

Memory-Augmented GRUs for Food Loss Forecasting: A Survey of Architectures and Design Trade-offs

Ru Poh Tan^{1*}, Siew Mooi Lim^{1*}, Kuan Yew Leong², Jun Kit Chaw³, Siaw Hong Liew⁴,
and Shee Chia Lee¹

¹ Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur 53300, Malaysia

[e-mail: tanrp-wm21@student.tarc.edu.my, siewmooi@tarc.edu.my, leesheechia@tarc.edu.my]

² Research and Development Department, A.I. System Research Co., Ltd., Kyoto, Japan
[e-mail: kyleong@aisr.dev]

³ Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[e-mail: chawjk@ukm.edu.my]

⁴ Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia
[e-mail: shliew@unimas.my]

*Corresponding author: Ru Poh Tan, Siew Mooi Lim

*Received January 14, 2026; revised March 29, 2026; accepted April 20, 2026;
published August 31, 2026*

Abstract

Forecasting food loss is pivotal to sustainable, efficient food systems, yet it demands models that can retain long-horizon patterns, handle exogenous drivers, and remain interpretable to practitioners. This survey investigates modern time-series methods through the lens of Gated Recurrent Units (GRUs) and argues that augmenting GRUs with explicit memory is a promising, underutilized strategy for food-loss prediction. We synthesize evidence across GRU-centered variants and five families of neural memory architectures - Neural Turing Machines (NTM), Differentiable Neural Computers (DNC), Relational Memory (RelMem), Entity-centric Memory (EntMem), and Modular Memory Units (MMUs) - organizing their design choices (controllers, read/write addressing, integration points) and reported trade-offs. Drawing on comparative studies, we outline a practical MAGRU blueprint that maps memory mechanisms to common properties of food-loss series (seasonality, shocks, missingness), and distill guidance on when external memory helps, how to balance accuracy with computational cost, and how to couple memory with attention and exogenous-variable embeddings for greater interpretability. Our review identifies key gaps - limited evaluations on realistic supply-chain datasets, the absence of standard benchmarks, and little exploration of lightweight or hybrid MAGRU designs, and proposes a research agenda to close them. Overall, memory augmentation consistently strengthens long-range dependency modeling in GRU-based forecasters while introducing complexity that must be managed, positioning MAGRU as a compelling foundation for actionable, trustworthy food-loss forecasting.

Keywords: GRU, Food Loss forecasting, Memory-Augmented GRU (MAGRU), time-series.

1. Introduction

Food loss, the decline in food quantity or quality during production and logistics, represents a critical threat to global sustainability, costing the global economy approximately USD 1 trillion annually and generating 8–10% of greenhouse gas emissions [1, 2, 3, 4]. These staggering figures highlight the urgent need for accurate forecasting tools [5] across the supply chain. While Gated Recurrent Units (GRUs) are efficient for sequential data, their standard design struggles with the long-horizon patterns inherent in food systems. Integrating memory-augmented mechanisms (MAGRU) into GRU models provides a promising pathway to overcome these limitations by capturing long-term dependencies and yielding more reliable and actionable insights [6, 7, 8].

Our comprehensive review examines existing memory-augmented architectures, including Neural Turing Machines (NTM), Differentiable Neural Computers (DNC), Relational Memory (RelMem) and Entity Memory (EntMem), Modular Memory Units (MMUs), to investigate their potential for enhancing food loss forecasting. By synthesizing literature from various time-series domains, we investigate how integrating these mechanisms enhance the performance, adaptability, and interpretability of traditional GRU models.

The paper is structured as follows: Section II reviews existing literature on enhancing GRU models with memory-augmented mechanisms; Section III establishes a methodological framework and comparative summary of key studies; Section IV analyzes the associated advantages and technical limitations; Section V identifies critical research gaps and future directions; and Section VI concludes the paper with final insights and implications.

2. Literature Review

This section reviews GRU-based time-series forecasting literature that enhances long-term dependency retention through memory-augmented architectures, with emphasis on their applicability to food loss prediction. Key developments, strengths, limitations, and future potential are examined across five architecture families.

2.1 Memory-Augmented GRU (MAGRU) for Long-Term Dependencies

Memory-Augmented GRU (MAGRU) architectures extend standard GRUs with external or internal memory mechanisms that enable selective storage and retrieval of information across extended time horizons. This capability is critical for food loss forecasting, where historical patterns, spoilage progression, storage conditions, and supply chain disruptions, must be preserved far beyond what a GRU hidden state can reliably hold. The five architecture families reviewed below differ along four dimensions bearing directly on MAGRU design: (i) memory access mode (how reads and writes are addressed), (ii) memory structure (flat matrix, relational slots, entity slots, or cell-integrated blocks), (iii) computational cost (ranging from near-GRU overhead to $O(N^2)$ attention), and (iv) GRU controller coupling (external-decoupled, attention-mediated, or cell-integrated).

2.1.1 Cell-Integrated Memory: GRU-MB, MGRU, and MMU

The GRU with Memory Block (GRU-MB) architecture (Fig. 1), introduced by Khadka et al. [9, 10], extends the standard GRU by decoupling memory interactions from the core hidden-state update through gated read/write mechanisms. This design enhances the model's ability to capture long-term dependencies while mitigating noise accumulation. Fig. 2 demonstrates

the full GRU-MB architectures, and **Fig. 3** compares GRU-MB and feedforward policies in 2-deep and 3-deep Extended T-Mazes, showing that robots using the GRU-MB network consistently outperform feedforward counterparts across all team sizes and task depths. These properties make GRU-MB well-suited for food loss forecasting, where long-range retention of spoilage patterns and supply chain events is essential.

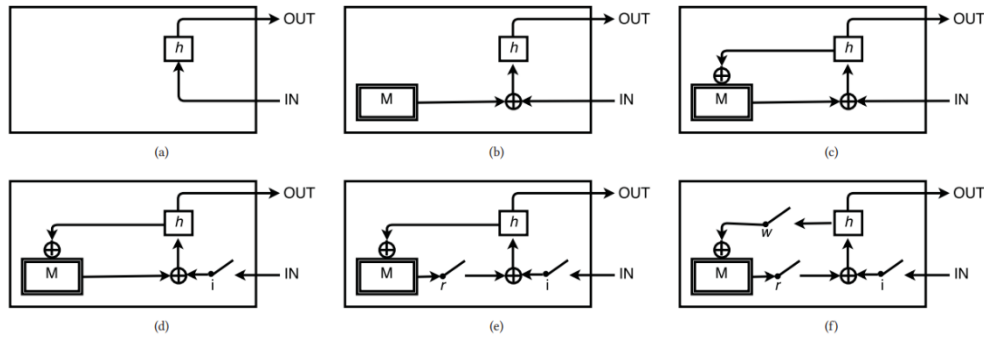


Fig. 1. Steps (a) to (f) illustrate a progressive enhancement of a GRU-based model, starting from a simple neural network and gradually incorporating memory blocks, feedback mechanisms, and gated controls. This process strengthens the model's capacity to store, regulate, and retrieve information over time [9].

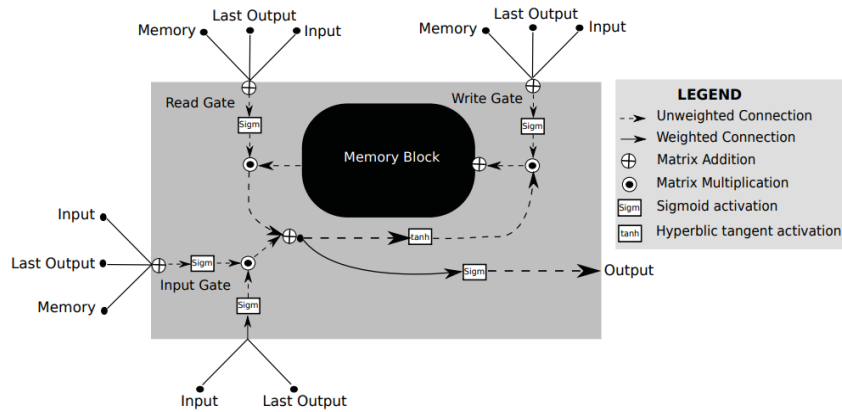


Fig. 2. Full GRU-MB neural architecture schematic detailing the weighted connections and activations [9].

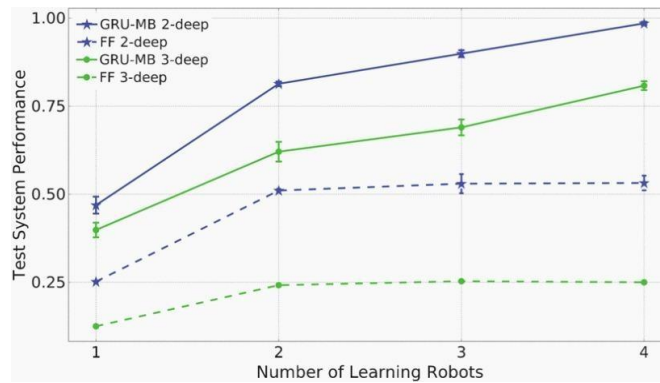


Fig. 3. Comparison of GRU-MB and feedforward policies in 2- and 3-deep Extended T-Mazes across varying numbers of learning robots [10].

In a more recent example, Memory-Augmented GRU (MGRU), as detailed by Yang, Li & Li [11] incorporates a fractionally integrated filter into the GRU cell to model long-memory effects in multivariate time series. As shown in Fig. 4, memory-augmented GRU with fixed-d (MGRUF) consistently outperforms both standard GRU and LSTM across evaluation metrics, demonstrating that even lightweight parametric filters meaningfully extend a GRU's effective memory horizon. The related Modular Memory Unit (MMU) architecture [12] generalises the same cell-integrated principle using neuroevolution to optimise dedicated memory blocks, achieving strong performance on deep memory benchmarks with minimal overhead. For food loss forecasting, where deployment in resource-constrained supply-chain environments is a practical requirement, cell-integrated designs offer the most favourable accuracy-to-cost ratio.

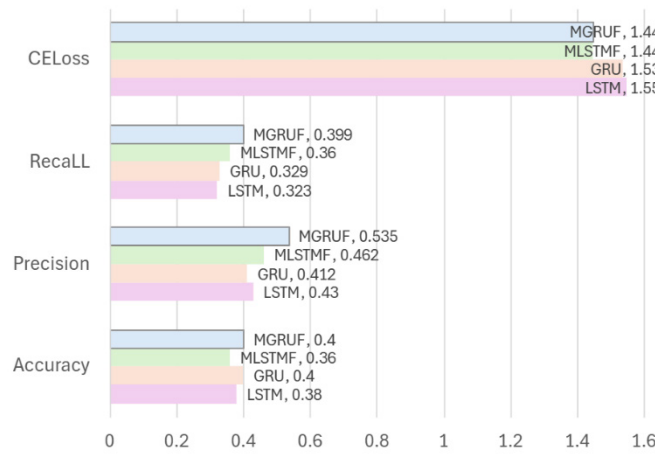


Fig. 4. Performance comparison with MGRUF achieving consistently superior performance across evaluation metrics [11].

Although grouped under the cell-integrated category, GRU-MB, MGRU, and MMU are independent designs: GRU-MB decouples memory via gated read/write mechanisms, MGRU extends the GRU cell with a fractional integration filter, and MMU uses neuroevolution to optimise modular memory blocks.

2.1.2 External Matrix Memory: Neural Turing Machines (NTM)

Proposed by Graves et al. [13] and reviewed comprehensively by Faradonbe et al. [14] (Fig. 5), NTMs combine a neural controller with an external memory matrix accessed via differentiable read/write operations, enabling end-to-end learning of algorithmic tasks such as copying, sorting, and associative recall. NTM is the most broadly validated external-memory architecture across domains and laid the conceptual groundwork for the Differentiable Neural Computer (DNC).

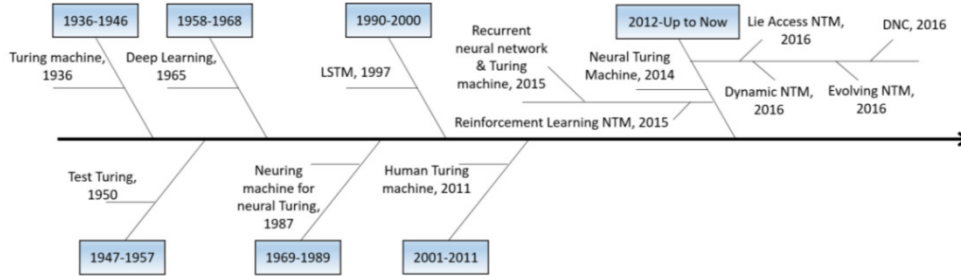


Fig. 5. The chronology of Neural Turing Machine development [14].

A. GRU + NTM

In 2024, Panggabean et al. [15] introduced a hybrid GRU-NTM model for detecting DoS and DDoS attacks, leveraging the sequential learning capabilities of GRU and the long-term memory of NTM. **Fig. 6** illustrates the architecture of the proposed model, which effectively captures complex temporal patterns in network traffic and achieves strong performance in intrusion detection tasks. The use of memory augmentation through NTMs enhances the model's ability to adapt to evolving threats, offering a robust solution for real-time intrusion detection; similarly, in food loss forecasting, NTMs can store and retrieve past spoilage events, storage conditions, and transport disruptions, allowing the GRU-NTM model to produce more informed, context-aware predictions under irregular and event-driven loss patterns.

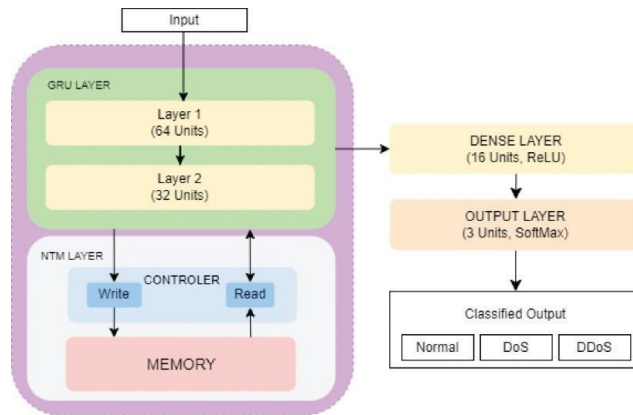


Fig. 6. Hybrid architecture integrating GRU layers, dense and softmax layers, with an NTM memory module for enhanced sequence processing and classification into Normal, DoS, or DDoS [15].

The Sparse Access Neural Turing Machine (SANTM) [16], which integrates sparse memory access, local multi-head self-attention, and ChebNet spectral graph convolution, is designed to improve long-term memorization while reducing computational overhead. When paired with GRU layers, SANTM can prioritize the most relevant temporal patterns while filtering out less informative sequences, enabling focus on critical historical patterns such as seasonal deterioration trends and supply chain bottlenecks.

Likewise, the Dynamic Neural Turing Machine (D-NTM) [17, 18], with trainable memory addressing and discrete attention, allows the GRU controller to retrieve highly specific historical contexts. The two architectures proposed in [19] for dialogue coherence, one with dual NTMs for each participant (**Fig. 7**) and another with a single NTM integrated into a language model (**Fig. 8**), demonstrate how NTMs maintain permanent, flexible storage of sequential history, improving response coherence. Applied analogously, a GRU-NTM hybrid

can maintain structured records of inventory rotations, handling schedules, and batch processes, supporting adaptive food loss forecasting.

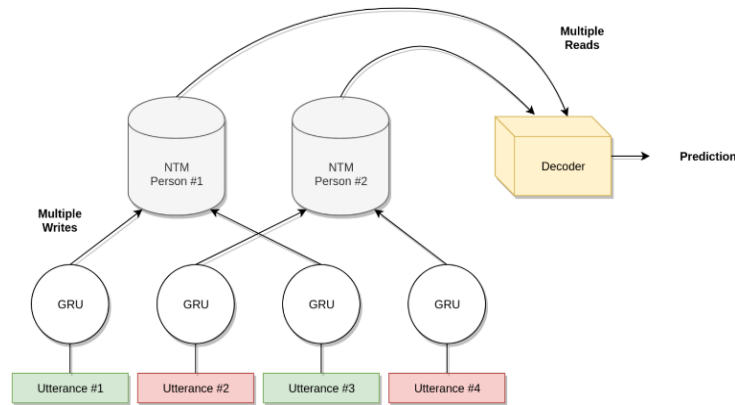


Fig. 7. Memory-augmented dialogue architecture with dual NTMs (D-NTMS) [19].

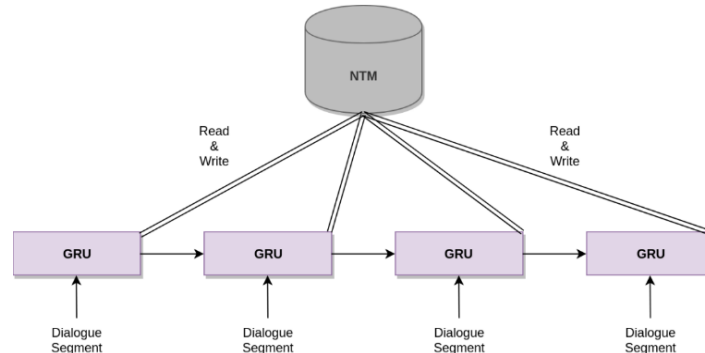


Fig. 8. Architecture of single-NTM language model dialogue system (NTM-LM) [19].

On the other hand, Xiong et al. [20] introduced a Multi-channel Encoder (MCE) integrating NTM external memory with GRU hidden states and raw word embeddings. The NTM-RNN-EMB variant achieves superior BLEU scores across varying sentence lengths (Fig. 9), demonstrating that blending external memory with recurrent states improves capture of compositional relationships in sequential data.

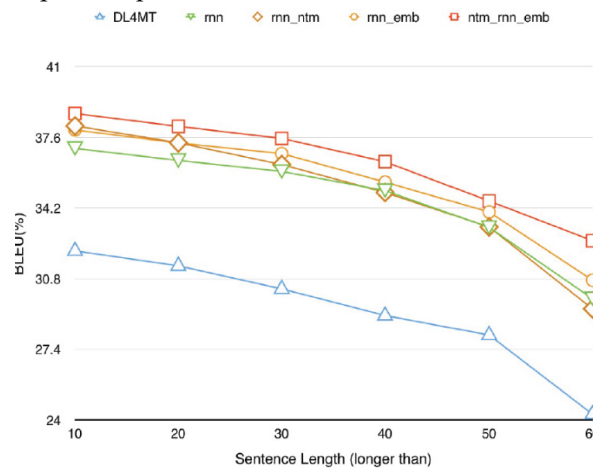


Fig. 9. Performance comparison of models across varying sentence lengths based on BLEU scores [20].

To develop a real-time pricing method for virtual power plants (VPP) demand response, Kong et al. [6] proposed a PER-DDPG algorithm integrating NTMs with GRUs and attention mechanisms to enhance the demand response evaluation and optimize electricity pricing strategies. Fig. 10 shows the training process of the Demand Response (DR) potential evaluation model based on NTM, Fig. 11 illustrates the overall VPP pricing architecture, and Fig. 12 demonstrates that NTM-augmented PER-DDPG achieves superior evaluation accuracy compared to baseline algorithms. The combination of memory and attention for real-time sequential decision-making mirrors the requirements of food loss estimation under dynamic supply chain conditions.

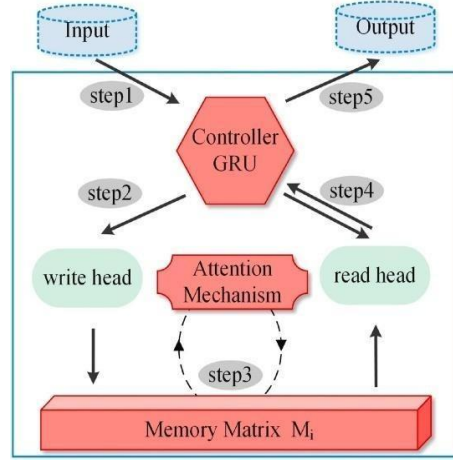


Fig. 10. Process of training DR potential evaluation model [6].

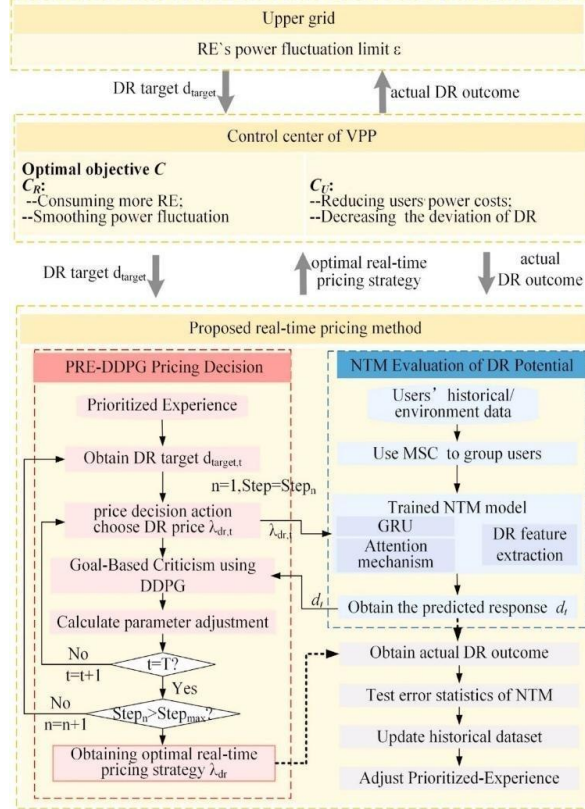


Fig. 11. Overall architecture of the VPP pricing method [6].

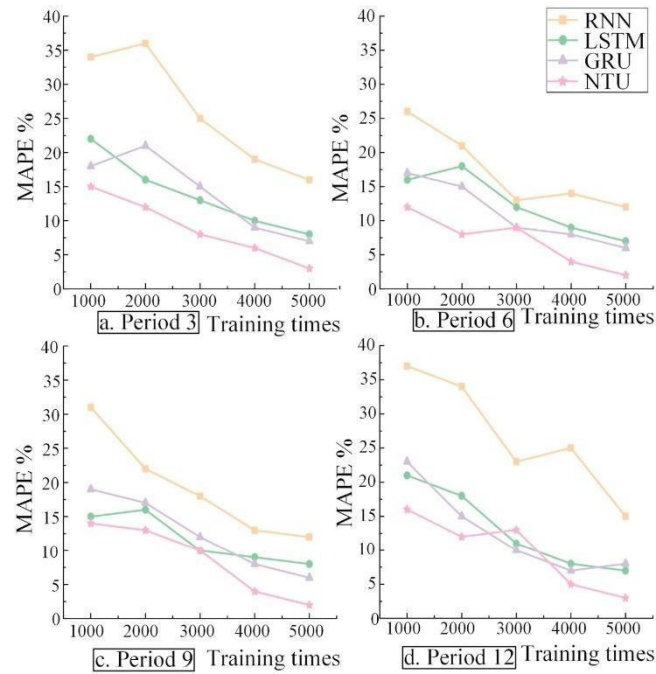


Fig. 12. Comparison of different algorithms across training iterations [6].

B. NTM Standalone and Non-GRU Studies

Studies [21] and [22] compare RNN, LSTM, GRU, and NTM for remaining useful life (RUL) estimation of mechanical components. Both show that NTMs outperform traditional sequence models through external memory with learnable read/write operations. Xiao et al. [7] developed an Interpretable Time-Series NTM (ITS-NTM) for diabetes prognosis that outperforms GRU, LSTM, RNN, SVM, and ANN on clinical time-series data (Fig. 13), demonstrating strong interpretability alongside predictive accuracy.

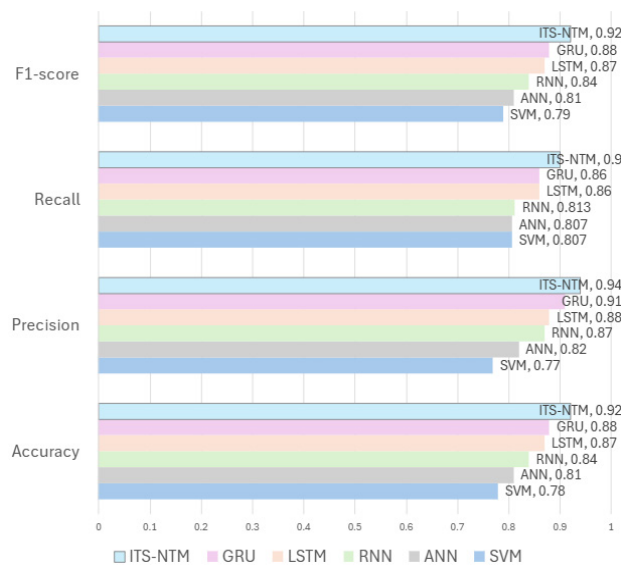


Fig. 13. Comparison of ITS-NTM and baseline models across F1-score, Recall, Precision, and Accuracy metrics [7].

Castellini [23] further shows that Feedforward Neural Turing Machine (FF-NTM) generalises binary arithmetic tasks better than LSTM-NTM, while Chen and Fan [24] demonstrate that an Adaptive Hard Sigmoid NTM (AH-NTM) reduces training time and improves efficiency over standard RNN, LSTM, GRU, and NTM models. Collectively, these works establish that NTM-style memory augmentation reliably improves accuracy and sequence learning efficiency across diverse sequential forecasting tasks, providing strong motivation for GRU-NTM integration in MAGRU designs.

2.1.3 External Matrix Memory with Temporal Linkage: Differentiable Neural Computers (DNC)

Differentiable Neural Computers (DNCs) extend NTM with temporal linkage matrices and usage-based allocation, enabling structured traversal of previously written memory locations. Unlike traditional RNNs, which rely solely on hidden states, DNCs can store and traverse structured information across extended sequences. This concept was originally introduced by Graves et al. [25].

Khan et al. [8] adapting DNC with dual controllers and a write-protected memory policy for predictive process analytics. This configuration supports tasks such as next-activity and suffix prediction, achieving competitive performance with fewer parameters compared to standard LSTM models and surpassing LSTM, GRU, and k-NN in accuracy and forecasting effectiveness. In food supply chains, DNC's structured memory traversal is well-suited to multi-step operational sequences and recurring degradation pathways, though its $O(N^2)$ cost makes it the most computationally demanding architecture reviewed here.

2.1.4 Slot-Based Memory: Relational Memory (RelMem) and Entity Memory (EntMem)

RelMem, introduced as the Relational Memory Core by Santoro et al. [26], applies multi-head self-attention across fixed memory slots, enabling inter-slot relational reasoning. It achieves state-of-the-art results in language modelling, reinforcement learning, and program evaluation. EntMem (EntNet), proposed by Henaff et al. [27], organises memory into discrete entity-keyed slots updated through gated mechanisms, supporting dynamic tracking of individual entity states over time.

Building on these concepts, [28] developed a Memory-augmented Dynamic Neural Relational Inference (MemDNRI), which employs dual memory pools (RelMem and EntMem) with GRU-based controllers to capture long-term dependencies in dynamic systems. Fig. 14 demonstrates the structure of MemDNRI. MemDNRI outperforms NRI and DNRI on trajectory prediction tasks, demonstrating that dual-pool slot-based memory with GRU controllers effectively captures long-term relational dependencies. For food loss forecasting, such architectures are particularly suited to tracking relational interactions among products, storage units, and handling processes, enabling detection of cascading degradation patterns.

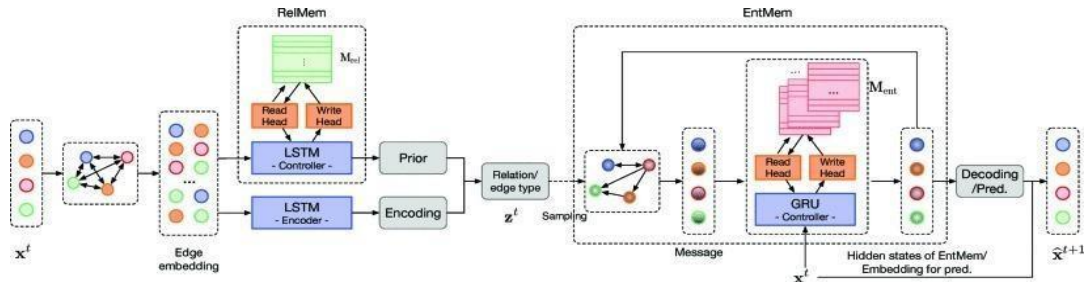


Fig. 14. Diagram of the proposed MemDNRI with dual RelMem and EntMem memory pools and GRU-based controllers [28].

2.1.5 Comparative Synthesis and MAGRU Design Implications

Three design trade-offs are relevant to MAGRU selection for food loss forecasting. First, regarding cost versus capability: NTM and DNC offer the strongest algorithmic memory but impose the highest training burden; MMU and GRU-MB achieve competitive long-range retention at near-GRU cost, favouring deployment-constrained settings such as farms and distribution centres. Second, regarding relational structure: when loss dynamics involve inter-entity dependencies, such as batch interactions and cascading spoilage, RelMem and EntMem provide explicit relational inductive biases absent in flat-matrix designs. Third, regarding GRU coupling: cell-integrated designs (MMU, GRU-MB) are directly compatible with standard GRU pipelines, whereas external-decoupled designs (NTM, DNC) require additional controller wrappers. Despite this integration overhead, the cross-domain validation history of external-memory architectures suggests meaningful transfer potential for food loss forecasting pipelines.

3. Methodological Review Framework

This section presents a methodological review of existing studies incorporating Memory-Augmented GRU (MAGRU) to improve long-term dependency modelling in time-series forecasting. Selected works are categorised by architectural design, integration strategy, and performance outcome.

Table 1 consolidates all 21 reviewed studies into a unified master table organised by the architecture family. The six columns capture architecture family, reference, application domain, specific model or method, key contribution or reported result, and whether GRU integration was employed. Row shading distinguishes the five architecture groups: cell-integrated memory (GRU-MB, MGRU, MMU), NTM-GRU hybrids, NTM standalone studies, DNC studies, and slot-based memory (RelMem/EntMem). **Table 2** provides the cross-architectural comparison across eight design dimensions that bears directly on MAGRU selection.

Table 1. Consolidated summary of reviewed studies on memory-augmented architectures for MAGRU.

Architecture family	Ref.	Application	Architecture / model	Key contribution / result	GRU integration
Cell-integrated memory: GRU-MB, MGRU, MMU					
GRU-MB / MMU	[9]	Deep memory tasks	GRU with Memory Block (GRU-MB); neuroevolution	Decoupled read/write gates improve long-range retention; outperforms feedforward across team sizes	Yes
GRU-MB / MMU	[10]	Multi-robot coordination	GRU-MB; cooperative coevolutionary algorithms (CCEA)	GRU-MB enables adaptive joint strategies in dynamic multi-agent environments	Yes
MGRU	[11]	Long-memory time-series forecasting	Memory-augmented GRU (MGRU) with fractional integration filter	MGRUF consistently outperforms standard GRU and LSTM on long-memory multivariate series	Yes
MMU	[12]	Deep memory tasks	Modular Memory Units (MMU); neuroevolution	Modular gated memory blocks outperform standard RNNs; efficient and generalisable	Yes (GRU/RNN base)
External matrix memory (NTM): GRU hybrid studies					
NTM + GRU	[15]	Network intrusion detection	Hybrid GRU-NTM	Combines GRU short-term and NTM long-term memory; achieves 99% accuracy on DoS/DDoS detection	Yes
NTM + GRU	[16]	Sequence learning and long-term memorisation	Sparse Access NTM (SANTM) with local multi-head self-attention	Reduces computational overhead while achieving superior long-term memorisation performance	Yes (GRU layers)
NTM + GRU	[17, 18]	Episodic QA and sequence learning	Dynamic NTM (D-NTM) with continuous and discrete addressing; GRU	Learnable addressing improves recall over standard NTM and LSTM on diverse tasks	Yes
NTM + GRU	[19]	Dialogue coherence	Memory-augmented recurrent networks with NTM (D-NTMS / NTM-LM);	Dual-NTM and single-NTM dialogue systems improve conversational context retention	Yes

			GRU		
NTM + GRU	[20]	Neural machine translation	Multi-channel encoder (MCE): NTM + GRU + word embeddings	NTM-RNN-EMB variant achieves superior BLEU scores by capturing compositional structure	Yes
NTM + GRU	[6]	Real-time VPP demand-response pricing	PER-DDPG; NTM; GRU; attention	NTM augmentation improves demand-response potential evaluation and real-time pricing accuracy	Yes
External matrix memory (NTM): standalone / non-GRU studies					
NTM	[13]	Algorithmic task learning	Neural Turing Machine (NTM)	Foundational NTM proposal; learns copying, sorting, associative recall via gradient descent	No (original)
NTM	[14]	NTM survey (ML/DL)	NTM; reinforcement learning	Comprehensive review of NTM architecture, tasks, and learning capabilities with external memory	No (survey)
NTM	[21]	RUL estimation (mechanical)	Neural sequence models: RNN, LSTM, GRU, NTM (comparative)	NTM and GRU both outperform RNN/LSTM; external memory bank improves preemptive maintenance	No (GRU compared, not integrated)
NTM	[22]	RUL estimation (mechanical)	NTM; deep learning	NTM achieves better RUL accuracy than LSTM with fewer parameters on multi-fault sequences	No
NTM	[7]	Type 2 diabetes prognosis	ITS-NTM (interpretable time-series NTM)	Outperforms GRU, LSTM, RNN, SVM, ANN on clinical time-series; high interpretability	No (NTM replaces GRU)
NTM	[23]	Binary arithmetic learning	Feedforward NTM (FF-NTM)	FF-NTM generalises binary addition/multiplication better than LSTM-NTM and standard LSTM	No
NTM	[24]	Tensile property testing (metals)	NTM; hard sigmoid (H-NTM); adaptive curriculum (AH-NTM)	AH-NTM reduces training time and improves efficiency vs RNN, LSTM, GRU, standard NTM	No

External matrix memory with temporal linkage (DNC)					
DNC	[25]	Cognitive tasks / neural memory	Differentiable Neural Computer (DNC)	Foundational DNC; structured memory with temporal linkage and usage-based allocation	No (original)
DNC	[8]	Predictive process analytics	MANN; DNC with dual controllers and write-protected memory	Surpasses LSTM, GRU, k-NN on next-activity and suffix prediction with fewer parameters	Partial (DNC controller)
Slot-based memory: RelMem, EntMem					
RelMem	[26]	Relational reasoning in sequential tasks	Relational Memory Core (RMC); multi-head dot-product attention	State-of-the-art on language modelling, RL, and program evaluation via inter-slot attention	No (original)
EntMem	[27]	Language understanding and reasoning	Recurrent Entity Network (EntNet); dynamic memory	State-of-the-art on bAbI tasks; maintains world-state estimate through gated entity slots	No (original)
RelMem + EntMem + GRU	[28]	Dynamic relational inference	MemDNRI: dual memory pools (RelMem + EntMem); GRU-based controllers	Outperforms NRI and DNRI on trajectory prediction; effective long-term relational dependency modelling	Yes

Table 2. Comparative analysis of memory-augmented architectures across design dimensions relevant to MAGRU.

Memory type	Core mechanism	Access mode	Memory structure	Complexity	GRU integration	Strengths	Limitations
[12, 9, 10]	Dedicated gated memory block inside recurrent cell	Gate-controlled module-wise	Cell-integrated memory block	Low; near-GRU overhead	Cell-integrated (native)	Lightweight ; deployment-friendly; modular	Limited exploration across diverse domains
[13, 14]	Differentiable Turing-style controller + external matrix	Content + location-based	Flat external matrix	High; $O(N \cdot M)$	External-decoupled	Algorithmic generalisation; broadest domain coverage	Harder to train; limited scalability

[25]	NTM + temporal linkage matrix + usage-based allocation	Content-based + temporal linkage + usage vector	Flat matrix + link/usage matrices	High; $O(N^2)$	External-decoupled	Structured memory traversal; fewer params than LSTM	Most computationally expensive; limited domain evidence
[26]	Multi-head self-attention across fixed memory slots	Attention-pooled (slot-slot)	Relational slot array	Moderate-high; $O(S^2)$	Attention-mediated	Inter-slot relational reasoning; sequential tasks	Large memory footprint for complex tasks
[27]	Gated entity-slot updates keyed by content	Key-based content addressing	Discrete entity-keyed slots	Moderate; $O(E \cdot d)$	Attention-mediated	Explicit entity tracking; dynamic world-state	Less general-purpose than NTM/DNC

Note: NTM complexity $O(N \cdot M)$ refers to memory matrix size N and head dimensionality M ; DNC complexity $O(N^2)$ arises from the temporal linkage matrix. Relative cost depends on the specific values of N and M chosen at deployment.

4. Discussion

This section synthesises key insights from the reviewed studies to identify prevailing trends, strengths, and limitations in GRU-based forecasting models augmented with external memory. The findings consistently indicate that integrating memory-augmented components enables GRUs to better capture long-range dependencies, addressing one of the core limitations of standard recurrent architectures. A key explanation for this improved performance is the model's ability to prioritise and recall critical time steps through structured memory access rather than relying solely on compressed hidden states.

As summarised in [Table 2](#), the five architecture families differ substantially in how they achieve this capability. NTM and DNC store information in flat external matrices accessed via differentiable operations, enabling the GRU controller to learn algorithmic retrieval patterns. For example, DNC-based GRUs can traverse temporally linked memory locations, while NTMs provide content- and location-based addressing that mimics algorithmic processing. These augmentations help GRU models maintain stability and avoid vanishing gradients over long sequences. In contrast, RelMem and EntMem partition memory into relational or entity-specific slots, introducing structural inductive biases that are particularly effective when loss dynamics involve interactions among multiple entities, such as co-stored products, batch handling sequences, or cascading supplier disruptions. Cell-integrated designs (GRU-MB, MGRU, MMU) achieve comparable long-range retention at substantially lower computational cost by embedding memory directly within the GRU cell, making them the most deployment-friendly option for resource-constrained food supply chain environments.

Although memory-augmented architectures significantly enhance the GRU's ability to model long-term dependencies, the inclusion of memory blocks or modules also introduces considerable computational overhead and increased model complexity. Training these architectures can be more challenging due to memory addressing ambiguity and slower

convergence. These models additionally require larger datasets to avoid overfitting and fully leverage their enhanced capacity. The trade-off is therefore not simply accuracy versus simplicity, but a multi-dimensional balance across computational cost, GRU integration complexity, data requirements, and the type of dependency structure present in the target forecasting problem.

Some findings do not directly apply to GRU without modification, as certain memory-augmented methods were originally developed around LSTM or custom controllers. Nevertheless, the adaptation of such memory-enhancing mechanisms to GRU architectures, as demonstrated across [Table 1](#), has shown consistently promising results. Techniques such as attention mechanisms and hybrid frameworks further enhance the model's ability to learn complex temporal patterns and non-linear relationships, as evidenced by the PER-DDPG framework [\[6\]](#) and the MemDNRI architecture [\[28\]](#). These insights carry important implications for policymakers and food supply chain managers who require accurate, interpretable, long-term forecasts for effective planning and intervention.

5. Research Gaps and Future Directions

Despite the progress made in enhancing GRU-based models with memory-augmented structures, several critical research gaps remain that limit the full potential of this advancement. One key limitation is that existing studies examine memory mechanisms in isolation, leaving unexplored the synergies that could arise from combining complementary designs within a unified MAGRU framework. This is particularly consequential for food loss forecasting, where long-term degradation processes, rare shock events, and complex multi-entity supply chain dependencies may each benefit from a different memory inductive bias. A promising direction is a hierarchical MAGRU design in which a lightweight cell-integrated module, such as MMU or GRU-MB, handles routine long-range retention, while a slot-based RelMem or EntMem layer activates selectively for multi-entity dependency modelling. Such a design would preserve relational expressiveness while containing computational cost, making it more viable for deployment across the food supply chain. Without exploring such combinations, model designs risk remaining suboptimal by overlooking cross-component interactions that could yield more accurate and resilient forecasting.

A second important gap lies in the limited real-world applicability of models proposed in current literature. Many studies rely on controlled datasets or synthetic benchmarks that fail to capture the irregularities, noise, and missing values typical of food supply chain data. Critically, food supply chain time-series present a distinct combination of properties that standard sequential benchmarks do not replicate: compound seasonality arising from overlapping daily, weekly, and harvest-cycle patterns; irregular sampling intervals from unevenly spaced cold-chain sensor readings; structured missingness driven by operational gaps rather than random failure; and rare but high-magnitude shock events such as spoilage spikes caused by equipment failure or weather disruption. It therefore remains unclear how well existing MAGRU models generalise to operational settings where these conditions co-occur. This disconnect highlights the urgent need for comprehensive evaluations on large-scale, diverse, and realistic datasets drawn directly from food logistics, inventory tracking, and climate-influenced agricultural operations, with evaluation protocols that explicitly stress-test model robustness under each of these conditions.

Looking forward, future research should pursue four concrete directions. First, regarding attention integration, future work should move beyond generic attention-augmented MAGRUs and investigate architectures that employ sparse cross-temporal attention for efficient long-range alignment and cross-entity attention for modelling interactions among co-stored

products, batch handling sequences, and cascading supplier disruptions. Such targeted attention designs would enhance interpretability by exposing which time steps and entities most influence each forecast. Second, regarding lightweight memory design, achieving deployability in resource-constrained environments, such as farms, distribution centres, and developing regions, requires more than architectural simplicity. Future work should explore structured sparsity in memory access, quantisation-aware training for memory addressing modules, and knowledge distillation from larger MAGRU models into compact cell-integrated variants, providing a principled pathway from research prototypes to field deployment. Third, regarding standardised benchmarks, the field lacks publicly available datasets tailored to food loss forecasting. Future benchmark construction should include multivariate series capturing temperature, humidity, batch identifiers, and handling event logs, with standardised train-test splits that preserve temporal order and representative shock-event frequencies. This would enable consistent model comparison and accelerate progress toward generalisable solutions. Fourth, regarding interpretability, memory read-weight visualisation and slot activation tracing, mechanisms already present in NTM and RelMem architectures respectively, remain underexplored as tools for generating actionable insights. Future work should develop standardised interpretability protocols that allow supply chain managers to trace a forecast back to specific historical events stored in memory, transforming MAGRU outputs from opaque predictions into transparent decision support.

6. Conclusion

This paper has explored the potential of enhancing GRU-based architectures for food loss forecasting through memory augmentation, reviewing five families of neural memory architecture, NTM, DNC, RelMem, EntMem, and MMU/GRU-MB, and organising their design choices, reported trade-offs, and cross-domain evidence into a unified framework. Memory-Augmented GRU (MAGRU) architectures introduce dedicated mechanisms for retaining long-range temporal information that standard recurrent models tend to forget, enabling models to revisit and reuse relevant historical context more effectively. Existing studies consistently suggest that MAGRU variants outperform baseline GRUs in tasks requiring temporal reasoning over longer horizons, indicating strong suitability for the complex forecasting demands of food supply chains.

The comparative analysis across [Table 1](#) and [Table 2](#) highlights that no single architecture dominates across all dimensions. Cell-integrated designs (GRU-MB, MMU) offer the best cost-deployability trade-off; NTM and DNC excel at algorithmic and structured sequential reasoning respectively, supported by cross-domain validation histories that suggest meaningful transfer potential for food loss forecasting; and slot-based designs (RelMem, EntMem) uniquely support relational and entity-aware dependency modelling. The most promising direction for MAGRU design in food loss forecasting is therefore a hybrid approach that selectively combines these complementary strengths, integrating lightweight cell-integrated memory for baseline long-range retention with targeted relational or NTM-style modules for event-driven and multi-entity patterns, while coupling memory with attention mechanisms for improved interpretability.

Although their individual strengths are evident, combining these techniques within a unified MAGRU framework has not yet been fully explored, and validated evaluations on realistic food supply chain datasets remain absent from the literature. Future work that bridges these methodological gaps, developing hybrid MAGRUs, establishing domain-specific benchmarks, and testing on operational supply chain data, will be essential to translating the

theoretical promise of memory-augmented forecasting into actionable, trustworthy tools for sustainable food system management.

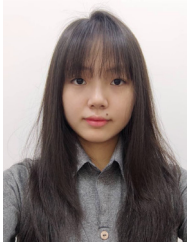
Acknowledgement

This work is funded by Tunku Abdul Rahman University of Management and Technology (TAR UMT) through the Internal Research Grant: UC/I/G2025-00164.

References

- [1] G. Gallo, and P. Navarro-Gambín, “Addressing food loss and waste in food systems: A critical review of definitions, drivers, and EU policies and strategies,” *Encyclopedia of Agriculture and Food Systems*, 3rd Ed., pp.177-192, Jan. 2026. [Article\(CrossRefLink\)](#)
- [2] O. Olabode, N. Kumar, and D. De, “Food loss and waste management in the retail food supply chain: Methods and framework to achieve environmental sustainability,” *Journal of Environmental Management*, vol.387, Jul. 2025. [Article\(CrossRefLink\)](#)
- [3] S. Negi, L.-S. Fan, H. Kim, T. Hidaka, A. Rani, and S.-Y. Pan, “Unlocking the potential of global greenhouse gas mitigation by reducing food loss and waste,” *Journal of Agriculture and Food Research*, vol.21, Jun. 2025. [Article\(CrossRefLink\)](#)
- [4] U. N. E. Programme, Food Waste Index Report 2024. Think Eat Save: Tracking Progress to Halve Global Food Waste, 2024. [Article\(CrossRefLink\)](#)
- [5] C. Muriana, “A focus on the state of the art of food waste/losses issue and suggestions for future researches,” *Waste Management*, vol.68, pp.557-570, Oct. 2017. [Article\(CrossRefLink\)](#)
- [6] X. Kong et al., “Real-time pricing method for VPP demand response based on PER-DDPG algorithm,” *Energy*, vol.271, May 2023. [Article\(CrossRefLink\)](#)
- [7] J. Xiao et al., “Interpretable time-series neural turing machine for prognostic prediction of patients with type 2 diabetes in physician-pharmacist collaborative clinics,” *International Journal of Medical Informatics*, vol.195, Mar. 2025. [Article\(CrossRefLink\)](#)
- [8] A. Khan et al., “DeepProcess: Supporting Business Process Execution Using a MANN-Based Recommender System,” in *Proc. of the 19th International Conference on Service-Oriented Computing (ICSOC 2021)*, Lecture Notes in Computer Science, vol.13121, pp.19-33, Springer, Cham, 2021. [Article\(CrossRefLink\)](#)
- [9] S. Khadka, J. J. Chung, and K. Tumer, “Evolving memory-augmented neural architecture for deep memory problems,” in *Proc. of the Genetic and Evolutionary Computation Conference (GECCO '17)*, pp.441-448, Jul. 2017. [Article\(CrossRefLink\)](#)
- [10] S. Khadka, J. J. Chung, and K. Tumer, “Memory-augmented multi-robot teams that learn to adapt,” in *Proc. of the 2017 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*, pp.128-134, Dec. 2017. [Article\(CrossRefLink\)](#)
- [11] M. Yang, M. Li, and G. Li, “On memory-augmented gated recurrent unit network,” *International Journal of Forecasting*, vol.41, no.2, pp.844-858, Apr.-Jun. 2025. [Article\(CrossRefLink\)](#)
- [12] S. Khadka, J. J. Chung, and K. Tumer, “Neuroevolution of a Modular Memory-Augmented Neural Network for Deep Memory Problems,” *Evolutionary Computation*, vol.27, no.4, pp.639-664, Dec. 2019. [Article\(CrossRefLink\)](#)
- [13] A. Graves, G. Wayne, and I. Danihelka, “Neural Turing Machines,” *arXiv preprint arXiv:1410.5401*, Dec. 2014. [Article\(CrossRefLink\)](#)
- [14] S. M. Faradonbe, F. Safi-Esfahani, and M. Karimian-kelishadrokh, “A Review on Neural Turing Machine (NTM),” *SN Computer Science*, vol.1, no.6, Oct. 2020. [Article\(CrossRefLink\)](#)
- [15] C. Panggabean, C. Venkatachalam, P. Shah, S. John, P. R. Devi, and S. Venkatachalam, “Intelligent DoS and DDoS Detection: A Hybrid GRU-NTM Approach to Network Security,” in *Proc. of the 2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*, pp.658-665, Sep. 2024. [Article\(CrossRefLink\)](#)

- [16] D. Shan, J. Zhu, and Y. Luo, "SANTM: A Sparse Access Neural Turing Machine with local multi-head self-attention for long-term memorization," *Applied Soft Computing*, vol.181, Sep. 2025. [Article\(CrossRefLink\)](#)
- [17] C. Gulcehre, S. Chandar, K. Cho, and Y. Bengio, "Dynamic Neural Turing Machine with Soft and Hard Addressing Schemes," *arXiv preprint arXiv:1607.00036*, Mar. 2017. [Article\(CrossRefLink\)](#)
- [18] C. Gulcehre, S. Chandar, K. Cho, and Y. Bengio, "Dynamic Neural Turing Machine with Continuous and Discrete Addressing Schemes," *Neural Computation*, vol.30, no.4, pp.857-884, Apr. 2018. [Article\(CrossRefLink\)](#)
- [19] D. Donahue, Y. Meng, and A. Rumshisky, "Memory-Augmented Recurrent Networks for Dialogue Coherence," *arXiv preprint arXiv:1910.10487*, Oct. 2019. [Article\(CrossRefLink\)](#)
- [20] H. Xiong, Z. He, X. Hu, and H. Wu, "Multi-Channel Encoder for Neural Machine Translation," in *Proc. of the AAAI Conference on Artificial Intelligence*, vol.32, no.1, Apr. 2018. [Article\(CrossRefLink\)](#)
- [21] G. D'Agostino, A. Falcon, G. Brajnik, C. Tasso, O. Lanz, and G. Serra, "Estimating the Remaining Useful Life via Neural Sequence Models: a Comparative Study," in *Proc. of the 2nd Italian Workshop on Artificial Intelligence and Applications for Business and Industries (AIABI)*, vol.3463, pp.1-8, 2022. [Article\(CrossRefLink\)](#)
- [22] A. Falcon, G. D'Agostino, O. Lanz, G. Brajnik, C. Tasso, and G. Serra, "Neural Turing Machines for the Remaining Useful Life estimation problem," *Computers in Industry*, vol.143, no.C, Dec. 2022. [Article\(CrossRefLink\)](#)
- [23] J. Castellini, "Learning Numeracy: Binary Arithmetic with Neural Turing Machines," *arXiv preprint arXiv:1904.02478*, Sep. 2019. [Article\(CrossRefLink\)](#)
- [24] X. Chen and W. Fan, "Testing technology for tensile properties of metal materials based on deep learning model," *Frontiers in Neurorobotics*, vol.16, 2022. [Article\(CrossRefLink\)](#)
- [25] A. Graves et al., "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol.538, no.7626, pp.471-476, Oct. 2016. [Article\(CrossRefLink\)](#)
- [26] A. Santoro et al., "Relational recurrent neural networks," in *Proc. of Advances in Neural Information Processing Systems*, vol.31, pp.7310-7321, 2018. [Article\(CrossRefLink\)](#)
- [27] M. Henaff, J. Weston, A. Szlam, A. Bordes, and Y. LeCun, "Tracking the World State with Recurrent Entity Networks," *arXiv preprint arXiv:1612.03969*, May 2017. [Article\(CrossRefLink\)](#)
- [28] D. Gong, Z. Zhang, J. Q. Shi, and A. Van Den Hengel, "Memory-augmented Dynamic Neural Relational Inference," in *Proc. of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.11823-11832, Oct. 2021. [Article\(CrossRefLink\)](#)



Ru Poh Tan received her Bachelor's Degree in Data Science (Honours) from Tunku Abdul Rahman University of Management and Technology (TAR UMT) in 2025. Currently, she is a Master's degree candidate in Computer Science, further deepening her expertise in the field. She also works as an AI Developer in Kuala Lumpur. Her current research interests lie at the intersection of artificial intelligence and real-world applications, with a strong focus on deep learning, natural language processing (NLP), time series forecasting, and memory-augmented neural architectures.



Siew Mooi Lim holds a Ph.D in Intelligent Computing from Universiti Putra Malaysia. She is currently an Assistant Professor at the Faculty of Computer Science and Information Technology, TAR UMT. She has led and participated in multiple research projects and has published extensively in her areas of expertise. She has also served as a mentor for student participation in national and international competitions, as well as an external examiner for Ph.D. scholars. She is recognized for her academic leadership, having received awards that reflect her commitment to advancing research and guiding future scholars. Her research interests include machine learning, predictive modelling, natural language processing, LLMs, and generative AI.



Kuan Yew Leong specializes in Wearable Technology, A.I., Computer Vision, and Deep Learning, with 20 years of combined academic and industry experience. He holds a Ph.D from Monash University (Australia), where he was nominated for the prestigious Mollie Holman Award. Based in Japan as a permanent resident, he works as a lead researcher at an A.I. corporation in Kyoto and serves as an adjunct professor at two universities. His research focuses on optimizing deep learning models, LLMs, RAG, and computer vision tasks like re-identification, image classification, and object detection.



Jun Kit Chaw received the B.Eng. degree in computer engineering and the Ph.D. degree in electrical engineering from Universiti Teknologi Malaysia (UTM), Johor, Malaysia. He is currently a Research Fellow with the Institute of Visual Informatics (IVI), UKM. Previously, he was a Senior Lecturer with the Faculty of Computing and Information Technology (FOCS), Tunku Abdul Rahman University of Management and Technology (TAR UMT). He has actively collaborated with industry partners, such as Advantech and HILTI Asia IT Services. His industry projects include optimization of manufacturing processes and computer vision applications. His primary research interests include computer vision and deep learning.



Siaw Hong Liew is currently working as a senior lecturer at the Faculty of Computer Science and Information Technology (FCSIT), Universiti Malaysia Sarawak (UNIMAS). She graduated with a Bachelor of Computer Science majoring in Artificial Intelligence in the year 2014 from Universiti Teknikal Malaysia Melaka (UTeM). She pursued her studies in the same university and completed her Master of Science in Information and Communication Technology in the year 2016. In 2021, she was awarded the PhD from UTeM, with a dissertation titled "Distraction Descriptor for Brainprint Authentication Modelling using Probability-based Incremental Fuzzy-Rough Nearest Neighbour Technique". Her research interests involve machine learning, deep learning and signal processing.



Shee Chia Lee received her Master's degree in Applied Statistics from the University of Malaya. She is currently pursuing a Ph.D. in Mathematical Sciences and is a lecturer at the Faculty of Computer Science and Information Technology, Tunku Abdul Rahman University of Management and Technology. Her research interests include statistics, forecasting, and predictive modelling.