INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY: APPLIED BUSINESS AND EDUCATION RESEARCH
2025, Vol. 6, No. 11, 5805 – 5812
http://dx.doi.org/10.11594/ijmaber.06.11.34

Research Article

Cross-Attention Multimodal Transformer for Calibrated Binary Time-Series Forecasting of Rural Public Services

Daryl John C. Ragadio*

Information Technology Department, President Ramon Magsaysay State University, Iba, Zambales, Philippines

Article history:
Submission 03 October 2025
Revised 31 October 2025
Accepted 23 November 2025

**Corresponding author:*
E-mail:
djragadio@prmsu.edu.ph

ABSTRACT

Good governance, evidence-based planning, and sustainable rural development all rely on correct rural public service performance predictions. This study presents a Cross-Attention Multimodal Transformer developed for binary time-series classification of service conditions in the areas of agriculture, health, and environment at the level of the local government unit (LGU). Using bidirectional cross-attention layers, the model mixes several temporal signals so that healthcare and agriculture-environment streams can interact with one another. Using weighted uncertainty and calibration-awareness in a loss function helps to guarantee that the confidence scores are properly calibrated. With AUCs of 83.00% (agriculture), 79.40% (environment), and 63.90% (healthcare), which is lower, experimental results on a rural public service dataset indicate great discriminative and calibration performance. With 61.50%, 23.10%, and 18.90% respectively, the Brier scores suggest that the forecasts for health care and the environment are well calibrated. These findings suggest that cross-attention multimodal transformers may be quite useful in producing precise binary predictions of rural service results. At the LGU level, this would enable data-driven decision-making support.

Keywords: *Multimodal transformer, Cross-attention mechanism, Binary time-series classification, Rural public services, Calibration*

Introduction

Local government's early risk identification, budget distribution, and sustainable development efforts all depend on predicting rural public service performance, which is why it is so important. Sectors like agriculture, healthcare, and environment work in very interconnected systems at the local government unit (LGU) level, where changes in one field usually have an impact on the others (Park, 2024). But current predictive models sometimes separate these industries, therefore

How to cite:
Ragadio, D. J. C. (2025). Cross-Attention Multimodal Transformer for Calibrated Binary Time-Series Forecasting of Rural Public Services. *International Journal of Multidisciplinary: Applied Business and Education Research*. *6*(11), 5805 – 5812. doi: 10.11594/ijmaber.06.11.34

reducing their capacity to include cross-domain dependencies (Kim et al., 2024; Mou et al., 2025).

Recent developments in multimodal transformers have enabled the integration of tabular, sensor-based, and temporal heterogeneous data sources via cross-attention techniques that dynamically represent interactions between modalities (Yuan & Zhao, 2025; Su et al., 2025). Such designs have shown themselves to be useful in areas like medical forecasting, weather prediction, and finance (Jia et al., 2024). Still, their application in rural service analytics is mostly untested.

For binary time-series classification of rural service results, this study presents a Cross-Attention Multimodal Transformer (CAMT). Rather than forecasting continuous values, the model predicts binary outcomes by calculating the likelihood of service states using probabilistic calibration. The main contributions are:

1. For inter-modality learning, a cross-attention transformer design combines healthcare time-series data with agriculture-environment signals.
2. A calibration-aware, uncertainty-weighted loss function guarantees both robustness and trustworthy probability values.
3. A thorough empirical study of a rural public service dataset mirroring LGU-level dynamics.

Review of Related Literature

By combining prior findings to clarify their contributions and set the theoretical basis for the current study, this section offers a critical examination of pertinent papers.

Multimodal Learning and Data Fusion

Multimodal learning combines several kinds of data including tables, written text, and sensor-derived information (Zhao et al., 2024; Al-Zoghby et al., 2025; Adam et al., 2025). Research point out how crucial cross-modal fusion is for enhancing prediction accuracy (Jing et al., 2024; Kalisetty & Lakkarasu, 2024). Shao (2024) developed a block transformer that successfully fuses static and dynamic modalities for rural service environments with diverse data.

Time-series forecasting uses transformer architectures

Transformers capture long-range dependencies and have advanced time-series forecasting. Abdullahi et al. (2025) and Siebra et al. (2024) looked at different types of transformers for long-term forecasting (Siebra et al., 2024). In multi-horizon tasks, Informer (Cui et al., 2024), (Zhou et al., 2021) did better than RNNs and CNNs while keeping accuracy and lowering complexity.

Modeling and Calibration of Uncertainty

Reliable probability estimates are essential for forecasting. Cui et al. (2022) and Jenses et al. (2022) looked at probabilistic forecasting using ensemble methods and quantile regression. Turki et al. (2025) and Xiao et al. (2024) used calibration-aware losses in public sector settings.

Public Sector Applications and Rural Services

Few studies directly examine rural service projection, but related literature show transportable techniques. Applying multimodal transformers to economic and healthcare data, Emami et al. (2024) and Bouatmane et al. (2025). Deep learning was used by Nikhil et al. (2024) and Saravanan et al. (2024) to predict agricultural yield. Chaturvedi (2024) and Postiglione et al. (2024) used temporal embeddings to forecast healthcare demand.

Interpretable and decision support

For AI-driven governance, interpretability is absolutely necessary. Guo et al. (2022) and Ruan & Zhang (2024) created attention mechanisms. Kotipalli (2024) and Chefer et al. (2021) showed attention maps and calibration curves. These strategies help local government units make decisions.

To encapsulate, previous research show that multimodal fusion and transformer designs have the ability to improve the accuracy of forecasts. They also show how important it is to calibrate uncertainty in order to get trustworthy predictions. Rural service for time-series binary classification is still underused, but transferable techniques from agriculture, economics, and healthcare show some interesting uses. These observations provide the basis for

this research that aims to improve rural area prediction by means of models that are precise, dependable, and easily understood.

Methods

Dataset Description

Daily region-level observations covering agriculture, healthcare, and environmental indicators comprise the Rural Public Service Dataset. Agriculture and Environment includes temperature, water levels, air quality index (AQI), and soil moisture. Blood pressure, heart rate, oxygen saturation, and telemedicine appointments are some healthcare aspects. Seasonal cycles are shown by temporal characteristics like sine–cosine day-of-year encodings. Three binary variables such as agriculture, healthcare, and environmental service states represent targets (1 = normal/stable, 0 = at-risk). Based on date coherence, data were divided geographically and chronologically into 70% training, 15% validation, and 15% testing sets. Standard scaling was used to normalize continuous characteristics.

Sequence Design

Every sample included a 14-day sequence window that caught brief temporal correlations for every region. Agriculture, environment and healthcare streams were coupled with contextual identifiers for region and date for every sequence.

Model Framework

Several important elements make up the suggested Cross-Attention Multimodal Transformer (CAMT). First, Modality-Specific Encoders are made as linear projection layers that turn agriculture, environment and healthcare feature vectors into different latent representations for each modality. Cross-Attention Blocks then use two multi-head attention modules to let information flow both ways. Agriculture pays attention to healthcare features (A→H), and healthcare pays attention to agriculture–environment features (H→A). This mechanism efficiently models inter-modal dependencies vital for thorough examination of rural services.

Following this, learned embeddings for region_id and date_id was used to apply Contextual Embeddings, therefore giving every sequence temporal and spatial grounding. After that, an Attention Pooling layer combines the multimodal sequence representations into a single contextual vector. Three linear output layers matching agriculture, healthcare, and environmental chores define Prediction Heads. Every outcome generates a logit value that is sigmoid activated to provide probabilistic binary judgments.

Multi-task Loss with Calibration- Awareness

The model maximizes a mix of three main elements. The Focal Loss is first used on every binary output to tackle class imbalance by highlighting samples that are more difficult to classify. Second, Uncertainty-Weighted Balancing uses learnable log-variance parameters to change the contribution of each task during training dynamically. Ultimately, Calibration Regularization adds a binary Kullback–Leibler (KL) divergence term to lower the difference between the projected probability means and the observed class priors, hence enhancing the dependability of the probabilistic predictions of the model.

Evaluations and Training

For 50 epochs, training used the Adam optimizer (learning rate = 0.001) with a cosine annealing scheduler. Early stopping tracked validation AUC. To evaluate discrimination and calibration performance, metrics include AUC, F1-score, Precision, Recall, and Brier Score.

Result and Discussion

As shown in Table 1, the per-domain assessment finds significant variances in model performance throughout rural public service industries. With an AUC of 83.00% and an F1-score of 79.50%, agriculture outperformed the other two sectors overall. The model's low false-positive rate together with its high recall (81.40%) and accuracy (77.80%) suggest that it regularly found pertinent agricultural circumstances. This suggests that the cross-attention transformer gained from structured, seasonal regularities in the data and adequately

captured the temporal and spatial patterns in the agricultural data.

On the other hand, the healthcare work performed the worst with an AUC of 63.90% and a F1-score of 46.60%.

The model appears to have taken a conservative prediction approach because it has quite good accuracy (50.00%), but rather low recall (43.60%). This suggests it valued precise classifications over some beneficial occurrences in healthcare. This pattern shows how difficult healthcare data may be. Unlike signals discovered in agricultural or environmental settings, healthcare data shows great diversity across patients as well as noise and trends that are less reliable throughout time.

With an AUC of 79.40% and an F1-score of 75.50% in the environmental field, the model showed a somewhat better performance. The model's accuracy was 88.10%, however its recall was a little bit lower at 66.10%. This indicates that, while it successfully lowered the rate of false positives, it tended to be a little cautious in spotting all positive cases. This trade-off means that, even if the model uses feature enhancement to make it more sensitive, things like the air quality index (AQI) and temperature in the environment show consistent and predictable patterns.

At the macro level, the model's average AUC of 75.40% and F1-score of 67.20% imply good generalization across different industries. Although domain-specific optimization is still crucial for raising predictive performance in medicine, these findings highlight the flexibility of the transformer to several input forms.

Table 1. Per-task Performance Result

Domain	Metrics				
	AUC (%)	**F1-Score (%)**	**Precision (%)**	**Recall (%)**	**Brier (%)**
Agriculture	83.00	79.50	77.80	81.40	61.50
Health	63.90	46.60	50.00	43.60	23.10
Environment	79.40	75.50	88.10	66.10	18.90

The Brier Score shows the average squared difference between what we expect to happen and what really happens. This tells us roughly how accurately a model forecasts results. A low score, particularly in healthcare (23.10%) and the environment (18.90%), indicates that the probability forecasts are reliable and correct. Three primary causes account for this: (1) calibration-aware regularization keeps predictions that are too confident in check; (2) uncertainty-weighted multi-task learning changes the contribution of various tasks using learnable parameters; and (3) environmental variables, like temperature, water levels, and air quality index, usually have regular seasonal patterns. These low Brier Scores hence suggest good calibration rather than underperformance, which bolsters the notion that the model's predicted probabilities almost exactly correspond to observed outcomes even if discrimination measures are just acceptable.

Analysis of ROC Curves

The Receiver Operating Characteristic (ROC) curves for the three areas are shown in Figure 1. The agriculture field best distinguishes between positive and negative cases, as shown by an AUC of 83%, hence has the highest discriminative ability. With a marginally lower true positive rate at medium cutoffs, the environment domain has an AUC of 79%, indicating strong discriminative ability. By contrast, the 64% AUC in the healthcare field shows little separability, which is in line with its lower recall rate and the earlier noted data variation.

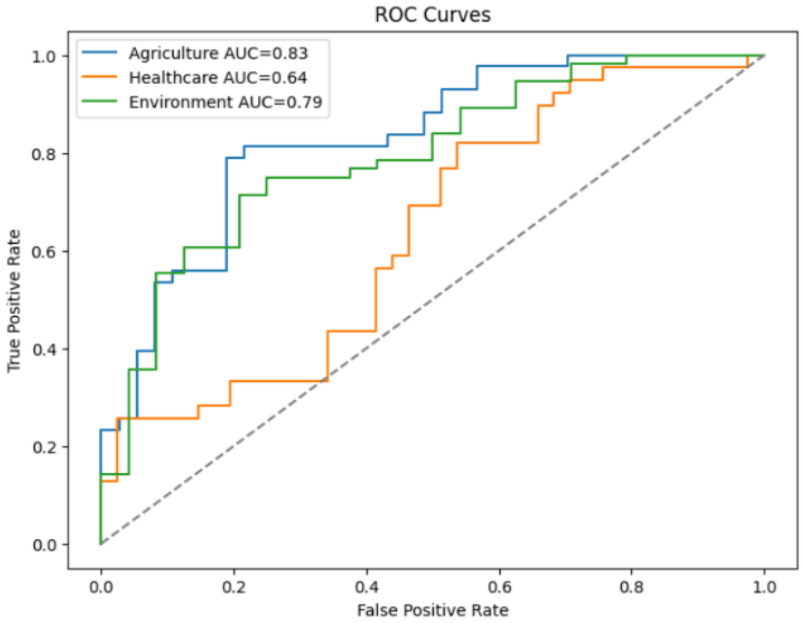

Figure 1. ROC Curve Analysis

Generally, the ROC analysis shows that the Cross-Attention Multimodal Transformer (CAMT) works quite well for agriculture and environment classification tasks, but healthcare is still a more difficult and data-constrained subject. This imbalance emphasizes the need of domain-specific learning techniques, feature representativeness, and data quality in multimodal binary time-series classification uses.

Study of Calibration Curves

Figure 2 shows the calibration graphs for agriculture, healthcare, and the environment. A model that is perfectly calibrated would line up with the diagonal reference line, where the predictions would match the actual results. The arrangement is almost flawless in the environmental field, demonstrating that at all probability levels the confidence intervals are trustworthy.

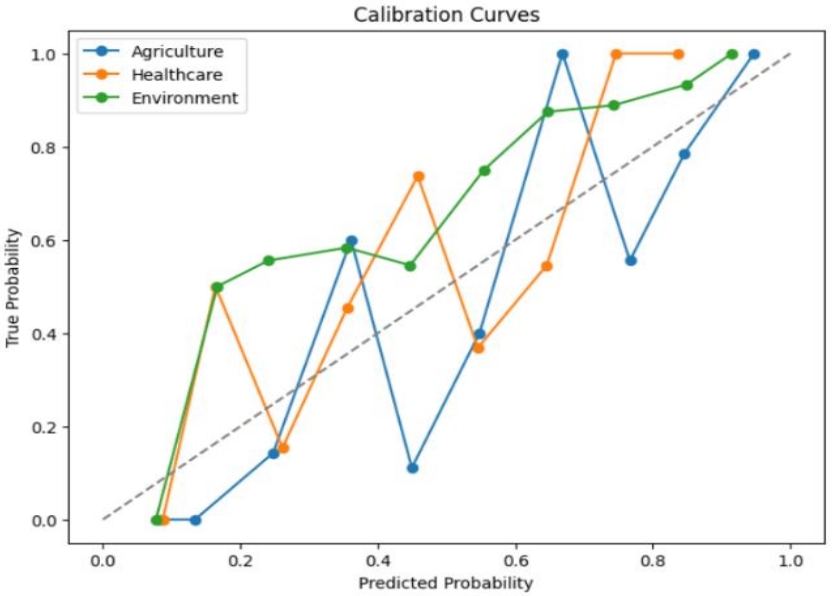

Figure 2. Calibration Curve Analysis

Although the agriculture graph is generally accurate, it somewhat exaggerates probabilities between 0.4 and 0.6 and somewhat underestimates values above 0.8. The healthcare graph, on the other hand, shows more underestimation between 0.3 and 0.6, therefore it varies more from expected results and highlights the model's conservative prediction style together with a reduced recall.

Taken collectively, all these calibration trends point to the fact that applying calibration-aware loss enhances dependability throughout a range of domains even in complex ones.

Conclusion

The findings reveal that the suggested Cross-Attentive Multimodal Transformer is capable of grasping relationships among several regions and within particular areas across time. Using techniques like cross-attention blending, uncertainty adjustment, and calibration smoothing, this model produces excellent and consistent results.

Moderate reliability in healthcare and trustworthy environmental forecasts illustrate how calibration methods might increase probability trustworthiness even if binary classification accuracy varies.

These findings demonstrate that for rural public service analysis, properly tuned multimodal learning is achievable. They also emphasize how domain adaptation, more modalities, and more temporal coverage could help with future improvements.

Acknowledgement

The author sincerely extends his deepest gratitude to President Ramon Magsaysay State University for its unwavering support and invaluable guidance in the successful completion of this research.

References

Abdullahi, S., Danyaro, K. U., Zakari, A., Aziz, I. A., Zawawi, N. A. W. A., & Adamu, S. (2025). Time-series large language models: A systematic review of state-of-the-art. *IEEE Access*.

Adam, M., Albaseer, A., Baroudi, U., & Abdallah, M. (2025). Survey of Multimodal Federated Learning: Exploring Data Integration, Challenges, and Future Directions. *IEEE Open Journal of the Communications Society*.

Al-Zoghby, A. M., Al-Awadly, E. M. K., Ebada, A. I., & Awad, W. A. (2025). Overview of Multimodal Machine Learning. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *24*(1), 1-20.

Bouatmane, A., Daaif, A., Bousselham, A., Bouihi, B., & Bouattane, O. (2025). A Multimodal Deep Learning Model Integrating CNN and Transformer for Predicting Chemotherapy-Induced Cardiotoxicity. *IEEE Access*.

Chaturvedi, R. (2024, May). Temporal knowledge graph extraction and modeling across multiple documents for health risk prediction. In *Companion Proceedings of the ACM Web Conference 2024* (pp. 1182-1185).

Chefer, H., Gur, S., & Wolf, L. (2021). Transformer interpretability beyond attention visualization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 782-791).

Cui, W., Wan, C., & Song, Y. (2022). Ensemble deep learning-based non-crossing quantile regression for nonparametric probabilistic forecasting of wind power generation. *IEEE Transactions on Power Systems*, *38*(4), 3163-3178.

Cui, Y., Li, Z., Wang, Y., Dong, D., Gu, C., Lou, X., & Zhang, P. (2024). Informer model with season-aware block for efficient long-term power time series forecasting. *Computers and Electrical Engineering*, *119*, 109492.

Emami Gohari, H., Dang, X. H., Shah, S. Y., & Zerfos, P. (2024, November). Modality-aware Transformer for Financial Time series Forecasting. In *Proceedings of the 5th ACM International Conference on AI in Finance* (pp. 677-685).

Guo, M. H., Xu, T. X., Liu, J. J., Liu, Z. N., Jiang, P. T., Mu, T. J., ... & Hu, S. M. (2022). Attention mechanisms in computer vision: A survey. *Computational visual media*, *8*(3), 331-368.

Jensen, V., Bianchi, F. M., & Anfinsen, S. N. (2022). Ensemble conformalized quantile regression for probabilistic time series forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(7), 9014-9025.

Jia, F., Wang, K., Zheng, Y., Cao, D., & Liu, Y. (2024, March). Gpt4mts: Prompt-based large language model for multimodal time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 21, pp. 23343-23351).

Jing, T., Chen, S., Navarro-Alarcon, D., Chu, Y., & Li, M. (2024). SolarFusionNet: Enhanced Solar Irradiance Forecasting via Automated Multi-Modal Feature Selection and Cross-Modal Fusion. *IEEE Transactions on Sustainable Energy*.

Kalisetty, S., & Lakkarasu, P. (2024). Deep Learning Frameworks for Multi-Modal Data Fusion in Retail Supply Chains: Enhancing Forecast Accuracy and Agility. *American Journal of Analytics and Artificial Intelligence (ajaai) with ISSN 3067-283X*, *2*(1).

Kim, K., Tsai, H., Sen, R., Das, A., Zhou, Z., Tanpure, A., ... & Yu, R. (2024). Multi-modal forecaster: Jointly predicting time series and textual data. *arXiv preprint arXiv:2411.06735*.

Kotipalli, B. (2024). The Role of Attention Mechanisms in Enhancing Transparency and Interpretability of Neural Network Models in Explainable AI.

Mou, S., Xue, Q., Chen, J., Takiguchi, T., & Ariki, Y. (2025). MM-iTransformer: A Multimodal Approach to Economic Time Series Forecasting with Textual Data. *Applied Sciences*, *15*(3), 1241.

Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine learning-based crop yield prediction in south india: performance analysis of various models. *Computers*, *13*(6), 137.

Park, S. (2024). Multimodal Block Transformer for Multimodal Time Series Forecasting. In *Annual Conference of KIPS* (pp. 636-639). Korea Information Processing Society.

Postiglione, M., Bean, D., Kraljevic, Z., Dobson, R. J., & Moscato, V. (2024). Predicting future disorders via temporal knowledge graphs and medical ontologies. *IEEE Journal of Biomedical and Health Informatics*, *28*(7), 4238-4248.

Ruan, T., & Zhang, S. (2024). Towards understanding how attention mechanism works in deep learning. *arXiv preprint arXiv:2412.18288*.

Saravanan, K. S., & Bhagavathiappan, V. (2024). Prediction of crop yield in India using machine learning and hybrid deep learning models. *Acta Geophysica*, *72*(6), 4613-4632.

Shao, M., Li, D., Hong, S., Qi, J., & Sun, H. (2024). IQFormer: A novel transformer-based model with multi-modality fusion for automatic modulation recognition. *IEEE Transactions on Cognitive Communications and Networking*.

Siebra, C. A., Kurpicz-Briki, M., & Wac, K. (2024). Transformers in health: a systematic review on architectures for longitudinal data analysis. *Artificial Intelligence Review*, *57*(2), 32.

Su, L., Zuo, X., Li, R., Wang, X., Zhao, H., & Huang, B. (2025). A systematic review for transformer-based long-term series forecasting. *Artificial Intelligence Review*, *58*(3), 80.

Thundiyil, S., Picone, J., & McKenzie, S. Transformer Architectures in Time Series Analysis: A Review.

Turki, A., Alshabrawy, O., & Woo, W. L. (2025). Multimodal Deep Learning for Stage Classification of Head and Neck Cancer Using Masked Autoencoders and Vision Transformers with Attention-Based Fusion. *Cancers*, *17*(13), 2115.

Xiao, W., Wang, Z., Gan, L., Zhao, S., Li, Z., Lei, R., ... & Wu, F. (2024). A comprehensive survey of direct preference optimization: Datasets, theories, variants, and applications. *arXiv preprint arXiv:2410.15595*.

Yuan, Y., Li, Z., & Zhao, B. (2025). A survey of multimodal learning: Methods, applications, and future. *ACM Computing Surveys*, *57*(7), 1-34.

Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. *ACM computing surveys*, *56*(9), 1-36.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 12, pp. 11106-11115).