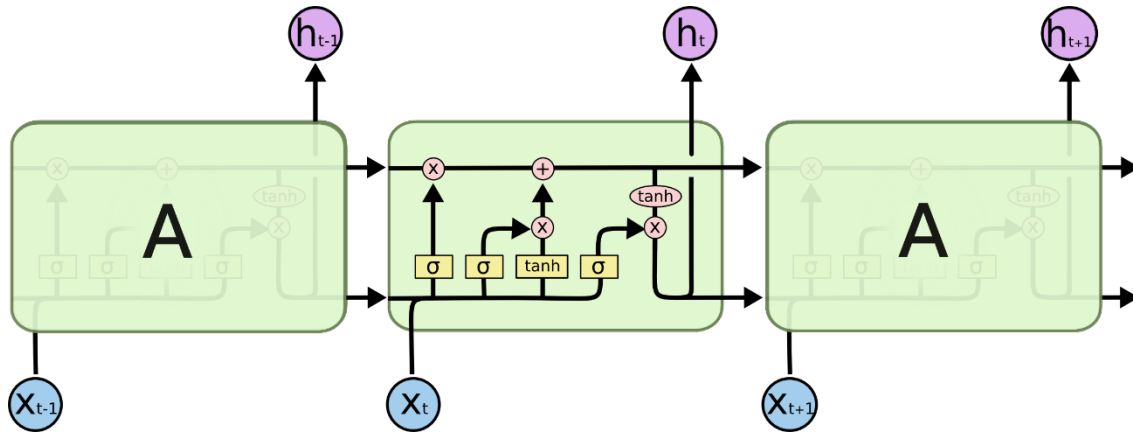


Figure 3. An unrolled LSTM depicting how the hidden state is calculated [13]

The provided illustration shows a diagram where each line carries a whole vector from one node's output to other nodes' inputs. The pink circles denote pointwise operations such as vector addition, while the yellow boxes represent trained neural network layers. The merging lines indicate concatenation, whereas the forking lines signify the duplication of content being sent to multiple destinations. The gates are formulated as follows:

$$\text{forget gate, } f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$\text{input gate, } i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\text{output gate, } o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

where the variables and function are x_t is the input vector at time step, t , h_t is the hidden state at time step, t , h_{t-1} is the hidden state at the previous time step, W_h, W_i, W_o are weight matrices for each gate, b_h, b_i, b_o are bias vectors for each gate, and σ is the activation functions. The current cell state is modified by multiplying the output of the forget gate with the previous cell state and adding it to the previous hidden state obtained from the input gate. At each time step, two calculations are performed to generate the output:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (5)$$

where C_t is the cell state, which is the raw candidate value, h_t is the preceding timestep's memory which is the final output after passing the candidate value via a tanh activation and then dotting an output gate, W_c is the weight matrices for the cell state while b_c is the bias vector for the cell state.

Both models were then compiled with the mean squared error loss function and the Adam optimizer. According to Kingma and Ba [15], the Adam (Adaptive Moment Estimation) optimizer is a deep learning technique that modifies the model's learnable parameters to increase the accuracy of neural networks. The optimizer has an adaptable learning rate and moment estimation capabilities. The popular stochastic gradient descent process, which is used to update the weights of a neural network, is extended by this method. Adam can improve performance and speed convergence by examining the past gradients to modify the learning rate for each parameter in real-time. It is employed because it is one of the most widely used optimisation algorithms.

Finally, the performance of the proposed model was evaluated by using several commonly used statistical assessment criteria such as R-squared, mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

RESULTS AND DISCUSSION

The findings of this research are discussed in this section. Various statistical studies were carried out to better understand the link between the dataset's years and death rates. Table 1 shows the mean, standard deviation, minimum, and maximum values for the mortality rate column which were computed using simple descriptive statistics.

Table 1. Descriptive statistics of the mortality rate column

Mean	6.3767
Standard deviation	3.1766
Minimum value	4.4900
Maximum value:	16.4070

In the mortality rate column, the number and percentage of missing values were calculated. There were no missing values found in the dataset. In this case, all 73 mortality rates from the year 1950 to 2022 were available for analysis. The mean value of 6.3767 suggests that, on average, the mortality rate is around 6.4. A higher standard deviation of 3.1766 indicates that the mortality rates vary more widely from the mean. In other words, the mortality rates are more spread out across the dataset. The lowest mortality rate achieved over the span of 74 years is 4.49 which was in the year 2003. The highest rate ever recorded was 16.407 in the year 1950. According to WHO, infectious illnesses were major causes of death in several nations, including Malaysia, throughout the 1950s. Illnesses such as pneumonia, tuberculosis, gastrointestinal infections, malaria, and other infectious illnesses were common causes of mortality during that period. Malnutrition and complications of labor were also common causes of death in some areas.

To understand the patterns or changes in the mortality rates over time, the trend of mortality rates over the years was visualized using a line plot. Figure 4 shows the mortality rate over the years.

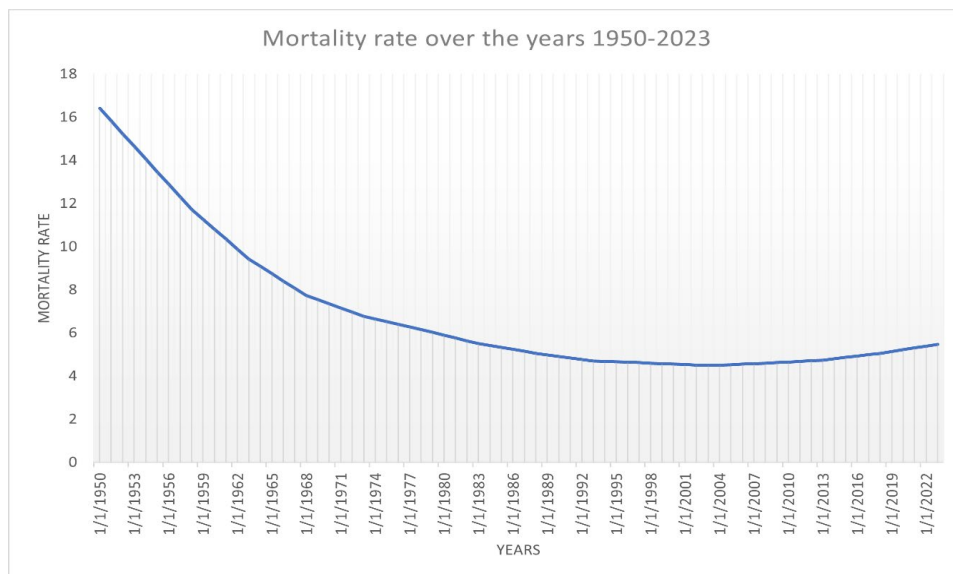


Figure 4. Mortality rate over the years

There is a general trend of decreasing mortality rates over time. The mortality rate has steadily declined from 16.407 in 1950 to 5.477 in 2023. In some periods, such as the late 1950s and early 1960s, the mortality rate remained relatively stable. However, this stability was within the context of an overall declining trend. There are some years where the mortality rate showed slight fluctuations such as in the late 1960s and early 1970s. However, the long-term trend still showed a decrease. From 1990 onwards, the annual decrease in mortality rates seemed to slow down, as the rate of decline became more gradual compared to earlier years.

Python was utilized in implementing a time series prediction model using a SimpleRNN (Simple Recurrent Neural Network) in Keras. For the RNN model, the training loss decreases well but the validation loss initially increases slightly and then decreases as seen in Figure 5, it suggests that the model might be overfitting the training data. Overfitting occurs when the model becomes too specialized in capturing the details and noise of the training set, to the point that it performs poorly on unseen data.

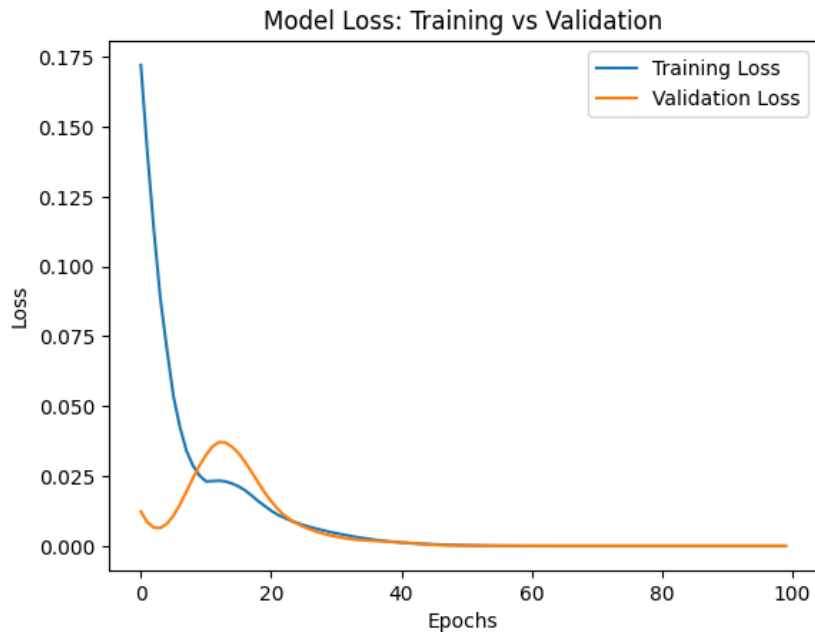


Figure 5. Training and validation loss of RNN model

The training and validation losses for the LSTM model both decrease over time, as shown in Figure 6, indicating that the model is improving its performance and learning to make better predictions. This is generally a good sign because it indicates that the model generalises effectively to both training and validation data.

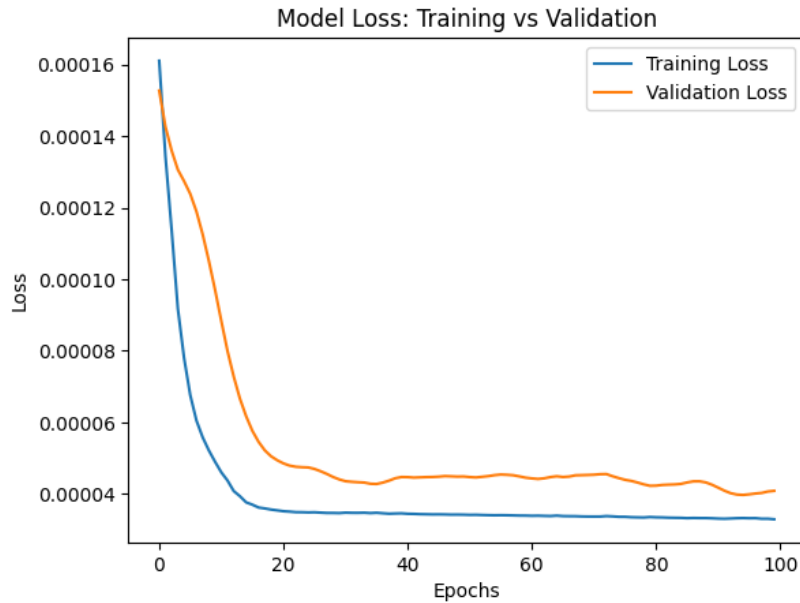


Figure 6. Training and validation loss of LSTM model

The evaluation results are reported in Table 2 below. The mean squared error, root mean squared error, mean absolute error and R-squared score of both RNN and LSTM models were used to evaluate their performance. These metrics provide information about the models' prediction skills.

Table 2. Evaluation results of the RNN and LSTM model

	Recurrent Neural Network (RNN)	Long-Short Term Memory (LSTM)
Mean Squared Error (MSE):	0.0253	0.0043
Root Mean Squared Error (RMSE):	0.1591	0.0653
Mean Absolute Error (MAE):	0.1563	0.0574
R-squared Score:	0.5518	0.9392

Based on the results, it can be seen that the LSTM model outperformed the RNN model across all criteria. When compared to the RNN model, the LSTM model had much lower mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). Furthermore, the R-squared (R2) score of the LSTM model was higher, indicating greater predictive potential. Figure 7 shows the graph of real mortality rates in comparison to the predicted mortality rates of both the LSTM and RNN model.

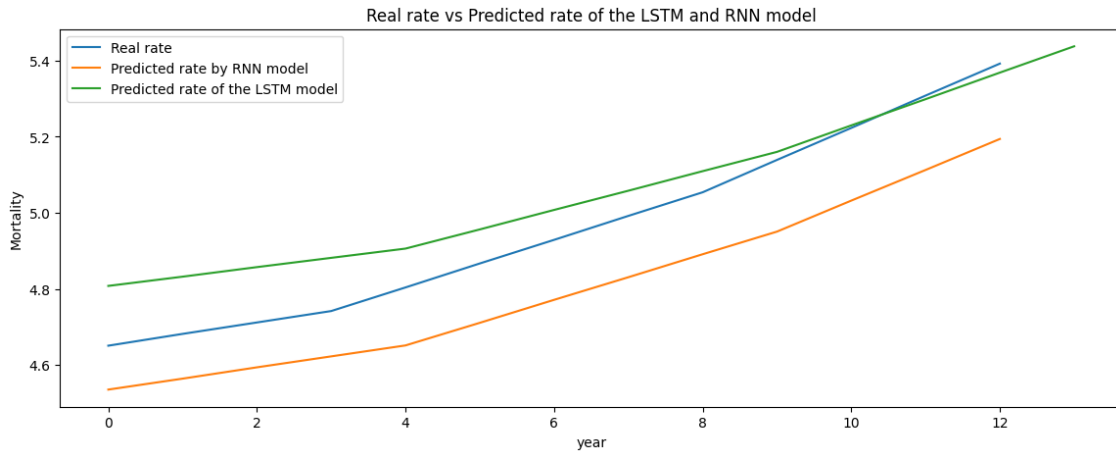


Figure 7. Real rate vs Predicted rate of the LSTM and RNN model

From the graph, both models exhibit similar trends in predicting the mortality rate. Both models' predicted values generally follow the shape and direction of the test values. There are, however, significant differences between the two models. The RNN model appears to accurately represent the overall trend of the mortality rate although it consistently underestimates the test values. This can be seen as the orange line (representing the RNN predictions) consistently lying below the blue line (representing the real rate). This observation suggests that the RNN model may struggle to capture the full range of variability in the test values, resulting in conservative predictions.

When compared to the RNN model, the projected values of the LSTM model are substantially closer to the real rate. The green line was further away from the general trend at the start of the graph, but it gradually approaches the test values around the year 2016 and then begins to underestimate the test values, though not as much as the RNN model. The LSTM model captures the patterns and variations in the data more accurately. The LSTM model has a strong ability to accurately approximate the mortality rate. Overall, the observations from the graph suggest that the LSTM model outperforms the RNN model in terms of capturing the patterns and variations present in the test data. Figures 8 and 9 show the LSTM model's predicted mortality rates with 2020 and 2021 rates included and excluded, respectively. Table 2 shows the predicted values for both models.

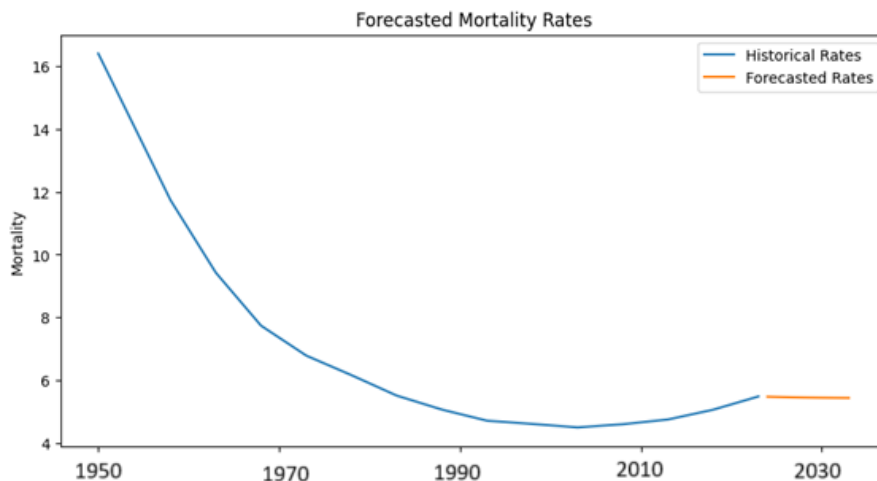


Figure 8. Forecasted mortality rates by using dataset with 2021 and 2022 rates

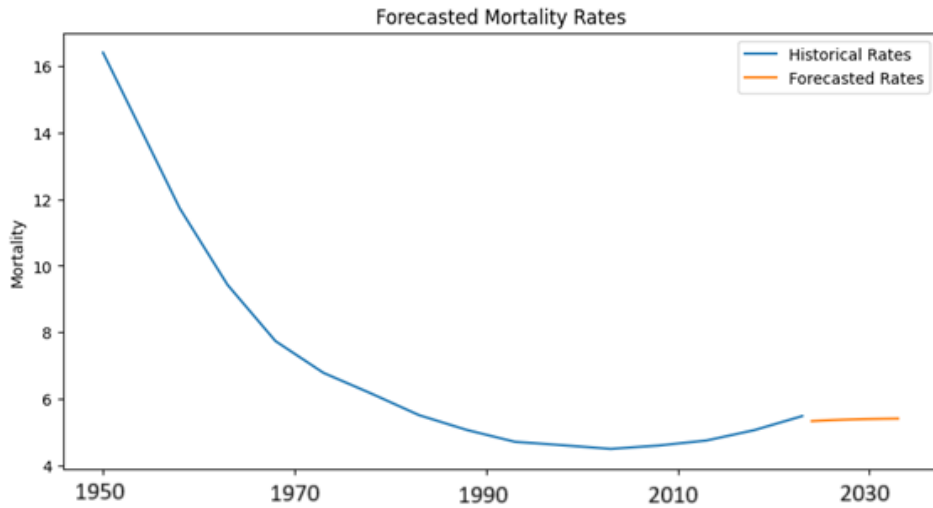


Figure 9. Forecasted mortality rates by using dataset without 2021 and 2022 rates

Table 2. Predicted mortality rates for 10 years using the LSTM model

	Dataset with 2021 and 2022 rates included	Dataset with 2021 and 2022 rates excluded
Year 1	5.4677	5.3263
Year 2	5.4600	5.3415
Year 3	5.4535	5.3543
Year 4	5.4481	5.3650
Year 5	5.4435	5.3739
Year 6	5.4397	6.3814
Year 7	5.4364	5.3876
Year 8	5.4338	5.3928
Year 9	5.4315	5.3972
Year 10	5.4296	5.4009

Figure 10 shows the result of the two-sample Z-test comparing the means of two groups which is the predicted mortality rates that used the dataset containing 2021 and 2022 rates hence called “Include” and the predicted mortality rates that used the dataset excluding the 2021 and 2022 mortality rates called “Exclude”.

```

Two-sample z-Test
data: Include and Exclude
z = 0.065445, p-value = 0.9478
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-2.567691 2.745091
sample estimates:
mean of x mean of y
5.27113 5.18243
    
```

Figure 10. Two sample Z-test between “Include” and “Exclude”

Since the p-value is 0.9478 which is greater than the level of significance, $\alpha = 0.05$, we fail to reject the null hypothesis. Hence, it can be concluded that the 2021 and 2022 mortality rates do not bring any significant difference to the predicted mortality rates.

CONCLUSION

This study aimed to predict mortality rates in Malaysia using Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) models and to investigate if post-pandemic mortality rates could affect the models when predicting mortality rates. The research utilized the 2022 Revision of World Population Prospects, extracting a dataset that includes years and the number of deaths per 1000 people in Malaysia data from the United Nations World Population Prospects and employed Python programming with TensorFlow and Keras libraries which were used for model development. The significance of this study lies in its potential to determine if the models could effectively forecast death rates, after which the models may be explored and reviewed further, providing valuable insights to stakeholders. By addressing the gaps identified in previous studies, such as the limited focus on the post-pandemic era and the lack of implementation of the LSTM technique, this research contributes to the analysis of Malaysian mortality rates.

The methodology outlines the research framework, including data collection, preprocessing, model development using RNN and LSTM, and performance evaluation metrics to assess the model's accuracy in predicting mortality rates. Performance evaluation metrics include the R-squared, mean squared error, root mean squared error and mean absolute error. A z-test is also conducted to analyze the significance of differences between mortality rates in 2021 and 2022. The study confirms the LSTM model's superiority over the SimpleRNN in accurately predicting mortality rates. The inclusion of post-pandemic data did not significantly affect the predictions. However, accurately predicting future mortality rates, especially in pandemic scenarios, remains challenging and requires further research. Overall, this study contributes to the understanding of Malaysian mortality rates and provides a foundation for future investigation in this field.

Predicting future mortality rates, especially considering the impact of events like the COVID-19 pandemic, remains challenging. Therefore, further research incorporating additional variables such as living conditions, smoking and drinking habits, and exposure to occupational toxins is necessary to enhance the predictive capabilities of mortality prediction models. Advanced techniques such as the convolutional neural network and Auto Regressive Integrated Moving Average (ARIMA) model could be implemented to forecast future mortality rates using post-pandemic data.

REFERENCES

- [1] World Population Review, World Population by Country 2023 (Live) [Online]. Available: <https://worldpopulationreview.com/>
- [2] S.S. Lim, T. Vos, A.D. Flaxman, G. Danaei, K. Shibuya, et al., "A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010," *Lancet*, vol. 380, no. 9859, pp. 2224–2260, 2012.
- [3] World Health Organization (2022). WHO COVID-19 Dashboard [Online]. Available: <https://covid19.who.int/>
- [4] The Star. (2022, October 31). Malaysia Experiencing Small COVID-19 Wave due to XBB Subvariant, says Health Ministry [Online]. Available: <https://www.thestar.com.my/news/nation/2022/10/31/malaysia-facing-small-covid-19-wave-due-to-xbb-variant-says-health-ministry>
- [5] G.V. Houdt, C. Mosquera and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, pp. 5929-5955, 2020.
- [6] T. Mikolov, M. Karafiát, L. Burget, J. Černocký and S. Khudanpur, "Recurrent neural network based language model," in Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), Japan, 2010, pp. 1045-1048.
- [7] A. Graves. (2014, April 2). Generating Sequences with Recurrent Neural Networks [Online]. Available: <https://www.arxiv-vanity.com/papers/1308.0850/>
- [8] Z.C. Lipton, D.C. Kale, C. Elkan and R. Wetzal, Learning to Diagnose with LSTM Recurrent Neural Networks [Online]. Available: <http://dblp.uni-trier.de/db/journals/corr/corr1511.html#LiptonKEW15>
- [9] N. Agarwal and R. Dutta, "Comparative predictive analysis of mortality rate after COVID-19 vaccination using various machine learning approaches," in Proceeding of 2022 International Conference on Computer Communication and Informatics (ICCCI), India, 2022.
- [10] Z. Xiao, "COVID 19 mortality rate prediction based on machine learning methods," in Proceeding of 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), China, 2021.
- [11] W.H. Hong, J.H. Yap, G. Selvachandran, P.H. Thong and L.H. Son, "Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest," *Complex & Intelligent Systems*, vol. 7, pp. 163-189, 2020.
- [12] A.E. Adegbosin, B. Stantic and J. Sun, "Efficacy of deep learning methods for predicting under-five mortality in 34 low-income and middle-income countries," *BMJ Open*, vol. 10, no. 8, pp. 1-8, 2020.
- [13] A. Elfassihi and A. Boumezoued, "Mortality trend prediction using machine learning," *Milliman White Paper*, 2022.
- [14] C. Olah. (2015, August 27). Understanding LSTM networks Github.io [Online]. Available: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- [15] D. Kingma and J. Ba, "Adam: a method for stochastic optimization," in Proceeding of the 3rd International Conference for Learning Representations, United States, 2015. pp. 1-15.