National College of Ireland

Configuration Manual:

Deep Learning–Enhanced CPU Usage Prediction for Cloud Resource Optimization

MSc Research Project
Msc cloud computing

Forename Surname-E
Hemanth kumar
Student ID: X22183744

School of Computing
National College of Ireland

Supervisor: Shaguna Gupta

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	E.Hemanth kumar
Student ID:	X22183744
Programme:	Msc cloud computing
Year:	2024-2025
Module:	MSc Research Project
Supervisor:	Shaguna Guppta
Submission Due Date:	20/12/2018
Project Title:	Configuration Manual
Word Count:	600
Page Count:	5

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	E.Hemanth kumar
Date:	2nd January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

E.Hemanth kumar

X22183744

1. Introduction

Cloud computing platforms host large-scale virtualized infrastructures that execute diverse workloads ranging from web services to data-intensive analytics. Efficient CPU resource utilization is critical for maintaining performance, scalability, and cost-effectiveness in such environments. However, CPU usage in cloud systems exhibits highly dynamic and non-linear temporal behaviour due to workload fluctuations and user demand variability.

Traditional forecasting techniques often fail to model these complex temporal dependencies. Deep learning–based time series models, including CNN, GRU, LSTM, and BiLSTM, have demonstrated superior performance in capturing sequential patterns. This project focuses on leveraging deep learning architectures, including an advanced BiLSTM-CATF model, to accurately forecast short-term CPU utilization using real cloud workload traces.

Accurate CPU usage forecasting enables improved cloud resource allocation, proactive scaling, and performance optimization.

2. Research Question

Primary Research Question

How effectively can deep learning models forecast CPU utilization in a cloud environment using time-series workload traces, and does the inclusion of advanced architectures improve prediction accuracy and performance efficiency?

Explanation of the Research Question

Cloud CPU utilization is influenced by:

- temporal workload variations,
- VM execution behaviour,
- underlying infrastructure capacity.

This study investigates:

- the ability of deep learning models to learn temporal dependencies from historical CPU usage,
- comparative performance of CNN, GRU, LSTM, BiLSTM, and BiLSTM-CATF models,
- forecasting accuracy and computational efficiency in a cloud-based setup.

3. Environment Setup

Proper environment configuration ensures reproducibility and consistent model performance.

3.1 Recommended Hardware

Component	Recommended
GPU	NVIDIA T4 / RTX series
RAM	Minimum 12 GB
CPU	Multi-core processor
Storage	~10 GB

3.2 Software Requirements

Supported operating systems:

- Windows
- Linux
- Cloud-based environments (Google Colab, AWS Cloud9)

3.3 Python Version

Python version ≥ 3.7 is required.

3.4 Required Libraries

Core libraries:

- numpy
- pandas
- matplotlib
- seaborn

Deep learning:

- tensorflow
- keras

Machine learning utilities:

- scikit-learn

System monitoring:

- time
- psutil

3.5 Installation Commands

```
pip install numpy pandas matplotlib seaborn scikit-learn
pip install tensorflow keras
pip install psutil
```

4. Dataset Description

Primary Dataset: Microsoft Azure Public VM Traces

The dataset consists of time-series CPU utilization measurements collected from Azure virtual machines at **5-minute intervals**. It represents real-world cloud workload behavior.

Key characteristics:

- time-indexed CPU usage values,
- long-duration VM execution traces,
- suitable for temporal forecasting tasks.

Target Variable: CPU utilization

(Screenshots of dataset loading and structure should be added here)

5. Data Pre-processing

Raw VM trace data is loaded and cleaned by handling missing values and removing irrelevant attributes. Timestamps are converted to datetime format, and CPU utilization values are scaled using Min–Max normalization. A rolling window technique is applied to transform the time series into supervised learning sequences.

6. Exploratory Data Analysis

Exploratory analysis is performed to visualize CPU utilization trends and workload variability over time. Time-series plots help identify temporal patterns and fluctuations in resource usage, providing insights into the forecasting complexity.

7. Model Architectures

Multiple deep learning models, including CNN, GRU, LSTM, BiLSTM, and BiLSTM-CATF, are implemented for CPU usage forecasting. These architectures are designed to effectively capture temporal dependencies in cloud workload data.

8. Training Pipeline

The dataset is temporally split into 90% training and 10% testing sets. Windowed and normalized

sequences are used to train models with the Adam optimizer, while execution time and memory usage are monitored during training.

9. Evaluation Metrics

Model performance is evaluated using MSE, RMSE, MAE, and R^2 score. Additional assessment includes actual versus predicted CPU usage plots along with latency, throughput, and execution time analysis.

10. Future Forecasting

Trained models are used to predict future CPU utilization values for multiple time steps. These forecasts support proactive cloud resource management and performance optimization.

References

Agrawal, S., Singh, D., 2024. Study Containerization Technologies like Docker and Kubernetes and their Role in Modern Cloud Deployments, in: 2024 IEEE 9th International Conference for Convergence in Technology (I2CT). Presented at the 2024 IEEE 9th International Conference for Convergence in Technology (I2CT), IEEE, Pune, India, pp. 1–5. https://doi.org/10.1109/I2CT61223.2024.10543986

Ahmadzadeh, M., Zahrai, S.M., Bitaraf, M., 2025. An integrated deep neural network model combining 1D CNN and LSTM for structural health monitoring utilizing multisensor time-series data. Struct. Health Monit. 24, 447–465. https://doi.org/10.1177/14759217241239041

Akoh Atadoga, Uchenna Joseph Umoga, Oluwaseun Augustine Lottu, Enoch Oluwademilade Sodiya, 2024. Evaluating the impact of cloud computing on accounting firms: A review of efficiency, scalability, and data security. Glob. J. Eng. Technol. Adv. 18, 065–075. https://doi.org/10.30574/gjeta.2024.18.2.0027

Ali, N.N., Zeebaree, S.R.M., 2024. Distributed Resource Management in Cloud Computing: A Review of Allocation, Scheduling, and Provisioning Techniques. Indones. J. Comput. Sci. 13. https://doi.org/10.33022/ijcs.v13i2.3823

Bala, B., Behal, S., 2024. A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques, in: 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). Presented at the 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, Kirtipur, Nepal, pp. 1755–1762. https://doi.org/10.1109/I-SMAC61858.2024.10714767

Betti Pillippuge, T.L., Khan, Z., Munir, K., 2025. Horizontal Autoscaling of Virtual

Machines in Hybrid Cloud Infrastructures: Current Status, Challenges, and Opportunities. Encyclopedia 5, 37. https://doi.org/10.3390/encyclopedia5010037