DOI: 10.24996/ijs.2025.66.1.30

ISSN: 0067-2904

Business Intelligence Approach-Based Hybrid Deep Learning Model for QS World University Ranking

Asaad R. Kareem, Hasanen S. Abdullah

Computer Science Department, the University of Technology, Baghdad, Iraq

Received: 20/6/2023 Accepted: 21/12/2023 Published: 30/1/2025

Abstract

The QS World University Rankings have become the most widely used basis for comparing universities across the world. Selecting a university is a difficult decision, and this may be particularly true for international students. The rankings of universities have been designed to help students make more informed choices. Informed decision institutions face challenges in analyzing huge data volumes and ensuring generalizability. This study presents a hybrid deep learning model (CNN-LSTM) that addresses these issues, minimizing data requirements and ensuring compatibility. The proposed hybrid deep learning model was compared with five other machine learning models. It demonstrated the highest efficiency, with a mean squared error of 0.0144 and an accuracy of 98.5%.

Keywords: business intelligence BI, Convolutional Neural Networks CNN; Long Short-Term Memory LSTM, Hybrid deep learning.

نهج ذكاء الاعمال القائم على نموذج التعليم العميق الهجين لتصنيف الجامعات العالمية QS

أسعد رحيم كريم ،حسنين سمير عبدالله

قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق

الخلاصة

أصبح تصنيف الجامعات العالمية QS هو الأساس الأكثر استعمالاً لمقارنة الجامعات في جميع أنحاء العالم ، ويعد اختيار الجامعة قرارًا صعبًا، وقد يكون هذا صحيحًا بشكل خاص بالنسبة للطلبة الدوليين. تم تصميم تصنيفات الجامعات لمساعدة الطلبة على اتخاذ قرارات أكثر استنارة. تواجه المؤسسات تحديات في اتخاذ القرارات في تحليل أحجام البيانات الضخمة وقد يكون النموذج محدودًا في قابليته للتعميم على أنواع مختلفة من البيانات. هذه الدراسة تقدم نموذجًا هجينًا للتعلم العميق (CNN–LSTM) مطبقًا على مجموعة بيانات تصنيفات الجامعات العالمية QS يعالج هذه المشكلات ، ويقلل من متطلبات البيانات ويضمن التوافق. تمت مقارنة نموذج التعلم العميق الهجين المقترح مع خمسة نماذج أخرى للتعلم الآلي. وأظهرت أعلى كفاءة بمتوسط خطأ مربع قدره 0.0144 ودقة قدرها 98.5%.

*Email: hasanen.s.abdullah@uotechnology.edu.iq

1- Introduction

Deep learning (DL) and business intelligence (BI) are two ideas that have recently attracted attention in the world of information technology. Deep learning (DL) is a subset of machine learning (ML) that employs neural networks to overcome complex problems; BI is a data-driven approach that makes use of tools to assist in decision-making [1]. DL eliminates the requirement for feature extraction steps by using NN models to automatically extract features from input data. This approach is an excellent option since it effectively categorizes huge and complicated data and does not need preprocessing procedures for feature description acquisition [2]. Traditional BI tools have limitations in terms of data diversity and capability, as well as the disability to learn, which means that the combination of BI and DL may transform how organizations evaluate data. Here are some potential applications that use the business intelligence approach based on deep learning: (I) Fraud Detection: DL models can be used to detect and prevent fraudulent activities. These algorithms can recognize the patterns and irregularities linked to fraudulent behavior. (II) Natural Language Processing (NLP) for Text Analysis: DL-based NLP techniques enable businesses to extract valuable insights from unstructured text data. (III) Image and Video Analysis: Deep learning models excel at image and video analysis tasks [3], [4].

It is used to train the hybrid deep learning model as well as five machine learning models: linear regression, decision tree regression, stochastic gradient descent regression, gradient boosting regression, and Bayesian ridge regression [5]. The training set data is first processed. The trained models are then used to predict the test set data after it has been processed, and the real value is compared with the predicted value, as described in more detail in the results section. Among the six prediction models, the proposed hybrid deep learning model achieved the lowest mean squared error of the real value and the predicted value, and it had the highest accuracy.

2. Related works

In 2020, Hind et al. [6] employed the entire business intelligence solutions cycle, including planning, design, and development. The proposed BI solution was evaluated through several iteration tests. The research aimed to discuss the role and impact of business intelligence (BI) in optimizing business decisions and strategies. The suggested BI solution was evaluated through multiple iteration tests, and the results of these iterations were presented. However, the specific findings or outcomes of these tests are not mentioned.

In 2022, Morteza et al. [7] have used two approaches. Machine learning involves algorithms that enable computers to learn and make decisions without explicit programming. Business intelligence involves technologies and applications used to collect, integrate, analyze, and present data for making strategic decisions and improving future results. The two approaches highlight the advantages of combining machine learning with business intelligence, such as improved operational processes and real-time data analysis. It also acknowledges some drawbacks, including cost, complexity, and time. The research does not explicitly mention specific results.

In 2021, Zhi-Xiong et al. [8] employed a descriptive survey method, which included a 43-item questionnaire and confirmatory factor analysis for data collection and validity analysis. The research aimed to investigate the effects of business intelligence on networked learning, innovation, and financial performance in startup companies. The results showed that business intelligence had no effect on networked learning but increased innovation by 0.99 and financial performance by 0.311. The intelligence of startups positively influenced network learning (0.537), which, in turn, enhanced innovativeness (0.632) and financial performance (0.397).

In 2022, Tarek et al. [9] employed machine learning and deep learning techniques, along with natural language processing tools such as stemming, normalization, and stop word removal. The research aimed to address the challenge of detecting customer satisfaction tones from big data collected from online businesses on social media platforms like Facebook and Twitter. The Random Forest classifier achieved the highest F1-measure of 99.1%, while the Support Vector Machine performed best without preprocessing techniques with an F1-measure of 93.4%. The Deep Neural Networks (DNN) algorithm demonstrated superior performance in the results.

In 2020, Daniela et al. [10] used Long Short-Term Memory (LSTM) for sentiment analysis, specifically to identify and categorize text based on conveyed sentiments and rate them on a scale from -100 to 100. The ultimate goal was to use business intelligence tools to evaluate products or services by analyzing notes, reviews, tweets, and feedback and categorizing them based on sentiments. The research does not provide specific results. However, it mentions that the newly suggested activation function for LSTM produced better results compared to existing artificial neural network (ANN) approaches.

In 2022, Anes et al. [11] employed object classification, descriptive analysis, and pattern recognition methods to investigate the correlation between decision-making and business intelligence in the field of management accounting. They also utilized a quantitative approach and acquired data from the SQL database of the organizations. The initial findings, based on numerical analysis, indicated a 10% increase in revenue and total revenue. These results suggested the potential impact of business intelligence on financial performance.

In 2022, Mahdi et al. [12] A model-driven approach has been proposed for automatically generating personalized Business Intelligence (BI) chatbots tailored to organizational needs. The approach consists of a modeling component for designing business-specific chatbots and an automation component for generating the chatbot code. The goal is to provide a solution that simplifies the process of creating personalized BI chatbots. The conducted case study demonstrates encouraging results in the development of interactive BI chatbots that effectively fulfill diverse organizational requirements.

In 2021, Zeel et al. [13] used Natural Language Processing (NLP) techniques, specifically DL-BERT and LSTM, and ML algorithms such as Logistic Regression, Stochastic Gradient Descent, Decision Tree, Multinomial Naive Bayes, and SVM. Business intelligence analysis is performed using Microsoft Power BI. This study allows companies to enhance efficiency and improve customer satisfaction. There is no discussion of the exact results or their importance.

In 2020, Ilayaraja [14] discusses the application of business intelligence (BI) in both business and higher education contexts. It proposes an integrated taxonomy derived from referenced maturity models, introduces the concept of applying BI, and provides guiding principles for fostering a data-centric culture. The goal is to understand what is happening, why it is happening, and how to address it through BI approaches. There is no discussion of the results or their importance.

3. Theoretical Background

3.1 Dataset

The dataset utilized in this article, sourced from the Kaggle datasets [15], is based on the QS World University Rankings. This dataset represents the annual publication of global university rankings conducted by Quacquarelli Symonds (QS). The International Ranking Expert Group (IREG) has approved the QS ranking, which is one of the top three most widely read university rankings in the world. QS is a British think-tank company with expertise in analyzing higher education institutions worldwide. Their ranking methodology relies on six factors for the ranking process [16]. The scores can indicate how well or poorly universities have performed compared to the previous year.

3.2 Pre- processing

- **Missing values**

Missing values are defined as the data value that is not stored for a variable in the observation of interest [17], [18]. The problem of missing data is relatively common in almost all research and can significantly affect the conclusions that can be drawn from the data [19], [20]. By applying this formula, the values in each column will be restricted to a range between 0 and 1.

$$\mathrm{m} = \frac{s}{n} \qquad \text{Eq. 1}$$

Where:
m: Mean
s: Sum of all data points
n: Number of data points

- **Ordinal encoder**

Ordinal encoding is a machine-learning technique that transforms categorical data into numerical data. It assigns a unique integer value to each category in the data. This integer value can then be used to encode categorical variables as input to neural networks. The advantage of employing ordinal encoding is its ability to maintain the order of categories [21].

- **Normalization**

Min-max scaling is a data transformation technique that resembles z-score normalization as it replaces each value in a column with a new value using a specific formula. [22]. The formula for min-max scaling is as follows:

$$\mathrm{m} = \frac{(\mathrm{x} - \mathrm{xmin})}{(\mathrm{xmax} - \mathrm{xmin})} \qquad \text{Eq. 2}$$

Where: m is our new value
x is the original cell value
xmin is the minimum value of the column
xmax is the maximum value of the column

Applying this formula ensures that the values of each column will be transformed to fall within the range of 0 to 1.

3.3 Convolutional Neural Network

The Convolutional Neural Network (CNN) stands as a leading deep learning model for diverse application-solving purposes [23]. A CNN model is composed of four different layers: convolution layers, activation layers, pooling layers, and fully connected layers. These layers work together to extract features from the input data and classify them into different categories. [24], [25]. The CNN architecture is made up of convolution layers that remove spatial features, activation layers that improve learning through non-linearity, pooling layers that reduce the number of dimensions, and fully connected layers that combine global and local features. Figure (1) depicts the CNN structure [26]. The convolution layer, connecting input x_i^k and output y_i^k, performs matrix multiplication between the previous layer's input and a filter bank w_{ij}^k, it also includes the addition of a regularization term (bias) b_j^k. Equation (3) can be used to compute the convolution layer [27]:

$$y_i^k = \sum_i \left(x_i^k * w_{ij}^k\right) + b_j^k \qquad \text{Eq. 3}$$

In the nonlinear layer, a leaky rectified linear unit (LeakyRelu) is utilized as the activation function. LeakyRelu differs from Relu by preventing the exclusion of negative inputs, thus addressing the dying Relu issue. Mathematically, LeakyRelu can be expressed as shown in Eq. 4.

$$f(x)LeakyRelu = \begin{Bmatrix} x, if\ x > 0 \\ mx, x \le 0 \end{Bmatrix} \qquad \text{Eq. 4}$$

The leak factor, denoted by m, is a small value often set to 0.001, introducing a slight negative slope for the activation function. This allows a small flow of information even when the input is negative, aiding in the learning process. In the maximum pooling layer, the highest value within a given matrix is selected (with non-overlapping features). The fully connected layer can incorporate a nonlinear activation function, like LeakyRelu. In classification tasks, the output layer is connected with a SoftMax activation function, while in regression tasks, a linear regression function is utilized. Equation (5) can be used to compute both the SoftMax function and the linear regression function.

$$\sigma\,(\mathrm{z})\mathrm{i} = \frac{e^{zi}}{\sum_{j=1}^{K} e^{zj}}\ \ for\ i = 1, \dots\dots\dots, k\ and\ z = (z1, \dots\dots\dots.zk) \in R^K \qquad \text{Eq. 5}$$

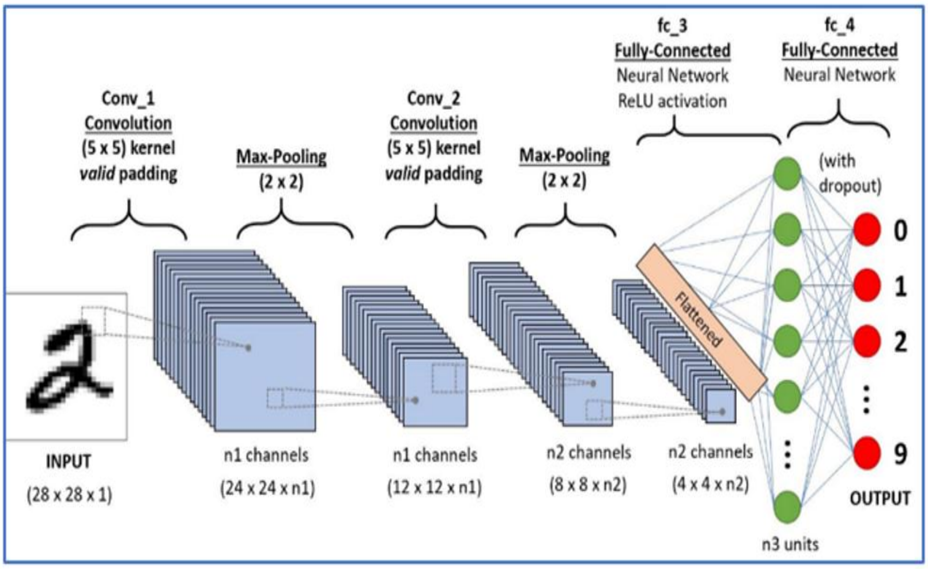


Figure 1: CNN Model Representation [29]

3.4 Long Short-Term Memory

Recurrent neural networks (RNNs) have a structure called long-short-term memory (LSTM). It adds memory cells to a hidden layer to control the memory information of time series data [28]. Information is transmitted among different cells of the hidden layer through several controllable gates (forget gate, input gate, output gate). This allows the memory and forgetting extent of the previous and current information to be controlled. Compared with traditional RNN, LSTM has a long-term memory function and avoids the gradient disappearance problem. The LSTM incorporates two gates that regulate the memory cell's state. The forget gate quantifies the extent to which memory from the previous time step is retained, while the input gate determines the amount of current input that is stored in the cell state. Additionally, these gates govern the proportion of historical information and current stimulus that are combined. Finally, the LSTM network includes an output gate that governs the quantity of information that is emitted regarding the cell status [29]. The representation of the LSTM structure is illustrated in Figure 2.

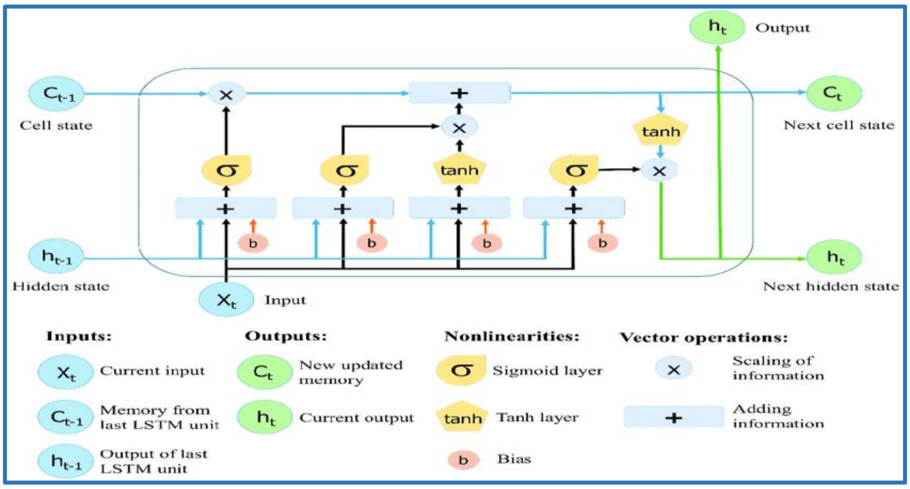


Figure 2: The Model Structure of LSTM [31]

σ is the sigmoid function shown in equation (13) in Figure 2. Its output is a value between 0 and 1, where 0 means let nothing pass and 1 means let everything pass.

$$f(t) = \sigma(wf\,.[ht-1, xt\,] + bf) \qquad \text{Eq. 6}$$

$$I(t) = \sigma(wf\,.[ht-1, xt\,] + bi) \qquad \text{Eq. 7}$$

$$\tilde{c}(t) = tanh(wc\,.\;[ht-1, xt\,] + bc) \qquad \text{Eq. 8}$$

$$c(t) = fi * ct - 1 + It * \tilde{c}t \qquad \text{Eq. 9}$$

$$o(t) = \sigma(wo\,.[ht-1, xt\,] + bo) \qquad \text{Eq. 10}$$

$$h(t) = ot * \tanh(c1) \qquad \text{Eq. 11}$$

$$sigmoid\;(x) = \frac{1}{1 + e^{-x}} \qquad \text{Eq. 12}$$

$$tanh\;(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad \text{Eq. 13}$$

The hyperbolic tangent function in Equation (12) is used to overcome the problem of gradient disappearance [30], [31]. Equations 6–13 describe the input and output of the network structure of the LSTM in Figure 2.

4 Methodology

Deep learning has become a powerful approach to addressing the challenges of managing massive amounts of data within strict time and cost constraints. By leveraging sophisticated algorithms and massive amounts of data, deep learning can efficiently analyze and process complex information, reducing the need for manual processing and improving the speed and accuracy of decision-making. This has the potential to revolutionize various sectors and drive significant improvements and increases in production, profitability, and efficiency. [32], [33], [34]. The most commonly used deep learning algorithms are convolutional neural networks (CNN) and long-short-term memory (LSTM). The advantage of CNN is feature extraction [35] and [36], while the LSTM component takes the spatially encoded features from the CNN and processes them sequentially. To make full use of their advantages, we combine these two models to obtain a new effective model (hybrid CNN-LSTM), which will be presented as follows: The structure of the hybrid deep learning model is shown in Figure 3.

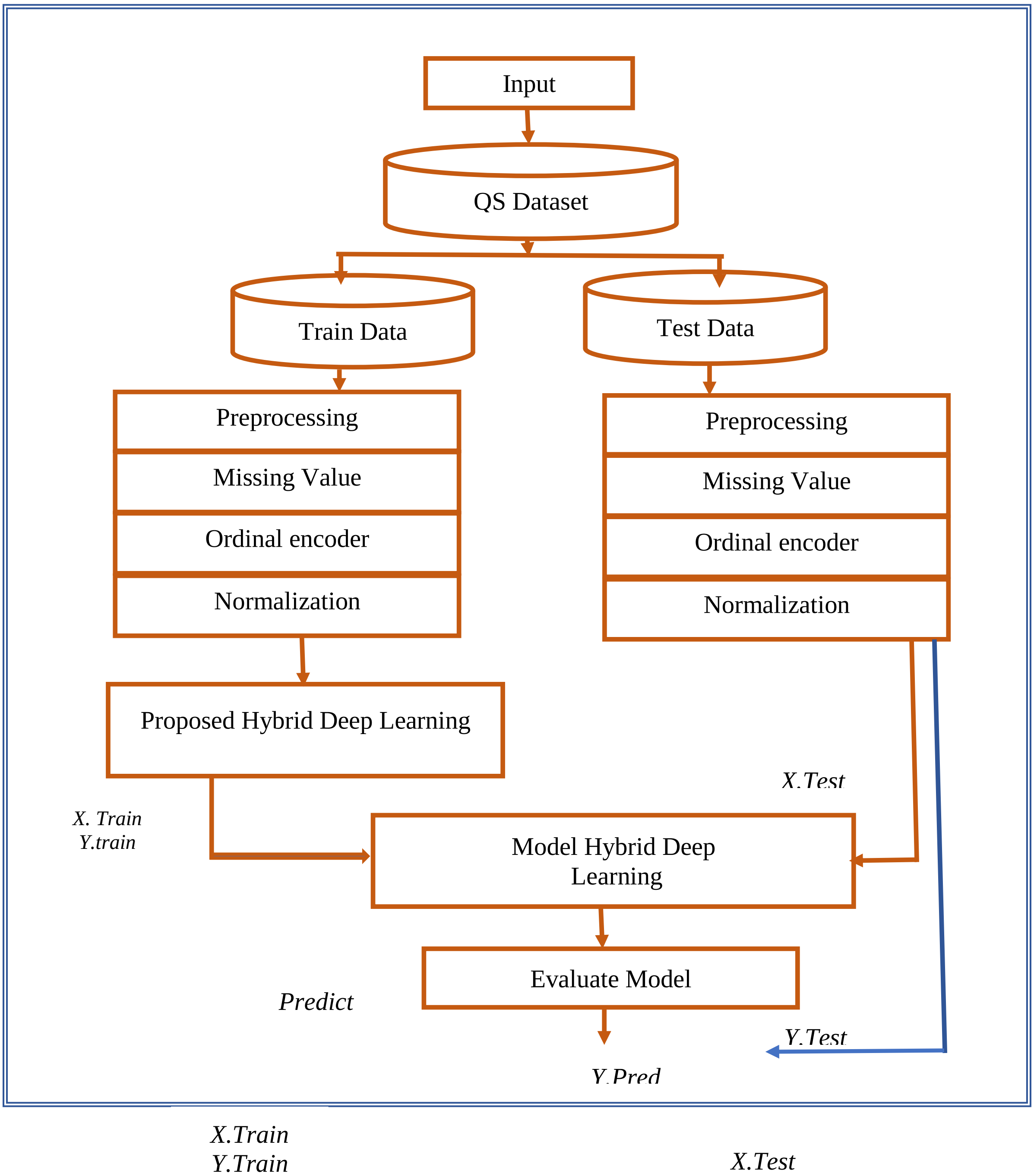


Figure 3: The Structure of the Hybrid Deep Learning Model

4.1 Proposed Hybrid Deep Learning Model

The convolutional neural network extracts the spatial features, while the long-short-term memory extracts the temporal features and employs the extracted features to generate the predicted results. The hybrid deep learning model can improve prediction accuracy by learning from mixed temporal and spatial features.

The model of hybrid deep learning contains twenty layers. Figure 4 illustrates these layers. The convolutional neural network model consists of four convolution layers. The first layer has 16 convolution filters with a kernel size of 3 × 1, strides of 1, and padding that is valid. The second layer has 32 convolution filters with a kernel size of 3 × 1, strides are 1, and padding is valid. The third layer has 64 convolution filters with a kernel size of 3 × 1, strides of 1, and padding that is valid. The fourth layer has 32 convolution filters with a kernel size of 3 × 1, strides of 1, and padding that is valid. The LSTM model consists of three layers. The first layer has 16 units, and the return sequences are true. The second layer, with 64 units and return sequences, is true. The third layer has 32 units, and the return sequences are false. The pooling layer consists of six max-pooling layers, with pool size being 1 and strides being 1. The activation function consists of six leaky, rectified linear units. The last layer, the dense layer, contains a single layer (one class) and a linear activation function. Table 1 shows the parameter setting of the hybrid deep model for this experiment.

Table 1: Parameter Setting of the Hybrid Deep Learning Model

Parameters	**Value**
Convolutional Layer Filters (conv-1, conv-2, conv-3, conv-4)	(16, 32, 64, 32)
Convolutional Layer kernal_size	3
Convolutional Layer Stride	1
Convolutional Layer Padding	valid
Pooling Layer pool_size	1
Pooling layer activation functions	Relu
LSTM Layer (output, return sequences)	((16, True), (64, True), (32, False))
LSTM Layer activation function	Linear
Learning Rate	0.001
Optimizer	Adam
Epochs	100
Batch_size	64
Loss function	(MAE, MSE, RMSPE)

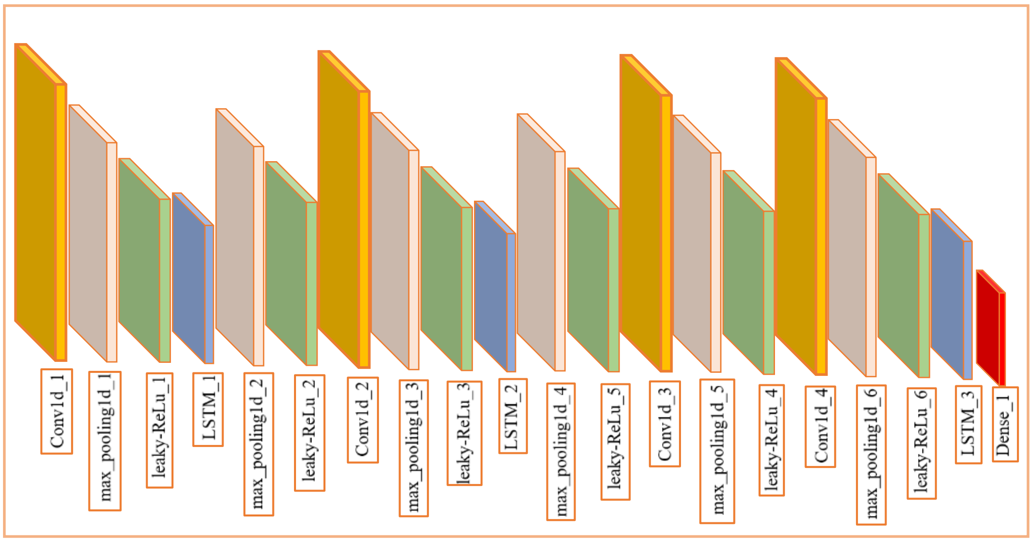


Figure 4. The Proposed Model Structure of Hybrid Deep Learning

5- Results and Discussions

5.1 Evaluation of the models

After determining the model's structure, the network is trained using the training set until convergence [37]. Three types of error metrics were used in this study to find out how well the models worked: mean absolute error (MAE), mean squared error (MSE), and root mean squared percentage error (RMSPE) [38]. Let Hk, k = 1,..., u denote the u validation data prediction scores derived for prediction methods (e.g., deep neural networks or machine learning methods) built using the training dataset. Further, let B k, k = 1,..., u denote the actual values of the target variable contained in the validation dataset. Last, let Ek = Bk - Hk, k = 1,..., u denote the u scoring (prediction) errors (over the validation dataset) associated with the chosen prediction method [39]. The following are commonly used scoring measures to evaluate prediction methods concerning the validation dataset:

$$\text{Mean Absolute Error: MAE} = \frac{\sum_{k=1}^{u} |E_k|}{u} \qquad \text{Eq. 14}$$

$$\text{Mean Square Error: MSE} = \frac{\sum_{k=1}^{u} E_k^2}{u} \qquad \text{Eq. 15}$$

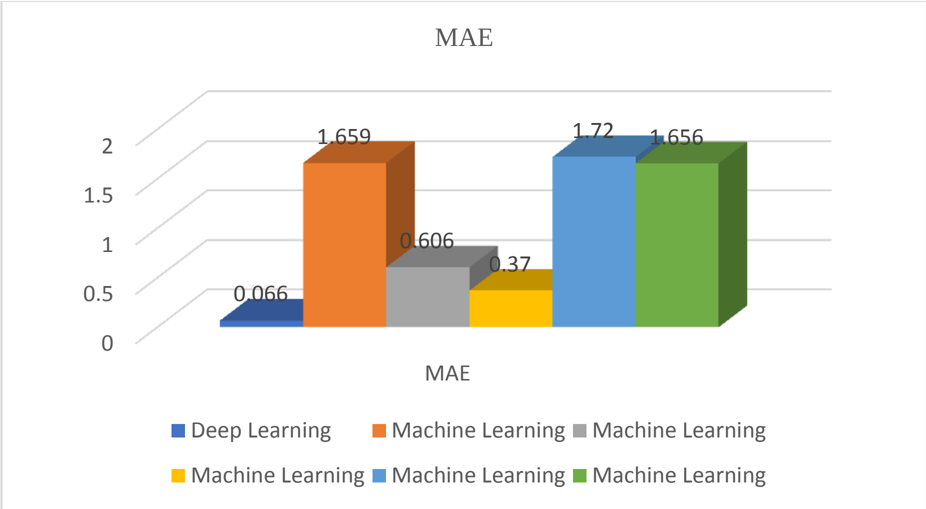
$$\text{Root Mean Square Precentage Error: RMSPE} = \sqrt{\text{MSPE}} \qquad \text{Eq. 16}$$

5.2 Performance Comparison of Hybrid Deep Neural with Machine Learning Models

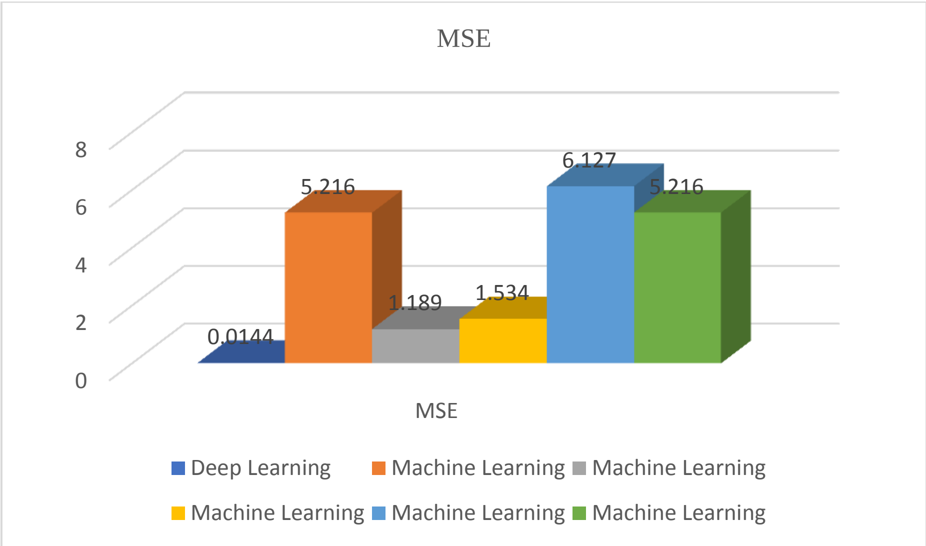
To validate the proposed hybrid CNN-LSTM deep learning model's efficiency, we compared it with five different machine learning models. We implemented these models on the same training dataset and test dataset. Table 2 lists the various machine learning model parameters to achieve the best performance for its model. Table 2 shows the results of the experiments performed. The hybrid CNN-LSTM model achieved the lowest mean absolute error (0.0659), mean squared error (0.0144), and root mean squared percentage error (0.121) compared to linear regression, decision tree regression, stochastic gradient descent regression, gradient boosting regression, and Bayesian ridge regression. Figure 5 shows the error metric in the hybrid deep neural learning model.

Table 2: Comparison of Hybrid Deep Learning Model with Machine Learning Models

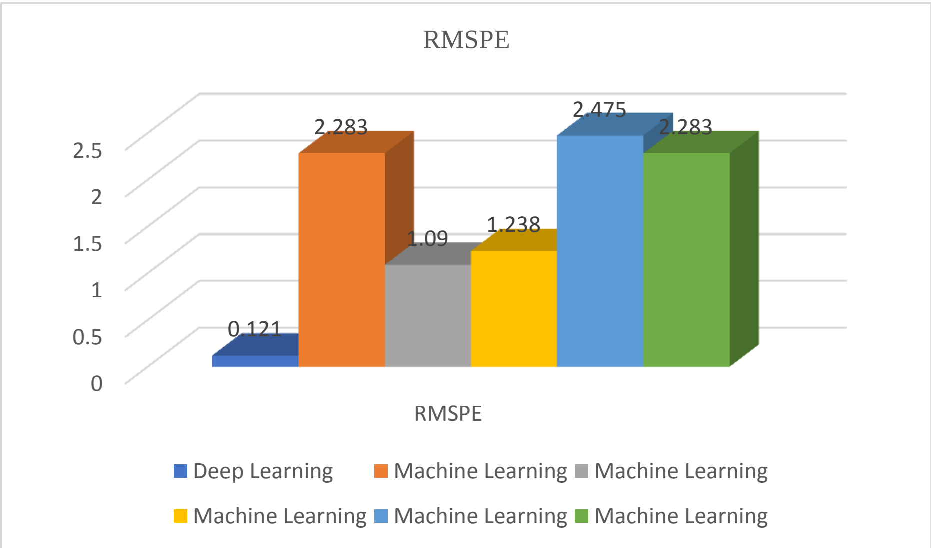
Error metrics	Deep Learning	Machine Learning				
	CNN+LSTM	Linear Reg.	Gradient-Bost-Reg.	Decision-T-Reg.	Stochastic-Grad-Reg.	Bayesian-Ridg-Reg.
MAE	0.066	1.659	0.606	0.37	1.72	1.656
MSE	0.0144	5.216	1.189	1.534	6.127	5.216
RMSPE	0.121	2.283	1.09	1.238	2.475	2.283



(A)



(B)



(C)

Figure 5: Represents the Performance Evaluation Results of Predicting Models
(A): Mean Absolute Error (MAE).
(B): Mean Squared Error (MSE).
(C): Root Mean Squared Percentage Error (RMSPE).

The results of the proposed system are presented as a BI dashboard. All the implemented visualizations represented answers to the questions that were asked in the deep learning stage. All the visualization in the proposed system has been done using Microsoft Power BI-free software.

1. Pie chart visualization is used to visualize the total number of faculty or academic staff at the university.
2. A tree map visualization is used to visualize the total number of students at each university. Each university's representation in the tree map is a rectangle, and the size of the rectangle corresponds to the number of students.
3. The doughnut chart visualization is used to visualize the type of university (private or public).
4. Line Chart visualization is used to visualize the rank given to the university; the X-axis represents the university, and the Y-axis represents the rank given to the university.
5. Geographical map visualization is used to visualize the total university ranking for each country that is distributed among the geographical areas based on university location in the world.

6- Conclusion

By utilizing the power of data and advanced neural networks, business intelligence and deep learning work together to detect significant insights, enhance data analysis, allow better decisions, and push prediction capabilities in a variety of fields. In this paper, the Proposed Hybrid CNN-LSTM model performed very well on the World University Rankings QS data set in terms of accuracy. The Proposed Hybrid CNN-LSTM model achieved 98.5 % accuracy and the Mean Squared Error: is 0.0145 as compared to traditional machine learning such as (Linear regression, Decision Tree Regression, stochastic Gradient Descent Regression, Gradient Boosting Regression, and Bayesian Ridge Regression).

Looking into the future, based on our analysis, we recommend the following actions and potential paths for further research and development:

1. To improve the use of data science and machine learning for university rankings, encourage cooperation and information exchange among researchers by sharing insights, ideas, and methods.
2. Perform continuous model monitoring and updates to ensure that the model remains correct and relevant as the dataset changes or new information becomes available. Periodically retraining the model can help maintain its accuracy and effectiveness over time.
3. Use data augmentation strategies to improve overall robustness by increasing the model's generalization on small datasets.
4. To address unbalanced data in a dataset, investigate techniques such as oversampling, under-sampling, or alternate assessment metrics to address unbalanced data and ensure that the final results are more accurate and reliable.
5. To improve the performance of the hybrid CNN-LSTM model, it would be worth exploring hyperparameter fine-tuning while considering various structures, learning rates, and regularization strategies. This could potentially lead to significant improvements in the model's accuracy and overall performance.

References

[1] M. D. Tamang, V. Kumar Shukla, S. Anwar and R. Punhani, “Improving Business Intelligence through Machine Learning Algorithms,” *International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom,* pp. 63-68, 2021. Doi: 10.1109/ICIEM51511.2021.9445344.

[2] S. A. Jebur, k. A. Huseen, H. K. Hoomod, L. Alzubaidi, and J. Santamaria, "Review on Deep Learning Approaches for Anomaly Event Detection in Video Surveillance," *Electronics*, vol. 12, no. 1, p. 29, 2022. https://doi.org/10.3390/electronics12010030.

[3] B. Nedelcu, "Business intelligence systems," *Database Systems Journal*, vol. IV, no. 4, pp. 12-20, 2013.

[4] S. Bashar, "Business Intelligence Based on Big Data Solver approach," *Master thesis at University of Technology/Baghdad, Computer of Science Department*, 95 pages, 2015.

[5] W. A. Kan, S. H. Chung, M. U. Awan, and X. Wen, "Machine learning facilitated business intelligence (Part I) Neural networks learning algorithms and applications," *Industrial Management & Data Systems*, vol. 120, pp. 164-195, 2020.

[6] H. F. Alaskar, T. Saba, "Application of Business Intelligence Solution Development and Implementation in a Small-Sized Enterprise," *First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia*, pp. 183-190, 2020. Doi: 10.1109/SMART-TECH49988.2020.00051.

[7] M. Hamzehi and S. Hosseini, "Business intelligence using machine learning algorithms," *Multimedia Tools and Applications*, vol. 81, pp. 33233-33251, 2022.

[8] Z. X. Huang, K. S. Savita, J. Z. Jie, "The Business Intelligence impact on the financial performance of start-ups," *Information Processing & Management*, vol. 59, no. 1, p. 102761, Jan. 2022. https://doi.org/10.1016/j.ipm.2021.102761.

[9] T. Kanan, A. Mughaid, R. AL-Shalabi, M. Al-Ayyoub, M. Elbes, and O. Sadaqa, "Business intelligence using deep learning techniques for social media contents," *Cluster Computing*, vol. 26, pp. 1285–1296, 2023. https://doi.org/10.1007/s10586-022-03626-y.

[10] D. Borissova, N. Keremedchieva, and D. Keremedchiev, "Business Intelligence Approach to Support Decision Making in Publishing Sector," *International Convention on Information, Communication and Electronic Technology (MIPRO)*, pp. 1268-1273, 2020. Doi: 10.23919/MIPRO48935.2020.9245424.

[11] A. Kelić, A. Almisreb, N. M. Tahir and J. Bakri, "Big Data and Business Intelligence - A Data-Driven Strategy for Business in Bosnia Herzegovina," *International Conference on Systems, Process & Control (ICSPC)*, pp. 226-230, 2022. Doi: 10.1109/ICSPC55597.2022.10001788.

[12] M. Banisharif, A. Mazloumzadeh, M. Sharbaf and B. Zamani, "Automatic Generation of Business Intelligence Chatbot for Organizations," *International Computer Conference, Computer Society of Iran (CSICC)*, pp. 1-5, 2022. Doi: 10.1109/CSICC55295.2022.9780490.

[13] Z. Desai, K. Anklesaria and H. Balasubramaniam, "Business Intelligence Visualization Using Deep Learning Based Sentiment Analysis on Amazon Review Data," *International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-7, 2021. Doi: 10.1109/ICCCNT51525.2021.9579786.

[14] I. Sreesurya, H. Rathi, P. Jain, T. K. Jain, "Hypex: A tool for extracting business intelligence from sentiment analysis using enhanced LSTM," *Multimedia Tools and Applications*, vol. 79, pp. 35641-35663, 2020. https://doi.org/10.1007/s11042-020-08930-6.

[15] https://www.kaggle.com/datasets/padhmam/qs-world-university-rankings-2017-2022.

[16] https://www.kaggle.com/datasets/divyansh22/QS-World-University-Rankings.

[17] M. N. Norazian, A. M. Al Bakri, Y. A. Shukri, R. N. Azam, "Estimation of missing values for air pollution data using interpolation technique," *Simulation*, vol. 75, pp. 94-99, 2006.

[18] JW. Graham, "Missing data analysis: Making it work in the real world," *Annual review of psychology*, vol. 60, pp. 549-576, 2009.

[19] National Research Council (US) Panel on Handling Missing Data in Clinical Trials, "The Prevention and Treatment of Missing Data in Clinical Trials," *National Academies Press (US)*, 2010. doi:10.17226/12955

[20] R. T. O'Neill, R. Temple, "The prevention and treatment of missing data in clinical trials," *New England Journal of Medicine*, vol. 367, pp. 1355-1360, 2012.

[21] X. Meng, C. Fan, Y. Ming, and H. Yu, "CORNet: Context-Based Ordinal Regression Network for Monocular Depth Estimation," *Transactions on Circuits and Systems for Video Technology*, vol. 32, pp. 4841-4853, 2022. Doi: 10.1109/TCSVT.2021.3128505.

[22] K. Bringmann, M. Künnemann, and K. Węgrzycki, “Approximating APSP without scaling: equivalence of approximate min-plus and exact min-max,” *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 943-954, 2019.

[23] K. Ovtcharov, O. Ruwase, J. Kim, J. Fowers, K. Strauss, and E. Chung, “Accelerating deep convolutional neural networks using specialized hardware,” *Microsoft Research Whitepaper 2*, vol. 11, pp. 1-4, 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84-90, 2017. https://doi.org/10.1145/3065386.

[25] D. Zhou, “Theory of deep convolutional neural networks: Downsampling,” *Neural Networks, Elsevier*, vol. 124, pp. 319-327, Apr. 2020. https://doi.org/10.1016/j.neunet.2020.01.018.

[26] G. Nair, S. Johnson, A. Verma, and A. Ebenezer, “A Fully Convolutional Deep Neural Network for Lung Tumor Identification,” *International Journal of Applied Engineering Research*, vol. 15, no. 5, pp. 431-435, 2020.

[27] J. Gu, Z. Wang, j. Kuen, L. Ma, A Shahroudy, B. Shuai, T. Liu, X. Wang, G. Cai, and T. Chen, “Recent advances in convolutional neural networks,” *Pattern Recognition*, vol. 77, pp. 354-377, 2018.

[28] T. Li, M. Hua, and X. Wu, "A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM2.5),” *in IEEE Access*, vol. 8, pp. 26933-26940, 2020. Doi: 10.1109/ACCESS.2020.2971348.

[29] H. Zhuang, R. Zhao, X. Luo, and C. Yang, “Research on Quantitative Stock Selection Strategy Based on CNN-LSTM,” *International Conference on Pattern Recognition and Artificial Intelligence (PRAI), Chengdu, China*, pp. 1142-1147, 2022. Doi: 10.1109/PRAI55851.2022.9904291.

[30] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, “Stock price prediction using LSTM, RNN and CNN-sliding window model,” *International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India*, pp. 1643-1647, 2017. Doi: 10.1109/ICACCI.2017.8126078.

[31] S. Yan, “Understanding LSTM and Its Diagrams,” June 2018. *Available online: https://medium.com/mlreview/understanding-lstm-and-its-diagrams-37e2f46f1714*.

[32] D. Goularas and S. Kamis, “Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data,” *International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), Istanbul, Turkey*, pp. 12-17, 2019. Doi: 10.1109/Deep-ML.2019.00011.

[33] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *in Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998. Doi: 10.1109/5.726791.

[34] J. H. Assi and A. T. Sadiq, “Modified Artificial Immune System as Feature Selection,” I*raqi Journal of Science*, vol. 59, No. 2A, pp. 733-738, 2018.

[35] T. Kim, and S. Cho, “Predicting residential energy consumption using CNN-LSTM neural networks. Energy,” *Energy, Elsevier*, vol. 182, pp. 72-81, 2019.

[36] M. A. Kadhim and A. M. Radhi, “Machine Learning Prediction of Brain Stroke at an Early Stage,” I*raqi Journal of Science*, vol. 64, no. 12, 2023.

[37] S. Jayawardena, J. Epps, and E. Ambikairajah, “Evaluation Measures for Depression Prediction and Affective Computing,” *International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK*, pp. 6610-6614, 2019. Doi: 10.1109/ICASSP.2019.8682956.

[38] Y. Zhong, A. Schwing, and J. Peng, “Disentangling Controllable Object Through Video Prediction Improves Visual Reinforcement Learning,” *International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain*, pp. 3672-3676, 2020. Doi: 10.1109/ICASSP40776.2020.9053819.

[39] S. Lei and M. Smith, “Evaluation of several Efron bootstrap methods to estimate error measures for software metrics,” *Canadian Conference on Electrical and Computer Engineering (CCECE)*, vol. 2, pp. 703-708, 2002. Doi: 10.1109/CCECE.2002.1013027.