APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 7, 2023, 168-177

# Forecasting Inpatient and Outpatient Visits for Depressive Disorders: A Comparative Study of Deep Learning Approaches

Nur Izzati Ab Kader [1*], Umi Kalsom Yusof [1] and Mohd Nor Akmal Khalid[2]

[1]School of Computer Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
[2]School of Information Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, 923-1292, Ishikawa, Japan

[*]Corresponding author: nurizzati.ucom13@student.usm.my

*Submitted 27 August 2023, Revised 24 October 2023, Accepted 29 October 2023, Available online 18 November 2023.*


**Abstract:** Forecasting inpatient and outpatient visits is essential for successful resource allocation and clinical decision-making. The techniques used by previous researchers for forecasting, primarily based on statistical approaches, which often require extensive data preprocessing and expert knowledge, can be time-consuming and difficult. Therefore, this study analyses three current deep learning (DL) algorithms, recurrent neural network (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU), for forecasting inpatient and outpatient visits for depressive disorders. These algorithms are among the most familiar DL techniques for time series and have been used with remarkable success in various contexts. The DL algorithms were evaluated using mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE). Based on the results comparison, LSTM has the best performance (lowest error values) compared to the RNN and GRU. The DL algorithms are also being compared to state-of-the-art algorithms, and the results show that the DL algorithms can accurately forecast inpatient and outpatient visits compared to the previously proposed algorithms. The findings from this study could be helpful in clinical decision-making and resource allocation in mental health care.

Keywords: Deep learning; Depression; Psychiatric Department; Time series forecasting.

## 1. INTRODUCTION

Forecasting is a technique of making future predictions based on knowledge acquired in the past through a methodical procedure or intuition. Forecasting has evolved and is widely used in many disciplines, such as business and economics, environment (including meteorology), sports, technology and politics, and health [1, 2]. A crucial area of forecasting is health forecasting, which is also a beneficial application for predicting future situations or events related to health, such as the demand for healthcare services and the need for treatment. It promotes preventative care and intervention methods by directing healthcare providers to take the necessary precautions to reduce risks and manage demand in advance [3, 4]. Considering the rise in life expectancy and diseases, healthcare authorities must effectively manage their resources to offer optimal care to patients. When the demand for resources surpasses the supply, hospitals become overcrowded, leading to adverse patient outcomes such as longer wait times and bed shortages [5]. Currently, there is extensive research on health forecasts for diseases like COVID-19, asthma, brucella, cardiorespiratory diseases, and dengue. Typically, health prognosis relies on routine data, daily data kept in the electronic patient record. Management in the hospital could be broadly divided into inpatient and outpatient visits. Hospital admission occurs when a medically qualified decision-maker determines that a patient requires inpatient treatment, when the patient is admitted by an inpatient specialist service for further treatment, and when the patient is administratively admitted to the hospital. Apart from this, outpatient treatment refers to patients who are treated but do not require hospitalization.

The previously mentioned health prognosis is categorised under physical illness. The current research revealed that only a few health forecasts pertained to mental or psychiatric conditions. Psychiatric illness, characterised by distressing behavioural and psychological symptoms or impairments in functioning, refers to severe clinical syndromes [6]. In addition, around 500 million people worldwide have a mental illness or disability, with over 150 million people suffering from depression and 24 million from schizophrenia. Additionally, billions of people abuse drugs, and 800,000 people commit suicide each year. Mental and behavioural issues are responsible for 12% of the global disease burden. It has been reported by the World Health

Organization (WHO) regarding an increase in the occurrence of psychiatric disorders globally [7]. A greater focus is highlighted on prominent psychiatric disorders, the primary and leading illnesses within the psychiatric category. One of the most prominent psychiatric disorders is depression, which has become increasingly prevalent. Statistics suggest that depression affects over 264 million individuals across the world. Signs of a depressed person include indifference to daily life and disinterest in activities, even if it is a hobby. People who are depressed are also less interested in social contacts and are afraid to meet other people.

Despite the increasing prevalence of depressive disorders, very few studies currently look at the prediction of depressive disorders. So far, only one study on this topic has been conducted [8]. Other studies conducted in the psychiatry department were related to other diseases. In addition, the study by [8] used aggregated data, where data from inpatients and outpatients had already been combined. However, looking at each treatment type in detail is essential to get more information rather than uniting the data directly, which is more general. Taking both data into account is vital to make a comprehensive forecast.

Previously, regression analysis and time series techniques, e.g., Poisson regression, autoregressive integrated moving average (ARIMA), negative binomial regression and generalised additive model (GAM), have been used to solve the problem [9, 10]. Numerous machine learning (ML) models are offered in literature and the community with specific components and architectures to address the sequential nature of the inputs. ML, such as random forest (RF) support vector machine (SVM), has also been used by previous researchers. However, current methods are restricted in their capability to model intricate non-linear data.

In the past few years, DL techniques have gained immense popularity and are widely used to address various issues, including the prediction of time series [11–14]. DL has shown the capacity to simulate complicated non-linear feature interactions, unlike traditional statistics-based models that can only represent linear correlations in the data [15]. Modern neural systems base their success on their deep structure, which stacks several layers on top of each other and densely connects many neural connections.

The increase in computing power in recent years has enabled the development of deeper models that have significantly improved their learning capabilities compared to shallow networks [16]. Moreover, adapting to the data immediately, without making prior assumptions, offers significant advantages when dealing with limited time series information. Therefore, this study proposes to develop a forecast of inpatient and outpatient visits for depressive disorders. This study aims to analyse the performance of deep learning models for predicting inpatient and outpatient visits due to depressive disorders. The contribution of this study lies in several aspects. First, it provides a more detailed analysis of the prediction of inpatient and outpatient visits compared to the previous study. In addition, it assesses the usefulness of DL algorithms in predicting inpatient and outpatient visits for depressive disorders, a topic that has not been thoroughly investigated in previous research. The literature indicates that RNN, LSTM and GRU have been successfully used in many domain studies and have achieved a good result [17, 18]. Furthermore, the paper provides evidence that DL algorithms, especially the LSTM algorithm, can accurately predict visits to mental health facilities, highlighting their potential use in clinical decision-making and resource allocation in psychiatric treatment.

This paper is divided into the following sections: Previous research on prediction in psychiatry and DL models is covered in Section 2, followed by an explanation of the proposed approach in Section 3, and finally, the findings are presented in Section 4. A discussion of the findings can be found in Section 5, and the study is concluded in Section 6.

## 2. RELATED WORKS

### 2.1 Forecasting in Psychiatric Department

Previous studies have done substantial work in forecasting the psychiatric department. The studies on predicting psychiatric departments can be divided into general and specific. In the general studies, all types of psychiatric disorders have been aggregated [19, 20]. For the specific category of studies, predictions are made based on individual studies, e.g., predictions based on bipolar disorder [8, 21–26].

In the generic nature of the study, [19] makes routine data accessible at admission to anticipate organisational characteristics of psychiatric hospital care. In the study of 45,388 inpatients, the model performed well in predicting compulsory treatment and 1:1 observation but poorly in predicting short lengths of stay and non-response to treatment. Two years later, [20] conducted a study to predict the number of patients admitted to mental health facilities before and during the epidemic COVID -19. Their research may explain the fluctuations in hospital admissions in 2019 and 2020 due to the impact of the COVID-19 pandemic. Apart from this, previous researchers have focused on psychiatric disorders such as bipolar disorder, schizophrenia, anxiety, and depression in specific studies. Currently, only one study by [8] focuses on depression. The research was done at Kwekwe General Hospital on recently diagnosed instances of severe depression from January 2010 through December 2019 to predict cases from January 2020 through December 2021. The study offers policy recommendations to improve the treatment of major depression. Besides, [24] conducted a study on the prevalence of schizophrenia. The study suggests a significant increase in the prevalence of schizophrenia after 2016, with varying trends across months. Few other researchers have also conducted prognosis studies on schizophrenia, such as [21, 26].

In addition, in the study for the psychiatric unit, the researchers make predictions based on one type of data, either inpatient or outpatient, and some of the authors used both data. For example, [19, 20, 25] used to make predictions based on inpatient data, [23] made predictions based on outpatient data, while [21, 26] considered both inpatient and outpatient data. On the other hand, [8, 22, 24] used both data types but in aggregated form. Regarding the techniques used, researchers in this domain study mostly forecast based on time series, a succession of values with specific times and dates to record them [27]. Typically, data is gathered regularly. Forecasts in time series forecasting are projected using models based on previous time series data. In time series forecasting, trends are combined with the observed value. In most cases, trend and seasonality patterns are used [28].

Researchers have previously forecasted psychiatric illnesses using statistical and machine-learning approaches. The statistical method is an established method that has been utilised for quite some time and has a solid track record. Statistics are used to describe the data as well as to infer meaning from the data. In addition, inferential statistics are typically used to answer questions regarding the data, test hypotheses, quantify effects, and characterise correlations or relationships within the data. It is common practice to employ point estimates when assessing the importance of connections [29].

The ARIMA model is a commonly used time series model that employs frequency statistics for forecasting purposes. It is widely popular due to its statistical properties, such as the Box-Jenkins method for modelling, which is a well-known approach [30]. The Box-Jenkins approach does not presume the existence of any pattern in the data of the history series that will be used to make the forecast. In its place, it utilises an iterative three-step approach consisting of model recognition, estimation of parameters, and diagnostic inspection to choose the most effective model from a vast class of ARIMA models [31]. In addition to ARIMA, other statistical techniques such as SARIMA, Exponential Smoothing, Prophet, Logistic Regression, and hidden Markov model (HMM) were also used for time series forecasting. However, the statistical approach requires significant time-consuming pre-processing tasks [32].

Therefore, the other alternative used by previous researchers in this field was ML models. It is defined in the scientific world as a system's capacity to learn from experience. ML is now widely used in artificial intelligence-based intelligent systems. The purpose of machine learning is to simplify the process of constructing analytical models and solving related problems by learning from problem-specific training data [33]. Gradient boosting with trees (XGB), SVM, and elastic networks were compared to predict psychiatric hospital admissions [20]. In addition to these algorithms, Radial Basis Function (RBF) and Multi-Layer Perceptron Networks (MLP) have also been used by previous researchers. Although ML models are faster in training and prediction, their accuracy is not as high as statistical approaches [32]. Recently, DL approaches have shown great potential for accurate forecasting, especially when time series are involved, as they can capture non-linearity [34]. However, DL has not yet been explored in this prediction area for psychiatric units. Nonetheless, this research offers valuable insights into the potential of advanced DL approaches for hospital resource planning and management. These approaches can improve resource allocation and patient outcomes in mental health care.

### 2.2 Deep Learning Approaches

DL is an advanced version of a neural network with excellent learning capability. DL models have been found to outperform superficial ML models and conventional data analysis methods in many applications [33]. Throughout the last several decades, ML has undergone a great deal of development in terms of advanced learning algorithms and efficient pre-processing approaches.

DL typically has multiple hidden layers and is arranged in nested network architectures, allowing more complex relationships between predictors and output to be discovered [35]. The DL's multiple layers will enable it to process non-linear and highly dynamic data efficiently and effectively [36]. Moreover, advanced neurons are often found in them. In addition, DL has been used for time series data applications. Over time, various DL architectures have evolved [37]. Although practically every architecture may be utilised for any task, some do better with certain kinds of data, such as photos or time series. Predictions based on time series data often include a combination of trend-observing and regression-based procedures. Time series forecasting is often used for predicting the behaviour of a process or the financial markets [38]. Most architectural variations are neural units, the type of connection used and multiple layers. DL methods most used for time series include RNN, LSTM and GRU.

#### *2.2.1 Recurrent Neural Networks*

RNNs are primarily used with sequential data structures, for instance, natural language, time series, and event sequences. Sequential pattern learning can create a memory, which allows it to represent temporal relationships. This is possible because the design of sequential pattern learning systems incorporates internal feedback loops. RNNs are a crucial component of DL, which emphasizes solving time series issues. This is because RNNs have many benefits over sequence dependency [39]. RNN is referred to as recurrent because it executes the same task for each constituent of a sequence and bases its performance on previous computations. RNNs can also be viewed as having a memory in which information about prior calculations is preserved. In theory, RNNs can use information from any length sequence but only look back a few steps [40].

#### *2.2.2 Long Short-Term Memory*

The problem of an exploding or vanishing gradient in RNN inspired the development of the LSTM model to solve the problem of RNN [41]. LSTMs make use of gates as a means of controlling the flow of information. This enables them to handle disappearing gradients while maintaining the error flow [42]. Memory blocks that can recall data from earlier inputs are another property of LSTMs. This enables it to handle long sequences of dependencies more effectively [43]. Memory cells are an alternative to the standard LSTM unit, a more straightforward logic circuit implementation. The memory cell's many internal states send a continuous flow of mistakes. The memory cell makes up a linear core unit with a permanently established self-connection. The learning capabilities of the LSTM have significant practical and theoretical consequences in a wide variety of domains, which helps to distinguish this model as a ground-breaking innovation [44].

#### *2.2.3 Gated Recurrent Unit*

GRU falls under the category of RNN to solve the problem of vanishing and exploding gradients in traditional RNNs in long-term dependencies [45]. GRU combines the receiving and forgetting ports into a single updating port. Its architecture is more straightforward than LSTM as it has only two gates - a reset gate and an update gate. It lacks an output gate. Although it evolved from the LSTM unit, its calculation and application are much more straightforward. This method maintains the LSTM’s resilience to the gradient diminishing problem. Because of its more straightforward internal structure, it is less complicated to instruct and requires fewer computations to keep its internal states current. The update port governs how much information from prior states is carried into the present state. In contrast, the reset port determines whether data from past states should accompany the current state. On the other hand, in comparison to the LSTM, the GRU has a lower number of parameters, which has several benefits, the most notable of which is that it can generally be trained more quickly and with less effort [46].

## 3. METHODOLOGY

### 3.1 Data Collection

The dataset used in this study was from Hospital Universiti Sains Malaysia (HUSM), located at Kubang Kerian, Kelantan, Malaysia. The data is provided by the Department of Psychiatry at HUSM and covers five years, from January 2016 until December 2020. For inpatient data, the data was collected based on daily data, with a total of 1827 records. Besides, for the outpatient data, the data collected was based on monthly data, with 60 records. For the dataset, ethical approval was granted by Jawatankuasa Etika Penyelidikan Manusia Universiti Sains Malaysia (JEPeM-USM) on 1 April 2022.

### 3.2 Experimental Setup

The initial phase in this experiment is data preprocessing. Doing this allows the data to be used for the remainder of the process. The data collected from the HUSM is directly extracted from their electronic health record (HER); thus, they were not in good condition for the experiment. Thus, data cleaning is the preparation phase's initial stage, eliminating any outliers, missing values and duplicate records. The dataset was free of outliers and missing values, but three duplicate records were found. The duplicated records are removed from the dataset and maintain only the original record.

Next, the data is normalised using MinMaxScaler. Data normalisation is needed as the number of patients varies greatly, especially for outpatient visits. Thus, using data normalisation helps scale the number to a consistent scale, enhancing data quality. This process involves estimating the lowest and highest values and reducing the data's size from what it was initially to a new scale where all values are in the range of 0 to 1. For the model development, the data collection was partitioned into training and test dataset. The samples from the training comprised seventy per cent of the total, while the samples from the test set comprised the remaining thirty per cent. To carry out the final evaluation and assess the effectiveness of the training procedure, the test set was utilised. The following structure sets the model’s inputs: sample size, time phase, and number of features.

The univariate method only considers a single characteristic or independent variable, whereas the multivariate method considers at least two additional time series variables. This study considers univariate input variable. The data are reshaped using the Python function NumPy reshape [47]. The Python program ran on Jupyter Notebook and involved pandas, numpy, pyplot, Keras and TensorFlow libraries. No graphics processing unit (GPU) was used for this study. All three algorithms used in this study are supervised learning algorithms. In supervised learning, all models necessitate labelled data samples that consist of an input component ($x$) and an output component ($y$). The algorithms rely on these data samples to function properly. For this research, the input component ($x$) pertains to the explanatory variable values from the preceding day. Figures 1, 2 and 3 show the visualisation of the training and test data. The algorithms are then developed. All algorithms were developed based on default values without any parameter tuning. The details of the parameters are shown in Table 1.

### 3.3 Evaluation Metrics

Three performance measures are chosen based on the benchmark used in earlier forecasting studies to be employed in this study: the mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). The MSE is a metric used to examine the range of forecasting imprecision by measuring the mean squared disparity among the actual and anticipated values as shown in Equation (1). $j$ represents each individual observation ranging from 1 to $n$ (the total number of observations), $y_j$ is the actual observed value and $y_j'$ is the predicted/estimated value.

$$\text{MSE} = \sum_{j=1}^{n} \frac{\left(y_j - y_j'\right)^2}{n} \tag{1}$$

The residuals or standard deviation of the anticipated error are calculated using RMSE [48]. To be more specific, it illustrates how effectively the information is focused on the best-fit line. The formula for calculating the RMSE is defined in Equation (2), where it is essentially the square root of MSE.

Table 1. Details of parameters setting for the DL algorithms

| Parameters | Value |
|---|---|
| Learning rate | 0.01 |
| Epochs | 50 |
| Batch size | 64 |
| Dense | 1 |
| Hidden node | 100 |
| Dropout rate | 0.5 |

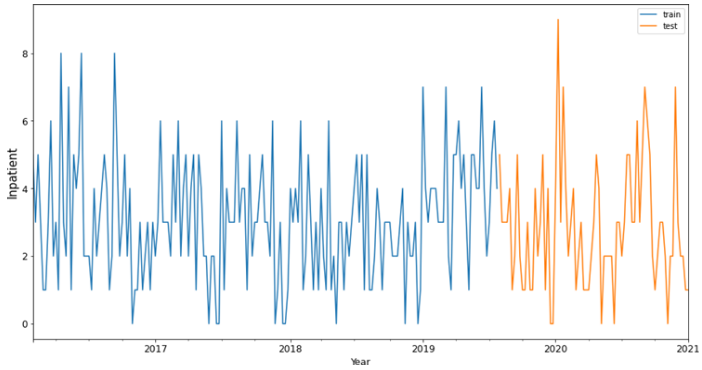

Figure 1. The visualization of the training and testing data for inpatient based on weekly data

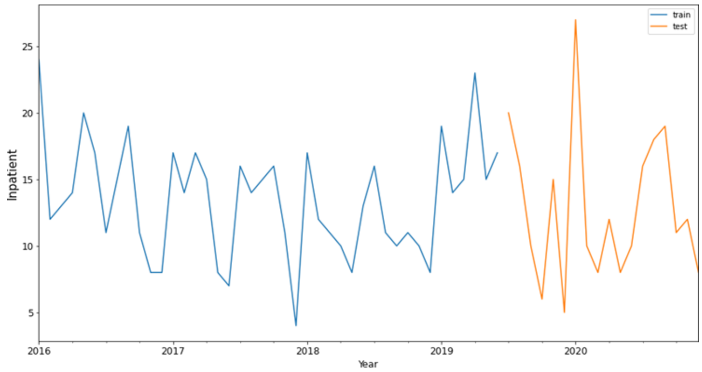

Figure 2. The visualization of the training and testing data for inpatient based on monthly data

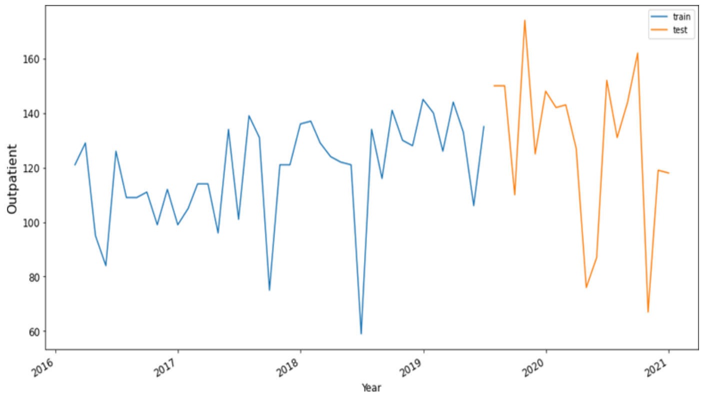

Figure 3. The visualization of the training and testing data for outpatient based on monthly data

$$\text{RMSE} = \sqrt{\sum_{j=1}^{n} \frac{(y_j - y_j')^2}{n}} \tag{2}$$

MAE is the average magnitude of mistakes in a series of predictions, regardless of error direction. It is the mean of the absolute differences between estimations and actual observations across the test sample [49].

$$\text{MAE} = \sum_{j=1}^{n} \frac{|y_j - y_j'|}{n} \tag{3}$$

## 4. RESULTS

This section discusses the results obtained from the experiments. Table 2 shows the comparison of the predictive accuracy between RNN, LSTM and GRU for inpatient forecasting. The evaluations were based on the evaluation metrics, MSE, RMSE and MAE. Lower values of the metrics indicate better algorithm performance. The inpatient forecast was divided into weekly and monthly for a complete analysis. In contrast, the outpatient forecast was based on the monthly forecast only, as the data is only available monthly. The algorithms performed almost similarly in inpatient forecasting, with MSE values between 3.37 and 3.45, RMSE between 1.84 and 1.86, and MAE between 1.84 and 1.86 for weekly forecasting, while for monthly inpatient forecasting ranging between 34.06 to 34.79 for MSE, 5.84 to 5.9 for RMSE and 5.84 to 5.9 for MAE. LSTM had the lowest error score for weekly inpatient prediction, while GRU had the lowest error yield for monthly inpatient prediction. The RNN method obtained the worst MSE and RMSE values for predicting inpatients. The three algorithms predicted inpatients with almost equal MSE and MAE values. Nevertheless, it was more difficult for the models to predict monthly inpatients than weekly inpatients, as shown by the higher RMSE values.

Table 3 tabulates the performance of the DL algorithms in predicting outpatients. The results show that the LSTM models predict outpatients well, with lower MSE, RMSE and MAE values than RNN and GRU models. This means the LSTM model is better than the RNN and GRU models to capture temporal relationships and patterns in monthly outpatient hospitalisation data. All three models had lower MSE, RMSE and MAE values for monthly prediction for inpatients than for outpatients, suggesting that they predict inpatient hospitalisations more accurately. For weekly inpatient and monthly outpatient prediction, the LSTM algorithm achieved better performance than other algorithms. It is important to note that the differences in performance between the three algorithms are modest.

The effectiveness of the LSTM algorithm is explored using a loss function to examine the top-performing algorithm thoroughly. Evaluating the loss function can provide insight into the algorithm's performance. The loss function of LSTM models evaluates the accuracy of the model's predictions based on input during training. It calculates the variance between the predicted output and the actual output. The model improves its predictive abilities by minimizing the loss function. The MSE loss function of the LSTM is shown in Figure 4. It shows the small loss function of the LSTM and the decreasing error during training. The LSTM model also makes more accurate predictions and better captures the patterns and correlations of the input data. LSTMs can predict monthly data, as shown in Figure 5. This shows that the LSTM can predict a trend. Further modifications should be made to improve LSTM results.

Comparisons are also made between the results of the DL algorithm and those of previous methods. The comparative results are presented in Tables 4 and Table 5. For the inpatient weekly projection, LSTM had the lowest MSE of 3.37 and RMSE of 1.84. The SVM, Prophet, SARIMA and ARIMA models were all studied in the previous paper. The ARIMA model performed the best in monthly prediction for inpatients with a score of 30.48 MSE. It had the lowest RMSE and MAE errors of all the models, at 5.52 and 2.14 respectively. The LSTM model also performed better than all other models for outpatient forecasts with a score of 1024.91 MSE, 32.01 for both RMSE and MAE. In summary, LSTM is the best DL algorithm for weekly inpatient forecasts and outpatient forecasts, while ARIMA is best for monthly inpatient forecasts.

Table 2. Performance of deep learning algorithms for inpatient forecasting

| | Weekly | | | Monthly | | |
|---|---|---|---|---|---|---|
| Algorithms | MSE | RMSE | MAE | MSE | RMSE | MAE |
| RNN | 3.45 | 1.86 | 1.86 | 34.79 | 5.90 | 5.90 |
| LSTM | **3.37** | **1.84** | **1.84** | 34.69 | 5.89 | 5.89 |
| GRU | 3.45 | 1.86 | 1.86 | **34.06** | **5.84** | **5.84** |

Table 3. Performance of deep learning algorithms for outpatient forecasting

| | Monthly | | |
|---|---|---|---|
| Algorithms | MSE | RMSE | MAE |
| RNN | 1047.05 | 32.36 | 32.36 |
| **LSTM** | **1024.91** | **32.01** | **32.01** |
| GRU | 1033.41 | 32.15 | 32.15 |

Table 4. Comparison of the performance of deep learning algorithms with previous studies for inpatient forecasting

| | Weekly | | | Monthly | | |
|---|---|---|---|---|---|---|
| Algorithms | MSE | RMSE | MAE | MSE | RMSE | MAE |
| RNN | 3.45 | 1.86 | 1.86 | 34.79 | 5.90 | 5.90 |
| LSTM | **3.37** | **1.84** | 1.84 | 34.69 | 5.89 | 5.89 |
| GRU | 3.45 | 1.86 | 1.86 | 34.06 | 5.84 | 5.84 |
| SARIMA [8] | 6.08 | 2.47 | 1.46 | 46.65 | 6.83 | 2.39 |
| SVM [20] | 3.64 | 1.91 | 1.50 | 36.52 | 6.04 | 4.76 |
| Prophet [21] | 3.56 | 1.89 | 1.53 | 38.59 | 6.21 | 4.92 |
| ARIMA [22] | 4.47 | 2.11 | **1.34** | **30.48** | **5.52** | **2.14** |

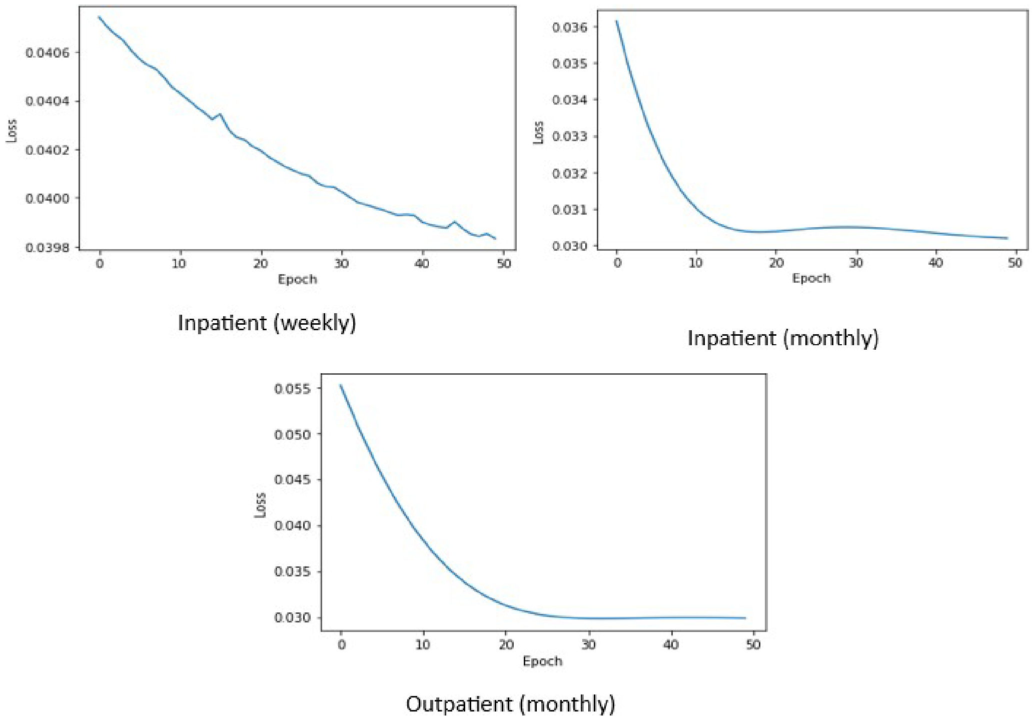


Figure 4. The loss function of LSTM

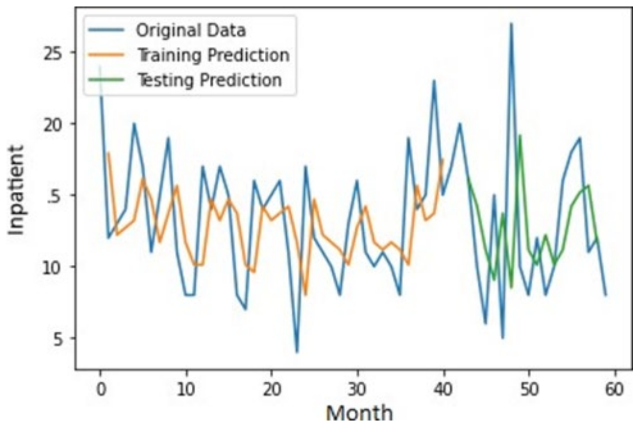


Figure 5. Example of LSTM plot in monthly forecasting

Table 5. Comparison of the performance of deep learning algorithms with previous studies for outpatient forecasting

| | Monthly | | |
|---|---|---|---|
| Algorithms | MSE | RMSE | MAE |
| RNN | 1047.05 | 32.36 | 32.36 |
| **LSTM** | **1024.91** | **32.01** | 32.01 |
| GRU | 1033.41 | 32.15 | 32.15 |
| SVM [20] | 1104.12 | 33.23 | 26.19 |
| Prophet [21] | 1132.55 | 33.65 | 27.55 |

## 5. DISCUSSION

Based on the result presented in Section 4, LSTM performed better than RNN and GRU in forecasting inpatient and outpatient visits. The prediction is better in this study with LSTM because it matches the nature of non-linear data in this study with a combination of the ability to detect long-term data dependencies. After all, each memory cell can retain the information longer. Unfortunately, compared to LSTM, RNN has shorter memory and can only retain recent information [50], which makes it ideal for temporary dependencies. Besides, for GRU, even though it can also retain more extended information, it is not as efficient as LSTM in capturing long-term term dependencies [51].

All three algorithms predict monthly health care worse than weekly forecasting, indicating that monthly prediction requires more data. In addition, with existing statistics on the data, it fluctuates from month to month, making prediction more challenging. Performance variables from a particular dataset may not apply to other datasets or applications. Hyperparameters, dataset quantity and quality, and pre-processing can affect the algorithm's performance [52]. Therefore, it is crucial to analyse the different DL algorithms and select the optimal algorithm for the task.

There are several limitations to this study. First, the data obtained from the HUSM was limited, especially in terms of outpatient data. The outpatient data found in the electronic health record was based on monthly data only. Therefore, the outpatient prediction is limited to monthly data. Also, this study was conducted only in the psychiatric department of one hospital. It would be good to have more data, e.g., from a few psychiatric departments, to make more reliable predictions. Another limitation is that some patients have more than one psychiatric illness other than depressive disorders, for example, a patient who have anxiety and depressive disorders; thus, this study might also contain some numbers of patients who have a psychiatric illness other than depressive disorders.

Some recommendations have been elaborated for future researchers to apply DL algorithms in this field. First, researchers should focus on the algorithm's hyperparameters DL as they strongly influence its performance. The number of hidden layers, the number of neurons per layer, the learning rate and the activation functions can be changed to improve the method. Selecting these hyperparameters using optimisation methods such as grid search, random search, or metaheuristic optimisation algorithms is crucial for optimal performance. The amount and quality of data collection used to train and evaluate the DL algorithms can also influence performance. More extensive and better data sets usually lead to better results. When the model is overfitting the training data, it may produce less accurate results when applied to the test data. The preparation of the data for analysis can also affect performance. Scaling or normalising the data, dealing with missing values and classifying categorical variables can all affect the technique's performance. Attention should be paid to all these tasks before model development.

This study also suggests resolutions for hospital management. Based on the forecasting, the hospital management could obtain information about the future trend of inpatient and outpatient visits. Thus, based on the graph, they can identify when the inpatient and outpatient visits peaked, which means the potential for overcrowding. Thus, they can make early preparations by allocating resources more efficiently by ensuring staff and beds are available when the demand is expected to be high. With the limited capacity, especially for the inpatient, which can accommodate 32 patients at a time, the management must have a proper plan to have a good patient outcome and avoid overcrowding.

## 6. CONCLUSION

The predictive power of three DL models (RNN, LSTM, and GRU) was compared for forecasting inpatient and outpatient visits. The LSTM model outperformed the other two models in predicting health visits from routine data. The precision of these estimates enables better health outcomes and more efficient use of resources. The study highlights the importance of the opportunity for DL to improve patient outcomes and resource allocation in mental health care. The results of this study add to the growing body of work on the promising possibilities of DL in healthcare and highlight the need for accurate prediction to improve patient outcomes and maximise resource utilisation.

## ACKNOWLEDGEMENT AND FUNDING

The authors would like to extend their gratitude to the Ministry of Higher Education of Malaysia for providing funds for this research. Project Code: FRGS/1/2019/ICT02/USM/02/3 from the Fundamental Research Grants Scheme supported this study.

## DECLARATION OF CONFLICTING INTERESTS

The authors declare no potential conflicts of interest with respect to the research and publication of this article.

## REFERENCES

[1] P. Junfeng, C. Chen, M. Zhou, X. Xie, Y. Zhou and C. -H. Luo, Peak outpatient and emergency department visit forecasting for patients with chronic respiratory diseases using machine learning methods: retrospective cohort study, *JMIR Medical Informatics*, 8, 2020, e13075.

[2] B. Morel, G. Bouleux, A. Viallon, M. Maignan, L. Provoost, J. -C. Bernadac, S. Devidal, S. Pillet, A. Cantais and O. Mory, Evaluating the increased burden of cardiorespiratory illness visits to adult emergency departments during flu and bronchiolitis outbreaks in the pediatric population: retrospective multicentric time series analysis, *JMIR Public Health and Surveillance*, 8, 2022, e25532.

[3] S. Lakshmy, Effective demand forecasting in health supply chains: emerging trend, enablers, and blockers, *Logistics*, 5, 2021, 1-21.

[4] N. Zhao, K. Charland, M. Carabali, E. O. Nsoesie, M. M. Giroux, E. Rees, M. Yuan, C. G. Balaguera, G. J. Ramirez and K. Zinszer, Machine learning and dengue forecasting: comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia, *PLOS Neglected Tropical Diseases*, 14, 2020, e0008056.

[5] X. Liu, F. Gu, Z. Bai, Q. Huang and G. Ma, Forecasting of daily outpatient visits based on genetic programming, *Iranian Journal of Public Health*, 51, 2022, 1313-1322.

[6] X. Han, F. Jiang, Y. Tang, J. Needleman, M. Guo, Y. Chen, H. Zhou and Y. Liu, Factors associated with 30-day and 1-year readmission among psychiatric inpatients in Beijing China: a retrospective, medical record-based analysis, *BMC Psychiatry*, 20, 2020, 1-12.

[7] M. A. Khafaie, M. Sayyah and F. Rahim, Extreme pollution, climate change, and depression, *Environmental Science and Pollution Research*, 26, 2019, 22103-22105.

[8] D. S. P. Nyoni and M. T. Nyoni, Modeling and forecasting major depression cases at Kwekwe General Hospital, Zimbabwe empirical evidence from a box jenkins catch all model, *JournalNX*, 6, 2020, 134-141.

[9] A. Nori-Sarma, Amruta, S. Sun, Y. Sun, K. R. Spangler, R. Oblath, S. Galea, J. L. Gradus and G. A. Wellenius, Association between ambient heat and risk of emergency department visits for mental health among US adults, 2010 to 2019, *JAMA Psychiatry,* 79, 2022, 341-349.

[10] L. Jakub, K. Piotrowicz, P. P. Hughes and M. Makara-Studzińska, Weather and aggressive behavior among patients in psychiatric hospitals—an exploratory study, *International Journal of Environmental Research and Public Health,* 17, 2020, 9121-9233.

[11] J. Christian, P. Zschech and K. Heinrich, Machine learning and deep learning, *Electronic Markets*, 31, 2021, 685-695.

[12] A. Laith, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, *Journal of Big Data*, 8, 2021, 1-74.

[13] M. H. Alsharif, Y. H. Alsharif, K. O. M. Yahya, O. A. Alomari, M. A. M. Albreem and Abu Jahid, Deep learning applications to combat the dissemination of COVID-19 disease: a review, *European Review for Medical and Pharmacological Sciences*, 24, 2020, 11455-11460.

[14] H. Bi, L. Lu and Y. Meng, Hierarchical attention network for multivariate time series long-term forecasting, *Applied Intelligence*, 53, 2023, 5060-5071.

[15] Y. Yang, C. Fan and H. Xiong, A novel general-purpose hybrid model for time series forecasting, *Applied Intelligence*, 52, 2022, 2212-2223.

[16] E. Jangam, A. A. D. Barreto and C. S. R. Annavarapu, Automatic detection of COVID-19 from chest CT scan and chest X-Rays images using deep learning, transfer learning and stacking, *Applied Intelligence*, 52, 2022, 1-17.

[17] S. M. Cihan, C. Paoli and Ö. D. Incel, Time series forecasting on solar irradiation using deep learning, *In 10th International Conference on Electrical and Electronics Engineering (ELECO)*, Turkey, 2017, 151-155.

[18] A. Zeroual, F. Harrou, A. Dairi and Y. Sun, Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study, *Chaos, Solitons & Fractals,* 140, 2020, 110121.

[19] J. Wolff, A. Gary, D. Jung, C. Normann, K. Kaier, H. Binder, K. Domschke, A. Klimke and M. Franz, Predicting patient outcomes in psychiatric hospitals with routine data: a machine learning approach, *BMC Medical Informatics and Decision Making*, 20, 2020, 1-9.

[20] J. Wolff, A. Klimke, M. Marschollek and T. Kacprowski, Forecasting admissions in psychiatric hospitals before and during Covid-19: a retrospective study with routine data, *Scientific Reports*, 12, 2022, 15912.

[21] S. Ryu, H. J. Nam, J. -M. Kim and S. -W. Kim, Current and future trends in hospital utilization of patients with schizophrenia in Korea: a time series analysis using national health insurance data, *Psychiatry Investigation*, 18, 2021, 795-800.

[22] S. P. Humaira, I. Nursuprianah and D. Darwan, Forecasting of the number of schizophrenia disorder by using the Box-Jenkins of time series analysis, *Journal of Robotics and Control,* 1, 2020, 213-219.

[23] S. Vatinee, J. Thongkam and J. Leejongpermpoon, Time series forecasting in anxiety disorders of outpatient visits using data mining, *Asia-Pacific Journal of Science and Technology*, 20, 2015, 241-253.

[24] A. Payam, A. Ghaleiha, E. Zarean, M. Sadeghifar, M. E. Ghaffari, Z. Taslimi and S. Yazdi-Ravandi, Modelling the frequency of depression using holt-winters exponential smoothing method, *Journal of Clinical & Diagnostic Research*, 12, 2018, 24-27.

[25] B. Alafchi, S. Y.- Ravandi, R. N.- Vosough, A. Ghaleiha and M. Sadeghifar, Forecasting new cases of bipolar disorder using poisson Hidden Markov model, *International Clinical Neuroscience Journal*, 5, 2018, 7-10.

[26] S. Ryu, H. J. Nam, S.-H. Baek, M. Jhon, J.-M. Kim and S.-W. Kim, Decline in hospital visits by patients with schizophrenia early in the COVID-19 outbreak in Korea, *Clinical Psychopharmacology and Neuroscience*, 20, 2022, 185-189.

[27] Z. Li, X. Zhang and Z. Dong, TSF-transformer: a time series forecasting model for exhaust gas emission using transformer, *Applied Intelligence*, 53, 2023, 17211-17225.

[28] I. Lohrasbinasab, A. Shahraki, A. Taherkordi and A. D. Jurcut, From statistical-to machine learning-based network traffic prediction, *Transactions on Emerging Telecommunications Technologies*, 33, 2022, e4394.

[29] M. D. Prasanna, O. Zimba and A. Y. Gasparyan, Statistical data presentation: a primer for rheumatology researchers, *Rheumatology International*, 41, 2021, 43-55.

[30] A. Zheng, Q. Fang, Y. Zhu, C. Jiang, F. Jin and Xin Wang, An application of ARIMA model for predicting total health expenditure in China from 1978-2022, *Journal of Global Health*, 10, 2020, 10803-10810.

[31] J. Mohamed, Time series modeling and forecasting of Somaliland consumer price index: a comparison of ARIMA and regression with ARIMA errors, *American Journal of Theoretical and Applied Statistics*, 9, 2020, 143-53.

[32] Y. Ning, H. Kazemi and P. Tahmasebi, A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet, *Computers & Geosciences*, 164, 2022, 105126.

[33] C. Janiesch, P. Zschech and K. Heinrich, Machine learning and deep learning, *Electronic Markets*, 31, 2021, 685-695.

[34] S. Masum, Y. Liu and J. Chiverton, Multi-step time series forecasting of electric load using machine learning models, *Artificial Intelligence and Soft Computing: 17th International Conference, ICAISC 2018,* Poland, 2018, 148-159.
[35] B. C. Healy, Machine and deep learning in ms research are just powerful statistics– no, *Multiple Sclerosis Journal*, 27, 2021, 663–664.
[36] Z. Shen, Y. Zhang, J. Lu, J. Xu and G. Xiao, A novel time series forecasting model with deep learning, *Neurocomputing*, 396, 2020, 302-313.
[37] S. Leijnen and F. V. Veen, The neural network zoo, *Multidisciplinary Digital Publishing Institute Proceedings*, USA, 2020, 1-6.
[38] K. Heinrich, P. Zschech, C. Janiesch and M. Bonin, Process data properties matter: Introducing gated convolutional neural networks (GCNN) and key-value-predict attention networks (KVP) for next event prediction with deep learning, *Decision Support Systems,* 143, 2021, 113494.
[39] X. Li, X. Ma, F. Xiao, F. Wang and S. Zhang, Application of gated recurrent unit (GRU) neural network for smart batch production prediction, *Energies,* 13, 2020, 6121.
[40] C. Fan, J. Wang, W. Gang and S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, *Applied Energy*, 236, 2019, 700-710.
[41] O. Ozyegen, I. Ilic and M. Cevik, Evaluation of interpretability methods for multivariate time series forecasting, *Applied Intelligence*, 52, 2022, 1-17.
[42] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita and U. R. Acharya, Automated detection of atrial fibrillation using long short-term memory network with RR interval signals, *Computers in Biology and Medicine*, 102, 2018, 327-335.
[43] X. Yu and D. Li, Important trading point prediction using a hybrid convolutional recurrent neural network, *Applied Sciences*, 11, 2021, 3984.
[44] Z. Karevan and J. A. K. Suykens, Transductive LSTM for time-series prediction: An application to weather forecasting, *Neural Networks,* 125, 2020, 1-9.
[45] G. Shen, Q. Tan, H. Zhang, P. Zeng and J. Xu, Deep learning with gated recurrent unit networks for financial sequence predictions, *Procedia Computer Science*, 131, 2018, 895-903.
[46] N. A. P. Rostam, N. H. A. H. Malim, R. Abdullah, A. L. Ahmad, B. S. Ooi and D. J. C. Chan, A complete proposed framework for coastal water quality monitoring system with algae predictive model, *IEEE Access,* 9, 2021, 108249-108265.
[47] B. C. Mateus, M. Mendes, J. T. Farinha, R. Assis and A. M. Cardoso, Comparing LSTM and GRU models to predict the condition of a pulp paper press, *Energies*, 14, 2021, 6958.
[48] A. Arundarasi, V. S. Sahithi, C. Gupta, M. Yadav, S. Ahirrao, K. Kotecha, M. Gaikwad, A. Abraham, N. Ahmed and S. M. Alhammad, Detecting extremism on twitter during US capitol riot using deep learning techniques. *IEEE Access*, 10 2022, 133052-133077.
[49] B. Zhu, Y. Feng, D. Gong, S. Jiang, L. Zhao and N. Cui, Hybrid particle swarm optimization with extreme learning machine for daily reference evapotranspiration prediction from limited climatic data, *Computers and Electronics in Agriculture*, 173, 2020, 105430-105442.
[50] A. Sagheer and M. Kotb, Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems, *Scientific Reports*, 9, 2019, 19038.
[51] K. E. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji and T. M. Brenza, Comparative analysis of gated recurrent units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends, *Alexandria Engineering Journal*, 61, 2022, 7585-7603.
[52] B. Akay, D. Karaboga and R. Akay, A comprehensive survey on optimizing deep learning models by metaheuristics, *Artificial Intelligence Review*, 55, 2022, 1-66.