Forecasting Unemployment Rate for Multiple Countries Using a New Method for Data Structuring

Amjad M. Monir Aljinbaz[1], Mohamad Mahmoud Al Rahhal[2]
Department of Computer Science, University of York, York, UK[1]
Department of Applied Computer Science-College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia[2]

***Abstract*—Forecasting the Unemployment Rate (UR) plays a key role in shaping economic policies and development strategies. While most research focuses on predicting UR for individual countries, there has been limited progress in creating a unified forecasting model that works across multiple countries. Traditional time series methods are usually designed for single-country data, making it difficult to develop a model that handles data from various regions. This study presents a new data structuring technique that divides time series into smaller segments, enabling the development of a single model applicable to 44 countries using various economic indicators. Four forecasting models were tested: an artificial neural network (ANN), a hybrid ANN with machine learning (ML), a genetic algorithm-optimized ANN (ANN-GA), and a linear regression model. The linear regression model, which used lagged UR values, delivered the best results with an R^2 of 0.964 and 89.8% accuracy. The ANN-GA model also performed strongly, achieving an R^2 of 0.945 and 85.1% accuracy. These results highlight the effectiveness of the proposed data structuring method, demonstrating that a single model can accurately forecast multiple time series across different regions.**

***Keywords*—Unemployment rate; artificial neural network; time series; hybrid model; genetic algorithm**

I. Introduction

Unemployment rate (UR) is a critical macroeconomic indicator that plays a pivotal role in economic planning, social stability, and policy development [1]. Researchers have used several methods to forecast unemployment rate using stochastic methods, machine learning and neural network. However, most of these approaches are focused on developing country-specific models. It is rare to find studies attempting to build a global model applicable across multiple countries. Because of the importance of building global model for multiple countries, an attempt was made to build such [2].

Employment is inherently linked to macroeconomic conditions and is influenced by a broad spectrum of factors, including inflation, gross domestic product (GDP), and other economic indicators[3]. At the same time, microeconomic factors such as age, educational attainment, gender, and poverty status can significantly impact unemployment dynamics [4], [5], [6]. Consequently, understanding the UR requires a multifaceted approach that accounts for both macro- and microeconomic influences.

The development of single-country models has been prevalent in previous research, utilizing a variety of techniques. Several studies used stochastic time series techniques. In such techniques, they study the effect of independent variables on dependant variables regression to examine level of significance [7], [8].

Also, time series analysis, including autoregressive integrated moving average (ARIMA) and its variants (e.g., ARMA, SARIMA, VARIMA), have been employed to capture the linearity and stationarity in UR data over time [9], [10], [11]. However, these methods are inherently limited to individual country forecasts. Artificial neural networks (ANN) were also used to analyse time series. This was either done using feedforward neural network (FNN) [12], or recurrent neural network RNN which is able to capture more complex patterns [13].

A combination of different models forms hybrid models which were also used to analyse UR. Such models leveraging the strengths of both linear and non-linear modelling capability. Hybrid model includes using ARIMA-RNN [13], [14], RNN with genetic algorithm GA for feature optimization [13].

This research aims to build a comprehensive model that forecasts UR across multiple countries. Leveraging a new method for structuring time series data, this study addresses the challenge of building a single forecasting model applicable to multiple regions. The data, sourced from the World Bank, includes a diverse set of macroeconomic and microeconomic features for 44 countries over a span of 33 years. By segmenting the time series into smaller data slices, the proposed approach allows for the consolidation of country-specific data into a unified dataset, facilitating the development of a generalized forecasting model.

The remainder of this paper is organized as follows: Section II reviews the related work on UR forecasting methodologies, including machine learning, autoregressive models, and hybrid approaches. Section III details the methodology of the data structuring process and model development. Section IV presents the results and evaluates the performance of each model. Section V discusses the implications of the findings, and Section VI concludes the paper with future research directions.

II. Related Work

The study of unemployment rate (UR) forecasting has evolved significantly over the past decades. Researchers have applied various methods, primarily categorized into traditional machine learning regression, autoregressive time series forecasting, neural networks, and hybrid models. Each approach has distinct advantages and has contributed to understanding the dynamics of UR within individual countries.

A. Machine Learning Regression for Forecasting

Several studies used stochastic time series techniques in forecasting. In such techniques, they study the effect of independent variables on dependant variables to examine if there is significant difference. Zawojska used linear regression to study the correlation between several macroeconomic factors and UR and found significant correlation with many factors such as GDP [7]. In [15] Alternative regression methods, including the Toda-Yamamoto procedure, have been utilized to examine the impact of specific factors, such as oil prices and interest rates, on the unemployment rate (UR) in different scenarios.

Before applying these techniques, researchers often test the data for stationarity to ensure appropriate modeling. For example, a study involving 10 developing countries found that UR, GDP, population, remittances, external debt, exchange rate and expenditure on education were non-stationary, necessitating differencing for proper regression analysis [16]. This type of tests is also applied on autoregression time series forecasting (see B).

B. Autoregression Time Series Forecasting

Autoregressive time series methods focus on the inherent symmetry or asymmetry, as well as the linearity or non-linearity, of UR over time. If a time series exhibits stationarity—consistent mean or variance over time—these methods can effectively forecast future trends[17]. It was suggested that UR has hysteresis, which means that it stays in equilibrium level with equality likely movement in either direction [18]. For instance, the logistic smooth transition autoregressive (LSTAR) model has been used to analyze non-linear and asymmetric UR data in Australia [19]. In this study, model accuracy was high due to LSTAR ability to capture cyclic fluctuations of business and UR as result. Other studies analyse the effect over years using Autoregressive Distributed Lag Stationarity ADLS [20].

Autoregressive integrated moving average (ARIMA) model is very famous for time series forecasting. A modified version (SARIMA) captures seasonality and used in case time series has seasonality element. Another variation is generalised autoregressive conditional heteroscedasticity (GARCH) that analyse the volatility of time series [17]. The authors in [9] A study of monthly UR to study the impact of COVID-19 in 5 Asian countries (ASEAN-5) using ARIMA, SARIMA and GARCH found that ARIMA and SARIMA are superior in forecasting UR. Another study [21] using ARIMA model with monthly data in Australia was found to predict on month ahead have R^2 of 0.96.

C. Artificial Neural Network for Forecasting

Neural networks (NNs) have emerged as powerful tools for forecasting time series data, particularly when the data exhibits non-linear characteristics [22], which is found in UR [23]. Feedforward neural networks (FNN) have been applied both to evaluate independent variables affecting UR and to utilize lagged values (k-lags) for one-step ahead forecasting [24]. A study applied FNN by using several macroeconomic factors including GDP, inflation and other factors to forecast UR in Philippine using a data from 1991 to 2014 [25] with accuracy of 87.5%. Another study used neural network with backpropagation to predict UR, gross national product and employee number in Germany and found that the model has R^2 close to 1 when predicting three lags ahead (t+3) [12]. In [17] Recurrent neural networks (RNN) and their variations, such as long short-term memory (LSTM), have demonstrated higher accuracy in forecasting due to their internal memory capabilities, which are well-suited for for time series forecasting.

D. Hybrid Models

Hybrid models combines different types of models to leverage the strengths of both linear and non-linear modeling techniques to improve forecasting accuracy [26]. For example, combining two models where one of them to capture linear elements (such as ARIMA or regression) and other to capture nonlinear elements (such as artificial neural network) could provide a better performance model. Other studies used hybrid models differently by combining neural network models, with feature optimization models, such as genetic algorithms [13]. Hybrid models has been used for time series forecasting, such as stock market price forecasting [27]. It was also used to forecast UR such as combining ARIMA and RNN [14]. They used ARIMA in the first phase to catch linear pattern, followed by autoregressive neural network ARNN to analyse nonlinearity and nonstationarity patterns. This hybrid model outperformed either of these models alone. Hybrid models can be employed for feature optimization and integrated with other approaches. In [13] LSTM models enhanced with genetic algorithms (GA) for parameter optimization have demonstrated superior performance compared to traditional RNNs in unemployment rate forecasting for Ecuador, delivering greater accuracy.

E. Gaps in the Literature

Previous studies were used to build separate UR forecasting models for individual countries. The reason for this is that they used time series forecasting, where data is arranged on a continuous line and is split for training and testing the model. This way makes it not possible to incorporate multiple countries inti the same model. This research addresses this gap by proposing a novel data structuring technique, enabling the construction of a single forecasting model for multiple countries.

III. Data and Method

This study aims to develop a unified AI model that leverages macroeconomic and microeconomic data across multiple countries to predict unemployment rates. While previous research has focused on models for individual countries [8], [16], [28], this study introduces a novel data structuring approach to create a forecasting model applicable to various economic contexts and regions.

A. Data Analysis

This study utilized data from the World Bank [29], covering annual macroeconomic and microeconomic indicators from five key datasets: economy and growth, education, urban development, health, and climate change. The data spans the years 1991 to 2023, with the analysis focusing on 44 countries that have complete records for this period. Initially, 162 features were drawn from the datasets, but after excluding those with missing values, 153 features were retained. Table I outlines the datasets, their feature counts, and examples of the included variables.

TABLE I. DATASETS USED IN THE RESEARCH, AND EXAMPLES OF THEIR FEATURES

Dataset	Feature count	Examples of features
Economy & Growth [30]	81	GDP, current account balance and exports of goods and services
Education [31]	5	UR, government expenditure on education, school enrolment, literacy rate and population age groups
Urban Development [32]	7	mortality by traffic, PM2.5 air pollution, population in large cities and urban population
Health [33]	62	fertility rate per age group, birth rate, cause of death by injury, cause of death by communicable diseases and death rate
Climate Change [34]	7	access to electricity, agriculture land, cereal yield, annual fresh water, CO2 emission and NO emissions

B. Data Structuring

A new method of structuring the time series data was developed to enable multiple time series forecasting within a single model. The data were segmented into rolling windows or "slices," each containing three years of lagged values for each feature and UR value for target year. For example, the UR for a given year (y) was predicted using the three preceding years (y-1), (y-2), (y-3) for each feature. This means that 3 columns are added per feature to represent $feature_{y-1}$, $feature_{y-2}$ and $feature_{y-3}$ to forecast UR_y. And because we have data from 1991 to 2022, we can have 3 lags to forecast UR starting from 1994 to 2022. (i.e. the first UR will have $feature\_a_{1991}$, $feature\_a_{1992}$ and $feature\_a_{1993}$ to forecast UR_{1994}). This approach resulted in 29 slices per country, allowing the aggregation of all 44 countries into a unified dataset with 1,276 observations. The reason for using 3 years of lagged features is that factors might have effect on UR for multiple years in the future [7], [16]. This way, the forecasting model will predict UR of any year (y) based on previous n years lag (y-3) of any given feature.

This method effectively addresses the limitations of traditional time series forecasting, which typically requires contiguous, long time series data and is not easily adaptable for multi-country analysis.

The data were imported using Python's Jupyter Notebook, leveraging Pandas and NumPy APIs. Each dataset was structured into a multi-index DataFrame, where level 0 represented the country, and level 1 denoted the year. This organization ensured that each country had 33 rows, corresponding to the years 1991 to 2023. The data for 2023 was excluded from the analysis to serve as out-of-sample validation [35]. Fig. 1 provides a snapshot of the DataFrame used, which contained 1276 observations and 153 features. The total number of columns includes 1 target variable (unemployment rate), along with 3 lags per feature for each of the 153 features.

Country	Indicator Code / Year	Unemployment	NE.CON.GOVT.CD y-1	NE.CON.GOVT.CD y-2	NE.CON.GOVT.CD y-3	...
Argentina	1994	11.76	31984701702	6807403684	6302044992	...
	1995	18.8	33948366600	31984701702	6807403684	...
	1996	17.11	34445834100	33948366600	31984701702	...
	1997	14.82	34023284500	34445834100	33948366600	...
...	...	...	...	...	...	...
United Kingdom	2018	4	5.00097E+11	5.168E+11	5.72364E+11	...
	2019	3.74	5.32384E+11	5.00097E+11	5.168E+11	...
	2020	4.472	5.44508E+11	5.32384E+11	5.00097E+11	...
	2021	4.826	6.09767E+11	5.44508E+11	5.32384E+11	...
	2022	3.73	6.99684E+11	6.09767E+11	5.44508E+11	...

1276 rows x 460 columns

Fig. 1. Image of the actual DataFrame used in the research.

C. Model Development

As outlined in Table II, four AI-based models were developed and evaluated:

Model 1 (ANN-All): Applied artificial neural networks (ANN) to all available features.

Model 2 (Hybrid ANN-ML): An ensemble model where separate neural networks were trained on each of the five datasets, and their outputs served as inputs for a machine learning regression model.

Model 3 (Hybrid ANN-GA): A hybrid approach integrating genetic algorithms for feature selection with an ANN for forecasting.

Model 4 (UR-ML): A machine learning model relying solely on lagged unemployment rate (UR) values (y-1, y-2, y-3) to forecast the following year's UR.

TABLE II. MODELS AND FEATURES USED

Model	*# Datasets*	*# countries*	*# features*
ANN-All	5	44	153
ANN-ML	5	44	153
ANN-GA	5	44	153
UR-ML	Unemployment Rate only	44	1 (UR)

The models were trained and validated on the structured dataset, with the testing conducted on out-of-sample data for 2023. Training and validation were split into 77.3% and 19.4% of the data, respectively, while the testing set comprised 3.3%.

D. Performance Evaluation of Methods

The models' performance was evaluated using four metrics:

1) The coefficient of determination (R2): Assess the proportion of variance explained by the model. One of the main advantages of R^2 is that there is no need to compare the result with other models to evaluate model fit [36]. Best value equals 1. This was the main metric used in this paper.

2) *Accuracy: defined as (% 1 – MAPE) [37].* Mean Absolute Percentage Error (MAPE) has high efficiency to minimize risk in regression forecasting, especially if target value is always positive (which is the case with UR) [38], [39]. Using (% 1 – MAPE) provides easy way to assess model accuracy. Best value equals 100%.

3) *Root mean squared error (rmse) [36]: Best value equals 0.*

4) *Mean absolute error (MAE) [36]: Best value equals 0.*

IV. Experimental Results

This section represents the results of each of the models used in the research and evaluate their performance.

A. Results

The performance of each model was assessed on validation datasets using R^2, accuracy (1-MAPE), RMSE and MAE as summarized in Table III and IV. Among the models, the UR-ML model, which uses only lagged UR values, demonstrated the best performance, achieving an R^2 of 0.964 and 0.989 on validation and testing data, respectively. Its accuracy was 89.8% on validation data and 94.2% on testing data, indicating strong generalizability and robustness.

Following is a summary of each model performance.

1) *Model 1: ANN-All – Neural network for 5 datasets:* This model used 5 datasets with a total of 153 features to predict UR. This model performs better than model in in all 4 measures. In validation data, this model scored 0.927 and 82.1% of R^2 and accuracy respectively. These measures fell greatly to 0.857 and 71.4% respectively which means that the model has issue with generalization. The scatter plot in Fig. 2 plotted predicted versus true value for this model.

2) *Model 2: ANN-ML – Hybrid model using ANN and ML:* This model built five individual Neural Network model for each of the five datasets to predict UR. Fig. 3 shows the result of these five individual models to predict UR.

TABLE III. Performance of Models for Forecasting UR (Validation Data)

Model	R^2	Accuracy	RMSE	MAE
ANN-All	0.927	82.1%	1.217	0.894
ANN-ML (hybrid)	0.921	77.4%	1.256	0.985
ANN-GA (hybrid)	0.945	85.1%	1.049	0.775
UR-ML	0.964	89.8%	0.851	0.563

TABLE IV. Performance of Models for Forecasting UR (Testing Data)

Model	R^2	Accuracy	RMSE	MAE
ANN-All	0.857	71.4%	1.745	1.1401
ANN-ML (hybrid)	0.86	72.8%	1.729	1.365
ANN-GA (hybrid)	0.919	80.4%	1.316	0.986
UR-ML	0.989	94.2%	0.487	0.316

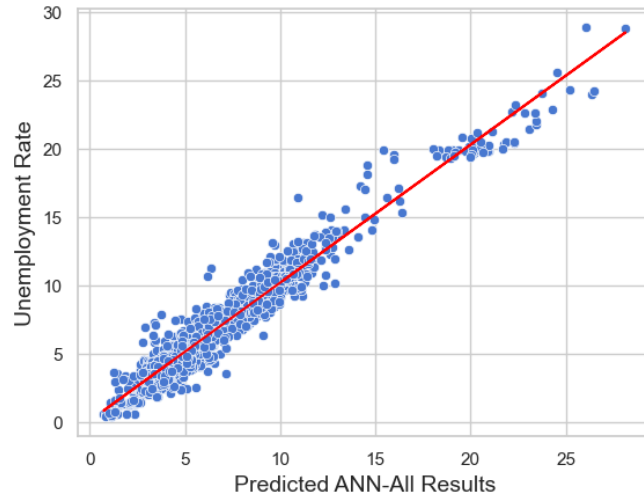

Fig. 2. ANN-All Predicted vs. True value.

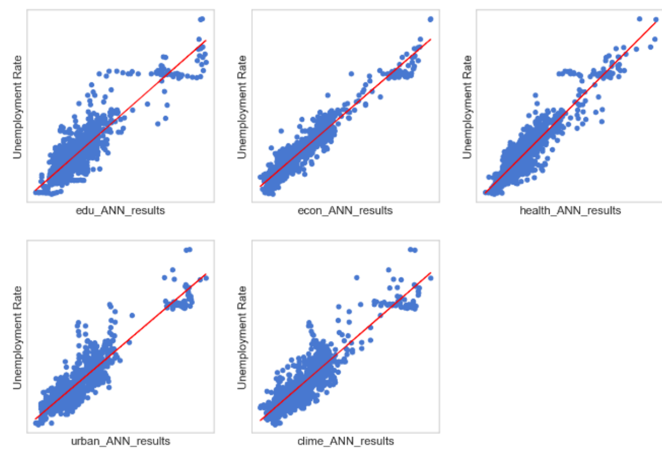

Fig. 3. The result of five individual ANN models to forecast unemployment rate.

Then, the resulted five predictions were fitted using machine learning linear regression to predict UR. The model has R^2 value of 0.921 which is similar to ANN-All, but has worst accuracy value of 77.4%. The model is also slightly worse than ANN-All in terms of RMSE and MAE. The model, however, is a little better in terms of generalization compared to ANN-All, as R^2 and accuracy for testing data were 0.86 and 72.8% respectively. Fig. 4 shows the predicted versus true value for this model.

3) *Model 3: ANN-GA:* This model used genetic algorithm for feature optimization, followed by building neural network model using best fitness features measured by R^2. Based on genetic algorithm result, 83 out of 153 features were selected and then used to build neural network model. The model result was better than all previous 3 models as it has R^2 of 0.945 and accuracy of 85.1% in validation data, and R^2 of 0.919 and accuracy of 80.4%. This means that those 83 features provide better performance compared to using all 153 features as in Model 1 ANN-All. Fig. 5 shows predicted versus true value for this model.

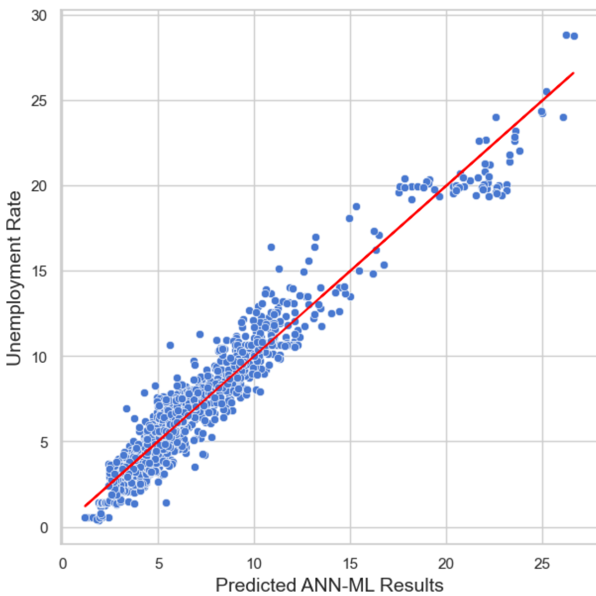

Fig. 4. ANN-ML Predicted vs. True value.

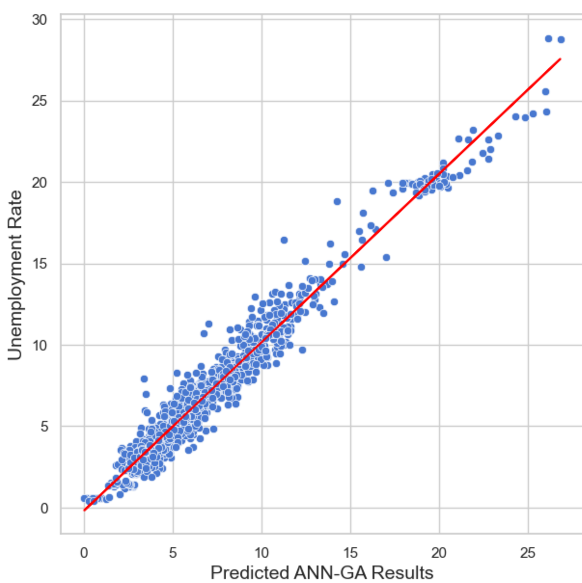

Fig. 5. ANN-GA Predicted vs. True values.

4) Model 4: UR-ML – Using ML regression for UR: The fifth model used only 1 feature, UR. And as all previous models, it uses this feature with 3 lags (UR_{y-1}, UR_{y-2} and UR_{y-3}) to predict UR_y. This model has shown the best overall performance in both validation data and testing data. In validation data it has R^2 of 0.964 and accuracy of 89.8%. This model demonstrated high generalization capability, as it scored 0.989 in R^2 and 94.2% in accuracy. This means that this model doesn't suffer from overfitting. This model also has the least score in RMSE and MAE, which were 0.487 and 0.316 respectively.

To illustrate the reason behind the high performance of this model, Fig. 6 shows the correlation of the n lags of UR (y-1, y-2 and y-3) with true UR. Each one of UR_{y-1}, UR_{y-2} and UR_{y-3} has high correlation with UR_y, making it very suitable for UR forecasting. Fig. 7 shows the predicted versus true value for this model.

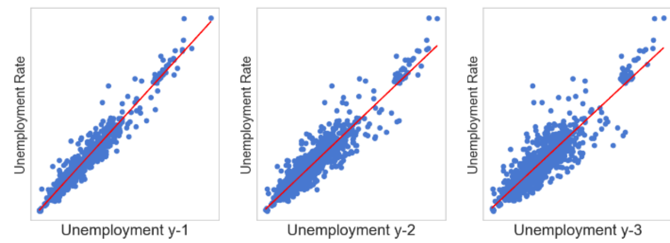

Fig. 6. Correlation between n lags UR and True UR.

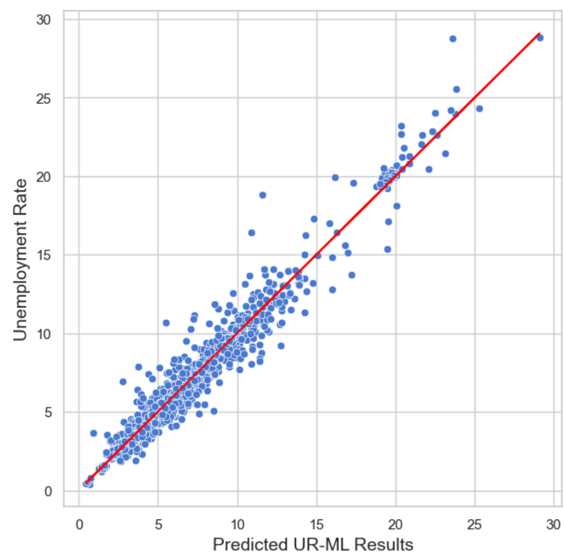

Fig. 7. UR-ML Predicted vs. True value.

B. Discussion

This study demonstrates that it is feasible to develop a single AI model that can accurately forecast UR across multiple countries using economic factors. The effectiveness of the new data structuring method, which segments time series into smaller slices, allows for a generalized model that maintains high accuracy and generalizability.

1) Discussing models performance: Neural network models are well-suited for UR forecasting due its nonlinear nature [23]. This research found that it is possible to build single neural network model that can be used effectively to forecast UR across several countries.

When comparing with other researches that use neural network model, we found that our model with R^2 of 0.87 and accuracy of 71.4% is compared to other models. Misil and Tarepe have built ANN model with 11 features (including GDP and inflation rate) to forecast UR in Philippine from 1991 to 2014 and has accuracy of 87.5% [25]. Their model has higher accuracy due to building the model for single country.

Our research found that hybrid model with genetic algorithm outperforms other models neural network models in forecasting UR as it reduce the number of features used and hence reduce model noise. This is consistent with other research that found this models superior to other models they compared [13], which used LSTM (Long Short Term Memory) with GA. However, their research has accuracy of 96% compared to our 80.4%. The superiority of their results is due to building the model on single country only (Ecuador) in addition to include UR (UR of previous years) as a feature, which will yield better results as will be discussed in 2).

The other hybrid model used in the research was ANN-ML, which combines neural network forecasting with machine learning regression. This model showed inferior results especially in terms of generalization on testing data with accuracy of 72.8%. This is not consistent with researches who used hybrid models in UR forecasting [14]. The reason is that we used different hybrid model. Chakraborty et al. [14] has combined ARIMA with RNN to study UR in individual countries, which wasn't applicable in our research to build single model for multiple countries.

Several researchers have found much superior results with different hybrid models. A study on individual countries including US, UK, Italy and France had accuracy ranges from 95.34 to 98.91% based on the country using a hybrid model of LSTM-GRU [40]. Another studied Asian countries using ARIMA-ANN model and had accuracy of 96.4 - 97.8% [41]. However, as these models used previous UR data only in forecasting, these should be compared to our UR-ML model (discussed in 2)), and their result is comparable to our 94.2% for UR-ML model.

2) Performance of ML model using UR only: Using UR from the past to forecast future UR is very common across research. Research have used several methods for such type of analysis including ARIMA (and its variations including ARMA, SARIMA, FARIMA, etc.), RNN (and its variations including GRU, LSTM, etc.).

In our research, the regression model that only uses UR (UR-ML) outperforms all other models. The model used 3 lags years (UR_{y-1}, UR_{y-2} and UR_{y-3}) to forecast UR_y result. The model has R^2 of 0.964 and accuracy of 89.8% on validation data. The model performs even better on testing data to forecast 2023 results with R^2 of 0.989 and accuracy of 94.2% which means it has high generalizability.

This result is aligned with many researches that were able to use previous UR solely to forecast future UR. Our research outperformed Davidescu et al. research who have used SARIMA model to forecast UR in Romania, and has accuracy of 86.6%. The same research has even worse accuracy using SETAR model, with accuracy of 84.23% [39]. As ARIMA model requires a continuous time series, and then using 80% of it for training and keeping the last 20% for testing. This makes the model more vulnerable to macro trends that occurs in last year that didn't exist in training data.

Higher accuracy could be obtained by building separate model for separate country or district. In Germany, accuracy ranged from 91.91% to 99.23% using SARIMA model [42]. Our model has slightly inferior results despite being built for multiple countries.

One of the reasons of why UR from previous year is a strong indicator of future UR, is that UR tends to change gradually and slowly over years, as it tend to stay in equilibrium [18]. To test this, we have created a dummy forecaster, where it forecast UR_y by simply returning UR_{y-1}. Fig. 8 shows a scatter plot of this Dummy model. This dummy model has R^2 of 0.958 and accuracy of 90.7%, which outperforms all 3 ANN models in this research, which hasn't included lagged UR as feature.

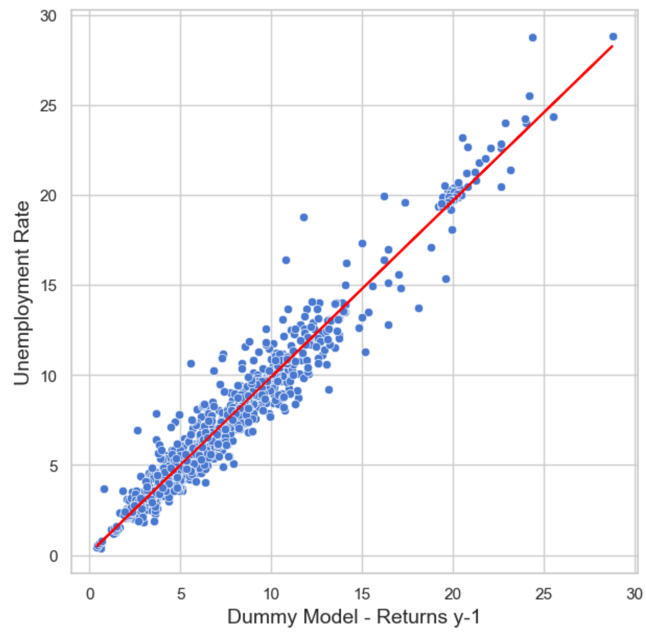

Fig. 8. Dummy Model that uses Y-1 as forecast value for Year Y.

For this reason, this research excluded UR from all ANN models, due to the high correlation between UR_{y-n} and UR_y. Also, this is the possible reason why ARIMA models have higher accuracy in one-step-ahead forecasting compared to multiple-steps-ahead.

3) Benefits of using new data structure: Methods like ARIMA and RNN mandates the presence of long and contiguous the time series [38], [43]. Because we only have 33 data point per country, we need to increase the number of observations to fit the model.

This research has used new way of structuring data by transforming continuous time series into an array of instances that contains slices of 3 years of lagged data. This method has the following benefits:

- This method enables mixing the data randomly for fitting the model. This has advantage of training data on a mix of time periods and make prediction more generalizable.
- As the Covid pandemic has made significant change in the labour market in 2020 and 2021 [9], [39].This make forecasting issue if out-of-sample forecasting included this period. This accounts for generalizability issue of the model and big difference between training and testing data conditions [39].

- Previous studies were forced to analyse each country separately [8], [14], [16], [28], [38], [40], [41], [44], [45], or even on separate county or city level [46] because of the limitation of ARIMA and RNN models. The new method enables us to train the model on multiple countries. This resulted in successfully building single forecasting model for multiple countries (in our case 44 countries).

- New approach enables us to have large number of observations to train and validate the model. Other studies either used the low number of instances [25] or increased the number of instances by using months instead of years [23], [39], [46]. Using monthly time periods increases the number to a certain degree, while adding another layer of complexity due to seasonality of UR.

However, this approach keeps memory of n lags of years to perform one step ahead forecasting. Other time series analysis methods keep longer memory which may increase the accuracy of multiple steps ahead forecasting [12].

V. Conclusion

This study successfully developed a robust AI-driven model for forecasting unemployment rates (UR) across multiple countries by employing an innovative data structuring technique. By segmenting time series data into smaller slices and aggregating data from 44 countries, the study produced a unified model demonstrating high accuracy and generalizability. Among the tested models, the UR-ML model emerged as the best performer, with the hybrid ANN-GA model also showing considerable potential for feature optimization. The proposed method holds promise for forecasting other economic indicators, such as GDP and inflation. However, this study is not without limitations. Firstly, the dataset excluded data from 2023 as it was reserved for out-of-sample validation, limiting the availability of real-time forecasting scenarios. Secondly, the exclusion of additional regional and socio-economic indicators, due to data availability constraints, may have impacted the model's ability to fully capture cross-country variations. Future research should explore advanced hybrid models, such as incorporating deep learning architectures (e.g., LSTM-GA combinations) to improve predictive capabilities further. Expanding the feature set to include additional socio-economic and environmental indicators may enhance accuracy and capture regional variations better. Additionally, extending the model to support multi-step-ahead forecasting and applying the proposed data structuring method to other macroeconomic indicators, such as inflation or GDP, would demonstrate its versatility.

Acknowledgment

This research was supported by the researchers supporting project number [RSPD2024R995], King Saud University, Riyadh, Saudi Arabia.

References

[1] Y. Wu, “The Importance of Studying the Unemployment Rate in China,” Journal of Education, Humanities and Social Sciences, vol. 24, pp. 465–470, Dec. 2023, doi: 10.54097/9maq2b67.

[2] P. M. Pincheira and A. M. Hernández, “Forecasting Unemployment Rates with International Factors,” Available at SSRN 3510597, 2019.

[3] G. C. Rodrigo, “MICRO AND MACRO: THE ECONOMIC DIVIDE,” International Monetary Fund. Accessed: Feb. 28, 2024. [Online]. Available: https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/Micro-and-Macro

[4] A. E. DePrince and P. D. Morris, “The Effects of Education on the Natural Rate of Unemployment,” Business Economics, vol. 43, no. 2, pp. 45–54, 2008, doi: 10.2145/20080205.

[5] L. Kilimova and O. Nishnianidze, “Socio-demographic characteristics as indicators of the unemployment rate,” Economic annals-XXI, no. 168, pp. 82–85, 2017.

[6] W. Baah-Boateng, “Determinants of unemployment in Ghana,” African Development Review, vol. 25, no. 4, pp. 385–399, 2013.

[7] A. Zawojska, “Effect of Macroeconomic Variables on Unemployment Rate in Poland,” ECONOMIC SCIENCE FOR RURAL DEVELOPMENT, 2010.

[8] G. Gozgor, “The impact of trade openness on the unemployment rate in G7 countries,” J Int Trade Econ Dev, vol. 23, no. 7, pp. 1018–1037, Oct. 2014, doi: 10.1080/09638199.2013.827233.

[9] K. Y. Ng, Z. Zainal, and S. Samsudin, “COMPARATIVE PERFORMANCE OF ARIMA, SARIMA AND GARCH MODELS IN MODELLING AND FORECASTING UNEMPLOYMENT AMONG ASEAN-5 COUNTRIES,” International Journal of Business and Society, vol. 24, no. 3, pp. 967–994, 2023.

[10] P. E. Tengaa, Y. M. Maiga, and A. M. Mwasota, “Modeling and forecasting unemployment rate in Tanzania: an ARIMA approach,” Journal of accounting, finance and auditing studies, vol. 9, no. 3, pp. 270–288, 2023.

[11] K. Vu, The ARIMA and VARIMA Time Series: Their Modelings, Analyses and Applications. 2007.

[12] B. Freisleben and K. Ripper, “Economic forecasting using neural networks,” in IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS PROCEEDINGS, NEW YORK, 1995, pp. 833–838 vol.2. doi: 10.1109/ICNN.1995.487526.

[13] K. Mero, N. Salgado, J. Meza, J. Pacheco-Delgado, and S. Ventura, “Unemployment Rate Prediction Using a Hybrid Model of Recurrent Neural Networks and Genetic Algorithms,” Applied Sciences, vol. 14, no. 8, p. 3174, 2024.

[14] T. Chakraborty, A. K. Chakraborty, M. Biswas, S. Banerjee, and S. Bhattacharya, “Unemployment Rate Forecasting: A Hybrid Approach,” Comput Econ, vol. 57, no. 1, pp. 183–201, 2021, doi: 10.1007/s10614-020-10040-2.

[15] H. G. Doğrul and U. Soytas, “Relationship between oil prices, interest rate, and unemployment: Evidence from an emerging market,” Energy Econ, vol. 32, no. 6, pp. 1523–1528, 2010.

[16] A. Siddiqa, “Determinants of unemployment in selected developing countries: A panel data analysis,” Journal of Economic Impact, vol. 3, no. 1, pp. 19–26, 2021.

[17] C. Huang and A. Petukhina, Applied Time Series Analysis and Forecasting with Python. Springer, 2022.

[18] O. J. Blanchard and L. H. Summers, “Hysteresis in unemployment,” in Economic Models of Trade Unions, Springer, 1986, pp. 235–242.

[19] M. Stevenson and M. Peat, “Forecasting Australian unemployment rates,” Australian Journal of Labour Economics, vol. 4, no. 1, pp. 41–55, 2001.

[20] M. S. Maqbool, T. Mahmood, A. Sattar, and M. N. Bhalli, “Determinants of unemployment: Empirical evidences from Pakistan,” Pak Econ Soc Rev, pp. 191–208, 2013.

[21] P. J. Wilson and L. J. Perry, “Forecasting Australian unemployment rates using spectral analysis,” Australian Journal of Labour Economics, vol. 7, no. 4, pp. 459–480, 2004.

[22] Y. Kajitani, K. W. Hipel, and A. I. McLeod, “Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian lynx data,” J Forecast, vol. 24, no. 2, pp. 105–117, 2005.

[23] S. Moshiri and L. Brown, “Unemployment variation over the business cycles: a comparison of forecasting models,” J Forecast, vol. 23, no. 7, pp. 497–511, 2004.

[24] G. P. Zhang and V. L. Berardi, "Time series forecasting with neural network ensembles: an application for exchange rate prediction," Journal of the operational research society, vol. 52, pp. 652–664, 2001.

[25] D. D. Misil and D. A. Tarepe, "A Time Series Forecasting of the Philippine Unemployment Rate Using Feed-Forward Artificial Neural Network.," Liceo Journal of Higher Education Research, vol. 14, no. 1, pp. 58–81, 2018.

[26] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," Appl Math Lett, vol. 22, no. 9, pp. 1467–1470, 2009, doi: https://doi.org/10.1016/j.aml.2009.02.006.

[27] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," Omega (Westport), vol. 33, no. 6, pp. 497–505, 2005.

[28] F. Pennoni and B. Bal-Domańska, "NEETs and Youth Unemployment: A Longitudinal Comparison across European Countries," Soc Indic Res, vol. 162, no. 2, pp. 739–761, 2022, doi: 10.1007/s11205-021-02813-5.

[29] The World Bank, "World Bank Open Data: Free and open access to global development data," The World Bank. Accessed: Mar. 05, 2024. [Online]. Available: https://data.worldbank.org

[30] The World Bank, "Economy & Growth," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/economy-and-growth?view=chart

[31] The World Bank, "Education," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/education?view=chart

[32] The World Bank, "Urban Development," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/urban-development?view=chart

[33] The World Bank, "Health," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/health?view=chart

[34] The World Bank, "Climate Change," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/climate-change?view=chart

[35] M. Joseph, Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning. Packt Publishing Ltd, 2022.

[36] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput Sci, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[37] C. Lee, F. Amini, G. Hu, and L. Halverson, "Machine Learning Prediction of Nitrification From Ammonia- and Nitrite-Oxidizer Community Structure," Front Microbiol, vol. 13, Jul. 2022, doi: 10.3389/fmicb.2022.899565.

[38] K. Dumičić, A. Čeh Časni, and B. Žmuk, "Forecasting unemployment rate in selected European countries using smoothing methods," World Academy of Science, Engineering and Technology: International Journal of Social, Education, Economics and Management Engineering, vol. 9, no. 4, pp. 867–872, 2015.

[39] A. A. Davidescu, S.-A. Apostu, and A. Paul, "Comparative analysis of different univariate forecasting methods in modelling and predicting the romanian unemployment rate for the period 2021–2022," Entropy, vol. 23, no. 3, p. 325, 2021.

[40] M. Yurtsever, "Unemployment rate forecasting: LSTM-GRU hybrid approach," J Labour Mark Res, vol. 57, no. 1, p. 18, 2023.

[41] L. Shi, Y. A. Khan, and M.-W. Tian, "COVID-19 pandemic and unemployment rate prediction for developing countries of Asia: A hybrid approach," PLoS One, vol. 17, no. 12, p. e0275422, 2022.

[42] A. Vosseler and E. Weber, "Forecasting seasonal time series data: a Bayesian model averaging approach," Comput Stat, vol. 33, pp. 1733–1765, 2018.

[43] A. L. Montgomery, V. Zarnowitz, R. S. Tsay, and G. C. Tiao, "Forecasting the U.S. Unemployment Rate," J Am Stat Assoc, vol. 93, no. 442, pp. 478–493, Jun. 1998, doi: 10.1080/01621459.1998.10473696.

[44] C. Katris, "Prediction of unemployment rates with time series and machine learning techniques," Comput Econ, vol. 55, no. 2, pp. 673–706, 2020.

[45] E. Olmedo, "Forecasting spanish unemployment using near neighbour and neural net techniques," Comput Econ, vol. 43, pp. 183–197, 2014.

[46] M. Wozniak, "Forecasting the unemployment rate over districts with the use of distinct methods," vol. 24, no. 2, 2020, doi: doi:10.1515/snde-2016-0115.