**Executive Summary**

This report introduces a proposal to predict children mortality rate through techniques such as Genetic Algorithm for searching the most aligned range of years in the dataset, K-means Clustering for grouping countries with similar breastfeeding and mortality characteristics and Long Short-Term Memory (LSTM) for predicting mortality rate for 4 consecutive years. This study uses python as coding with libraries, pandas, numpy, tensorflow, sklearn, hyperimputation and matplotlib. Pandas is for reading csv. Numpy is formatting and array. Tensorflow is for high-level coding of building LSTM. Sklearn is for K-means Clustering. Matplotlib is for plotting figure. Excel is used for tabulation and dataset reading.

Preprocessing and data manipulation such as merging tables, formatting, imputation and normalisation is presented. The result of GA is 1985-2021 which is believed to a be the nearly optimum depending on the interpretation of the fitness function. The algorithm converges with 20 generations. K-means clustering effectively segments regions with similar patterns in breastfeeding and mortality data that is proved by the improvement of accuracy of LSTM with clustering. The LSTM model performs well in countries with high children mortality rate. 58 out of 199 countries (mostly developing counties) achieve over 96% of predictions within 95% confidence intervals. Evaluation criteria such as MSE, MAE and RMSE are used to measure performance of the LSTM. Predictions followed by historical values is presented in the Appendix. The results in this report are presented in tables, graphs and plots. Considering the early convergence in GA, exploring the potential of more efficient algorithm such as Simulated Annealing to replace GA could save computational time. Experimenting different numbers and features for clustering or adding randomness could increase the effectiveness of cross-region learning. Investigation of fine tuning hyperparameters and architecture is needed to optimise model’s performance. Considering adding benchmark group could allow developing countries learn the existing pattern of developed countries for forecasting can potentially increase model’s accurate.[10]

**Introduction**

Despite of advancements in medical technology and raised public awareness, global child malnutrition remains closely linked to high rates of child mortality, particularly among children under the age of five. Research shows that more than half of global under-five deaths are preventable, while nutrition is playing a crucial role [1].

The understanding of between features such as infant breastfeeding rate and mortality rates(0-5 years) is essential for predicting trends in the future. By leveraging AI models and search algorithms, the NGO can analyse nonlinear relationships between features in datasets. In terms of breastfeeding rates and children mortality statistics, AI models can help to forecast high-risk regions, and prioritise interventions. Therefore, organisations can target resources, justify the case for funding, and reasonably contribute to achieve global health goals.

AI has revolutionised various sectors, including public health, through its ability to analyse large datasets and make accurate predictions. The capability of deep learning to compute high dimensional tasks is superior to conventional machine learning approaches database grows[2]. Neural Network “learns” from a training set by backpropagating the errors through its layers. This process updates the weights in vector form according to the gradients of the loss functions for minimising the error between predictions and the target labels. The foundation of Recurrent Neural Network is Backpropagation Through Time(BPTT) is introduced in 1974 [3]. Apart from backpropagating through layers, BPTT calculates error at each time step. The hidden state at each time step contains self-loops which hold memory from previous time step[4]. This property introduces memory in a neural network for sequence prediction, language modelling in time series.

Long Short-Term Memory (LSTM), a form of RNN which excels at learning patterns and dependencies in sequential and time-dependent data. LSTM is first developed by Hochreiter et al and Schmidhuber et al to address the vanishing gradient problem in RNN[5]. The ability of capturing long-term dependencies has raised the interest of later development of transformers.[6]

In public health, machine learning techniques, are widely used to make diagnostic predictions comparable to those of human experts [7]. With increasing nutrient and enhancements of medical development, trend of decreasing in mortality rates(0-5 years) globally can be seen in fig. 1. Undoubtedly, apart from region and number of deaths among children, time is a significant factor for predicting mortality. By applying LSTM, accurate forecasting of child mortality rate can be made by relating various features. Insights can be gained to guide targeted interventions to reduce malnutrition-related deaths.

In this report, a LSTM model is proposed to predict children mortality rate. Genetic Algorithm(GA) is used to determine the most fit year range across all counties to ensure an sufficient information from all countries can be considered. Imputation is used to fill in missing data in features from1985 to 2021.

K-mean Clustering is applied to group countries with similar breastfeeding and mortality rates. This classification is added to the dataset to allow LSTM learn pattern between counties in the same group. For comparison, a separate set of test are conducted with no clustering. The result shows improvement in accuracy by adding regional grouping for training.

A graph with blue lines

Description automatically generated

**Fig. 1** Decreasing trend of children mortality

**Literature Review**

Since 2013, some researchers use deep learning models to predict public health data with Artificial Neural Network (ANN) and (RNN). Forecasting the mortality rates for different age groups and genders in Malaysia, Michael et al. analyses performance of a hybrid model with Lee-Carter (LC)-ANN, against a LC-Autoregressive Integrated Moving Average (ARIMA) method and a LC-Random Forest(RF) model[9]. Time series of mortality rates are consistent without any temporal misalignment. The training and testing splits are 2000-2013 and 2014-2016, respectively. Models are evaluated using accuracy metrics, MAPE, RMSE, and Average Forecast Error (AFE). The LC-ANN model performs well by its adaptability to complex, non-linear trends. Unlike RNN, traditional ANN is also able to predict time-outputs effectively. However, there is difficulty in selecting the optimal hyperparameters, the number of neurons and hidden layers, which are keys of model’s accuracy. It is also complicated to interpret the logic and model structure behind its configuration.

The complexity of forecasting children mortality includes limited historical data, nonlinearity of region-specific trends and breast-feeding trends. Relating trends across is difficult that different countries have their own culture, socioeconomic status and their unique health issues. Without sufficient information of these indicators, classification remains a challenging task. In the case of developing countries, there can be sharp declines in children mortality rates because of improvements in nutrition and healthcare. However, these improvements are short-term, simply projecting them through regression may lead to unrealistic forecasts. To address this issue, researchers has used developed countries as benchmarks and assumes developing countries might converge with those in developed countries[10]. LSTM can learn time-dependent patterns with the ability to capturing long-term dependencies as well as short-term patterns. However, it is still difficult to presume countries have aligned healthcare access, disease prevalence, and socio-political conditions. Therefore, an effective grouping strategy is crucial for classifying features with nonlinear relationships. Researchers uses K-means clustering to explore demographic or regional differences. For example, Daniel et al. uses this technique to segment patients based on demographic information and health-related indicators such age and health condition index[11]. The elbow method and silhouette score which can find the optimal range of number of cluster is presented in this study. After K-means-clustering patients with similar characteristics, the LSTM model is applied separately on each cluster to predict medical costs. This study demonstrates that clustering can improve prediction accuracy by reducing data complexity and capturing group-specific trends.

It has been shown that with an optimal number of cluster, multidimensional data can be segmented into meaningful groups, each representing distinct characteristics. In this study, K-means clustering can be applied to group regions with similar breastfeeding-children mortality rate relationship. This approach allows the model learning patterns across counties with comparable breastfeeding and children mortality dynamics.

One of the tools that Jose et al. [10] use in the study is Keras with TensorFlow as the primary deep learning framework to implement the LSTM. This LSTM is with two stacked layers, each with 128 neurons, two fully connected layers and an output layer which uses linear activation function. Keras is a high-level API that allows building complex models in a simpler way. Tensorflow has stable support for RNN, especially LSTM layers.

Training LSTM model requires consistent and non-missing data for data preprocessing and time series splitting. Typical preprocessing includes standardising format, encoding categorical values, normalising data and imputation. To maintain data quality, Daniel et al. uses default value of “1900-01-01” for dates and “0” for other missing fields and uses threshold to filter out small values[11]. The lack of imputation for missing data could introduce bias and impact on data integrity, thereby less accurate predictions. Imputation such as K-nearest neighbours or regression are common techniques to preserve data integrity more effectively. HyperImpute has been developed to learn from the datatype and automatically apply the optimal imputation method based on the type[12]. Leveraging deep learning models, it can be a sensible tool to impute missing values in breastfeeding data and number of death because of its flexibility.

Daniel et al. used MinMaxScaler function to scale numerical values for inputs of the LSTM model[11]. In his study, statistical metrics such as RMSE and MAPE are used to measure against no clustering. Runing sperate model for each cluster in this study presumes that patients from different cluster are not having similar pattern. In this case, the model would lack the ability to learn pattern across clusters.

**Research Design**

The assumptions of the dataset are as follows:

* In dataset, Infant\_nutrition\_data\_by\_country, data across years is treated as data from the starting year.
* Both sex values of mortality rate are the average of those of Male and Female
* Both sex values of number of deaths are the sum of those of Male and Female

The steps of data pre-processing before running imputation are as follows:

1. Values of Both-sexes in mortality rate and number of deaths are removed
2. In Infant nutrition data by country , any year interval in the year field, the starting year will be taken as the value. (2015-2016 ->2015)
3. Merge Infant\_nutrition\_data\_by\_country and all Children\_mortality\_rates tables by add two columns, namely Early initiation of breastfeeding (%) and Infants exclusively breastfed for the first six months of life (%) to the main dataset.
4. Remove the range inside square brackets after the numeric values.
5. For every column with numeric values, remove any space or symbol within the value.

After unifying the datasets, there are still missing values in columns of number of death (male and female) and breastfeeding columns. HyperImputation is applied to fill in all missing values in breastfeeding[12]. The dataset is searched by GA for finding the maximum year periods that have minimum number of missing values across all countries. Hyperimputation is used again for filling in missing values from Children\_mortality\_rates and Number\_of\_death. K-mean clustering is then used for grouping regions with similar breastfeeding and mortality characteristics.

LSTM is selected because of its recursive neural network architecture which allows memory to be selectively stored or forgotten. In fig. 2, the long-term memory (line c\_t-1- c\_t) goes through the output from the short-term memory (line h\_t-1-h\_t) by a sigmoid function that the value from the input to value between 0 and 1. Here, the output determines the percentage of the long-term memory is remembered. This output will then join a sum of a product of a percentage of potential memory to remember (left) and a potential long-term memory in where an input gate determines how much potential long-term memory to retain. The updated long-term memory will be an input of a tanh function to become a potential short-term memory. Again, the last sigmoid function is being multiplied with this output to generate the new short-term memory. This mechanism can resolve the vanishing gradient problem which can be common seen in conventional RNN.

A diagram of a computer

Description automatically generated

**Fig. 2** Cell structure in LSTM

The model consists of an input layer, two LSTM layers, each of them followed with a dropout layer, and a fully connected output (dense) layer. All features and target are normalised by MinMaxScaler() at first and inverse transform is used to retrieve the output values during test. The input shape is [batch size, window size, total\_features\_count]. The features are summarized as follows:

|  |  |
| --- | --- |
| **Numeric features:** | Number of features |
| Early initiation of breastfeeding (%): | 1 |
| Infants exclusively breastfed for the first six months of life (%): | 1 |
| Number of deaths among children under-five (Male and Female) | 2 |
| Under-five mortality rate (Male and Female): | 2 |
|  |  |
| **One-Hot Encoded Features:** | Number of features |
| Countries | 199 |
| Clusters | 6 |

Batch size of 32 is selected because of its balance between computational time and gradient accuracy. The training, validation and test set are 1985-2012, 2013-2015 and 2016-2021, respectively. Both features and targets are structured in window of size 2. In testing set, input of 2016 and 2017 are used to predict output of 2018, input of 2017 and 2018 are used to predict output of 2019 and so on.

The testing set,2016-2021 have 4 predictions, 2018-2021, since 2016 and 2017 are the first two years used to predict output of 2018. The model is trained for each countries, then validatd for each countries and tested for each country. This order ensurea all countries in the training set year range are completed before validation and test.

The parameters of the LSTM are presented as follows:

|  |  |
| --- | --- |
| **parameters** | **size** |
| batch size | 32 |
| epochs | 100 |
| no. of neurons in LSTM layer | 100 |
| no. of neurons in Dense layer | 2 |
| total features count | 211 |
| window size | 2 |
| learning rate | 0.0001 |

Experimental Results and Analysis

The search space of GA is presented in table X. Two columns of breastfeeding rate are excluded in GA due to the number of missing values. This GA starts by random initialisation of population size, 50, in which each member is represented by a start year and an end year. Fitness function is defined as adding one point to each year when all values are presented in the four columns for all countries and giving penalty otherwise. Best two parents are picked by tournament selection followed by single-point crossover with 50%-crossover rate. Elitism is applied to ensure that the best 2% of the population retain in the next generation. The exploration is based on 30%-mutation rate to avoid premature convergence. The algorithm terminates after 100 generations are processed. Parameters are adjusted by trial and error incrementally. Penalty is strength a few times to effectively eliminate early years with more missing values. Most of the missing values of mortality rate and death counts occur before 1985, limiting the effectiveness of a single-point crossover for exploration. Due to this data distribution, a high mutation rate is used for exploration to compensate. In fig. 3, The GA converge within 20 generations could indicate that the search space is relatively simple. The end year can be intuitively known as the last year, 2021. The explorable variable is the start year. Thus, the optimal solution is obtained in the early generations.

The result of the optimal range of years is 1985-2021.



**Table 1** Search dimensions of GA

A graph with a line

Description automatically generated

**Fig 3.** GA convergence

A graph of a number of clusters

Description automatically generated

**Fig 4.** Elbow Method

A graph showing a number of dots

Description automatically generated with medium confidence

**Fig.5** Scatter plot of K-means clustering

According to the plot of elbow method in fig. 4, 6 is taken as the number of clusters. The algorithm uses StandardScaler() to scale the breastfeeding rate and the average mortality rate into a given range. The Euclidean distance between points is used to measure the similarity. The cluster number is then allocated to countries in similarity. The clustering plot is presented in fig.5. The cluster number is then added to the main dataset. The first results from the LSTM are for the comparison of whether using clustering in table . The parameters of the model remain constant as stated in Table X. To evaluate the accuracy of the model, the MSE, MAE and RMSE for Male and Female are computed for each year and reported as the criteria.

A table with numbers and lines

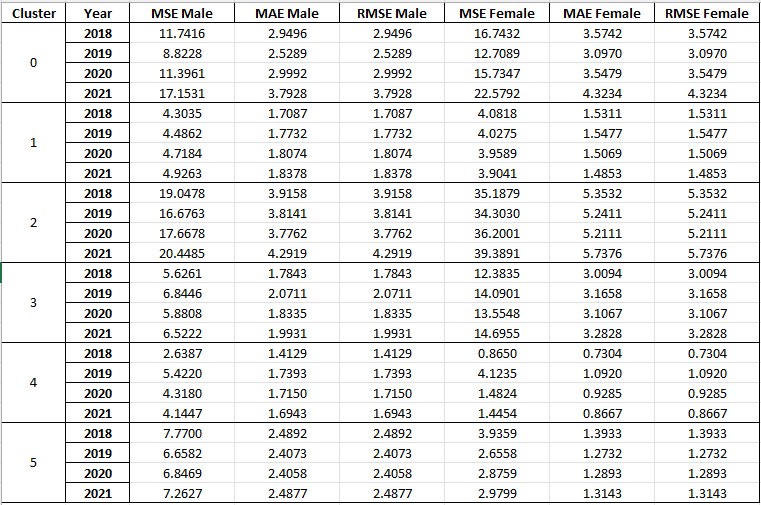
Description automatically generated

**Table 2** Performance of predictions of LSTMs with or with clustering

With clustering, a general improvement on accuracy can be observed in Table X. Model without clustering loses accuracy as time progresses. Notice that cumulative error stacks in non-clustering model across years. Whereas this is less significant in model with clustering that contributes to more consistent performance. Error metrics for females are higher than for males, both with and without clustering. This could be explained a general lower mortality rate for female.

Table 3 show observable variation in error metrics that each cluster meet much less errors than others. Cluster 1 and 4 show the best error consistently across years that illustrate the clustering can capture the relationship between low mortality and high breastfeeding rate. The model performance better in Cluster 4 with lower RMSE values as low as 0.7304 for female mortality in 2018. Cluster 1 also show stable errors across years. Cluster 5 demonstrates relatively stable errors without accumulating when time progresses. Cluster 0, 2 are the regions with higher mortality rate which show generally higher errors. These are developing countries where you they spread sparsely on the scatter plot.

Cluster 3 represents countries with high breastfeeding rate and high mortality that show better errors than Cluster 0 and 2 (developing countries) but less accurate than Custer 1 and 4 (developed countries). They exhibit variable breastfeeding and mortality characteristics that might lead to subtle differences in patterns across countries.



**Table 3** Performance of predictions of LSTM with clustering

Table 4 shows the statistics of the percentage of predictions that in with the 95% confidence intervals.

Interesting, the model performs the best in Cluster 0, 2 and 3. These regions are classified as with higher mortality rate based on plot X. This suggests the model achieves statistically consistent predictions as evidenced by the 95% confidence. The reliable performance in these clusters indicates that the model is effective to capture key patterns and factors contributing to high mortality trends in critical regions countries such as Nigeria. The improvement of model’s accuracy by taking factors which is influenceable by socioeconomic status can capture mortality trends more effectively. This finding resonates with similar research by Jose et al.[10]. This model would be restrictedly applied to high-mortality countries.

A table with numbers and text

Description automatically generated

**Table 4** Performance of predictions of LSTM with clusteringaginst confidence intervals

Conclusion

Many factors can affect children mortality trends, different factors such as healthcare access, socioeconomic status and nutrition are influential in a complicated way. Observable trends of high mortality in developing countries require extra attention to improve them.

Based on the dataset of children-under-five mortality rate from WHO, the proposed model manages to predict mortality rate in the coming four years. The results show statistically consistent predictions by a regular LSTM with K-mean clustering in pre-processing.

The first part of the solution proposed a global search method, GA, which can get global optimum to determine the nearly most aligned year range from a WHO dataset with missing values. The part of the solution is to demonstrate an unsupervised learning technique, K-means clustering that can segment countries in group with similar breastfeeding and morality characteristics. This regional grouping is then added to the dataset as an additional feature which allow the LSTM mode to learn patterns from countries within the same group. The model shows convincing result mostly in developing countries with high mortality rates. The evaluation follows statistical approach by comparing MSE, MAE, RMSE and confidence level given by the dataset. The model is trained, validated and tested by data from 199 countries.

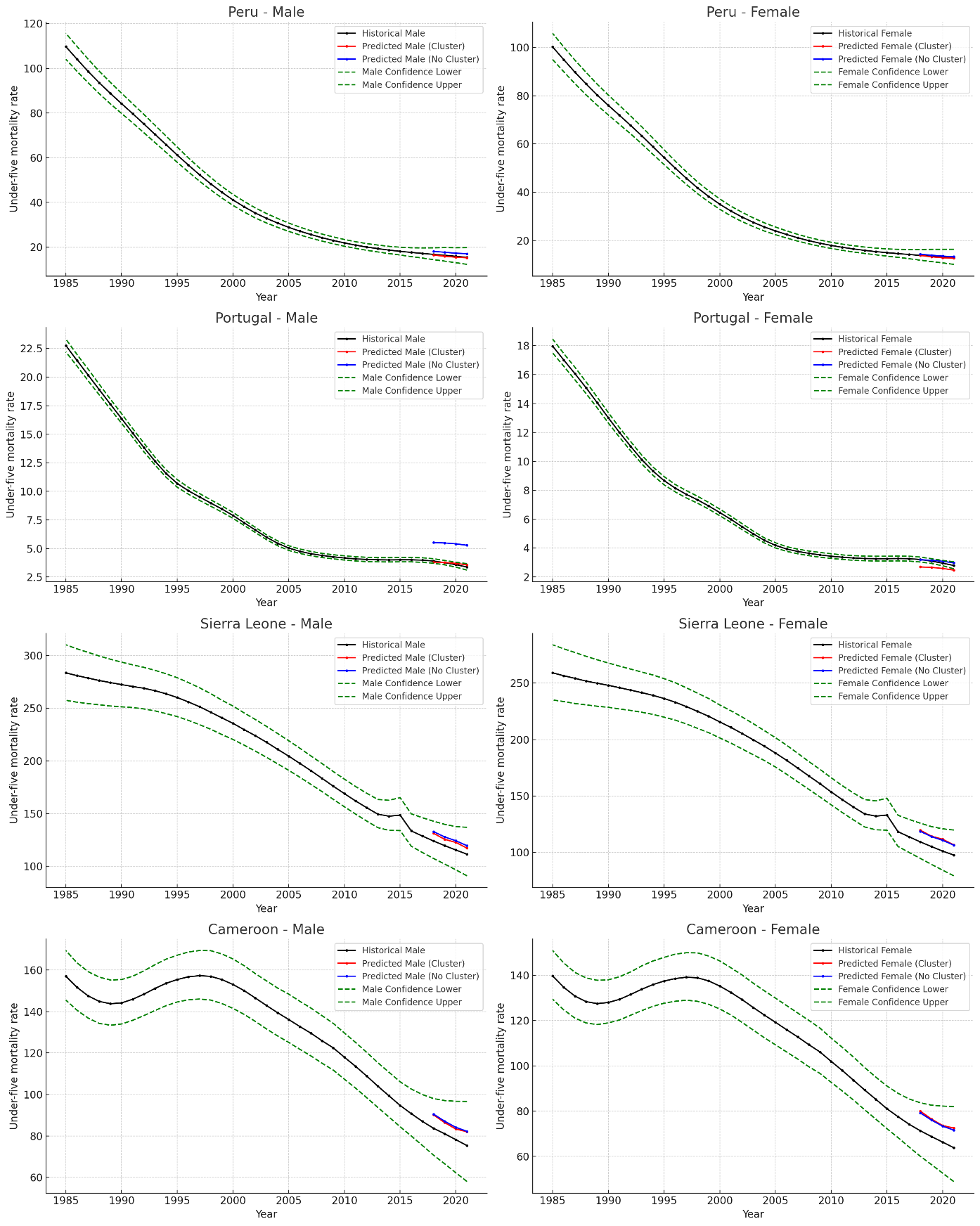
Limited time is put in fine-tuning hyperparameters such as window size, batch size and number of neuron of the model. Simulated Annealing has potential to get the best year alignment instead of using GA to reduce computational time and improve efficiency. Adding features of benchmark group for project patterns from developed countries to developing countries to the LSTM model can be studied in the future[10].

Appendix A

A table of information

Description automatically generated with medium confidence

**Table A1.** Countries in 6 clusters



**Fig. A1.** Historical, Confidence intervals and predictions for LSTM with clustering and without clustering (Male and Female)

A screenshot of a computer screen

Description automatically generated

**Table. A2.** Table form of Fig. A1

**Reference**

[1] L. Liu, S. Oza, D. Hogan, J. Perin, I. Rudan, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black, "Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: an updated systematic analysis," *The Lancet*, vol. 385, no. 9966, pp. 430-440, Jan. 2015.

[2] Krizhevsky, A., Sutskever, I., and Hinton, G. E., “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[3] P. J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D. dissertation, Harvard Univ., Cambridge, MA, USA, 1974.

[4] P. J. Werbos, "Backpropagation Through Time: What It Does and How to Do It," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.

[5] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.

[7] A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, pp. 1317-1318, Apr. 2018.

[8] M. A. Pienaar, J. B. Sempa, N. Luwes, and L. J. Solomon, “An artificial neural network model for pediatric mortality prediction in two tertiary pediatric intensive care units in South Africa: A development study,” Frontiers in Pediatrics, vol. 10, no. 797080, 2022. doi: 10.3389/fped.2022.797080.

[9] 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, no. 1, pp. 163–189, 2021, doi: 10.1007/s40747-020-00185-w.

[10] J. Garrido, Y. Shang, and R. Xu, "LSTM-Based Coherent Mortality Forecasting for Developing Countries," *Risks*, vol. 12, no. 2, pp. 1-24, Feb. 2024, doi: 10.3390/risks12020027.

[11] D. R. Sandoval Serrano, J. C. Rincón, J. Mejía-Restrepo, E. R. Núñez-Valdez, and V. García-Díaz, "Forecast of Medical Costs in Health Companies Using Models Based on Advanced Analytics," *Algorithms*, vol. 15, no. 4, pp. 106, Mar. 2022. doi: 10.3390/a15040106.

[12] D. Jarrett, B. Cebere, T. Liu, A. Curth, and M. van der Schaar, "HyperImpute: Generalized Iterative Imputation with Automatic Model Selection," *Proceedings of the 39th International Conference on Machine Learning (ICML)*, Baltimore, MD, USA, vol. 162, 2022, pp. 1-10.