



# International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

## Cross-Domain Generalization of a Regime-Aware Forecasting Framework Across Financial, Meteorological, and Physiological Time Series

Pradnya Prakash Shelar<sup>\*1</sup>, Dr. Monika Bhatnagar<sup>2</sup>

<sup>1</sup> Research Scholar, Department of Computer Science & Engineering, Oriental University, Indore (M.P.).

E-mail: [pradnyak.bachhav@gmail.com](mailto:pradnyak.bachhav@gmail.com), ORCID: 0009-0009-9593-3455

<sup>2</sup> Supervisor, Department of Computer Science & Engineering, Oriental University, Indore (M.P.). E-mail: [monikabhatnagar@oriental.ac.in](mailto:monikabhatnagar@oriental.ac.in), ORCID: 0009-0002-8239-6565

### Abstract

The infrastructure style benchmarks are a necessary set, but they do not necessarily give a good idea of how an architecture will perform on data that has a very different statistical nature. This paper does just that here. We use a five-stage regime-aware forecasting framework proposed in an accompanying paper (Causal-Regime Tokeniser, a Wavelet-State-Space Dual-Path Predictor, evidence-weighted fusion, counterfactual cross-regime validation and drift aware continual adaptation) and apply it to three public datasets without any changes specific to the domains: a 5-year sample of daily closing prices for the S&P 500, a dataset of 3-year NOAA hourly temperature records, and the WESAD physiological stress dataset. It has the lowest MAE and RMSE across all three domains, the lowest CRPS and highest coverage nearer the nominal value in the probabilistic calibration task, and the highest F1 score on the physiological stress-classification task, when compared to the three baselines used in the original validation of the framework: a progressive-learning transformer, a Taylor-series/reinforcement-learning hybrid, and a kernel-based non-Markovian predictor. The largest coverage difference appears on the financial series, an increase of some 6 percentage points compared to the next highest baseline, which is related to the dynamics of the framework's conformal recalibration when abrupt volatility regimes occur. Combined, these findings suggest regime discovery, multi-scale forecasting and reliability-weighted fusion as a general tool for non-stationary structurally discontinuous data which is not tailored specifically to the infrastructure domain the framework was originally developed to address — a feature that would directly impact anyone with a forecasting system running on multiple statistically different streams of data.

**Keywords:** Cross-Domain Forecasting, Non-Stationary Time Series, Regime-Aware Forecasting, Uncertainty Calibration, Financial Time Series, Physiological Signal Forecasting.

## 1. Introduction

The process of most forecasting architectures is validated against data very similar to what they were designed for, leaving another question unanswered: is the benefit from real understanding of non-stationarity, or merely implicit or explicit tuning of the architecture to a certain type of data? While financial markets, meteorological systems, and physiological signals are all textbook examples of non-stationary domains, the event of regime shifts and the statistical signatures thereof vary considerably from one to the next. Macroeconomic news, geopolitical events, and clusters of volatility can make assumptions of a financial series become invalid within hours of the release [1-4]. The sudden, large deviations (storm front, heat wave) of meteorological series are difficult to capture in conventional seasonal models, in addition to strong periodic structure [5-8]. The variability of physiological signals, e.g. heart-rate variability, electrodermal activity, is mostly circadian and behavioural, and transitions rapidly between rest and acute stress with uncertainty, so the consequences of the wrong forecast goes beyond worse health monitoring [9, 10].

A companion paper introduces the five-stage regime-aware forecasting framework that this paper is based on: CaRT for differentiable regime discovery, WSDP for dual-path multi-scale forecasting, EviFuse for evidence-weighted fusion, CF-XValid for counterfactual stress testing, and DriftShield-ACL for drift-aware continual recalibration. This paper extensively demonstrates the validation of this framework on benchmark systems of infrastructure, electricity load, traffic flow, energy consumption, and demand driven by weather, on which the framework minimizes error and enhances calibration compared to existing baseline models for both ordinary and compound stresses [11-13]. However, this validation applies to a general class of series: the grid or sensor patterned ones, with whatever differences they share, which have a broadly similar statistical structure.

Whether the same architecture, no special modification of the domain, would transfer to significantly different non-stationary regimes is a different question, and is the subject of this paper.

### Contribution

This paper doesn't introduce any new architectural components. What it offers is an empirical test of the transfer: a five-stage framework that was developed on three datasets selected deliberately to be as different as possible from one another in both the domain it was developed in and other data—for financial, meteorological, and physiological—where the main test claims of the five-stage framework (accuracy gains, calibration reliability, robustness from regime-aware fusion) are therefore tested, and it does so by examining how well they hold up when presented with input that is not in the framework's original domain of development. This is not just a re-run of the original experiments on new data. The issue here is not re-validation, but generalization.

## 2. Related Work

**Financial time-series forecasting.** Price forecasting methods from the deep learning realm have increasingly come to the realization that non-stationarity is a real feature, and not just something to be ignored and swept under the rug. Liu et al. suggested that directly modeling the non-stationary signal in the attention mechanism is preferable as opposed to normalizing the signal out of the model, as preserving the non-stationary signal will increase forecasting accuracy [Liu et al., 2022]. The deep learning literature for financial price forecasting calls attention to the fact that few architectures are evaluated with regard to the robustness of their ability to adapt their architecture in the event of a regime change, rather than simply the average model accuracy on mostly stationary historical windows [Zhang et al., 2024]. Another cogent argument comes from recent article in CNN: In fact, predictive accuracy is improved when the model is designed for a domain with a regime change – in this case the S&P 500 — as opposed to simply copying a framework from a different domain that was designed for a different sequence space [Ranjan et al., 2025].

**Meteorological forecasting.** For medium-range weather forecasting, data-driven approaches were found to be as skillful or even skilful than traditional numerical weather prediction has been demonstrated by the deep learning weather models and are exemplified by the recently developed GraphCast, which was able to skillfully forecast the global weather with history-only reanalysis data [Lam et al., 2023]. In more sweeping surveys of deep learning for time-series forecasting it appears meteorological forecasts are a recurring exercise in stress testing: while most architectures work well on the periodic structure, uncertainty estimates on sudden departures from it, when they are likely to be most valuable to an operator, break down easily [Li & Law, 2024].

**Physiological and health-monitoring time series.** The multimodal data of the physiological signals (blood volume pulse, electrodermal activity, respiration, temperature) collected in the baseline, amusement and induced-stress conditions of Schmidt et al. is a benchmark for stress and affect modelling from wearable signals, which is widely used to-date [Schmidt et al., 2018]. More current research has begun to shift from the classification to forecasting the stress. The promise and label-scarcity problem across this domain is the capability of self-supervised pretraining information on each subject's baseline dynamics to predict future stress events from multimodal biosignals, given a small number of labelled events [Islam & Washington, 2023]. The quality of the calibration has even real implications on the use of a physiological forecast; in this particular context, it is not a nuisance added to the point accuracy, it is probably the most important one.

**Positioning.** All of these three domains are different and have their own architectures already constructed/skewed for that niche, which work really well. What is actually rare is when everyone uses the same architecture – developed and validated on foundational infrastructure data – with no changes whatsoever, across all three. This is the space between. It's the space in the middle this paper fills.

## 3. Framework Summary

Here, the forecasting system is not re-derived from scratch, but rather its general description, which is crucial to understanding the cross domain results below, is summarized in the companion framework paper [11].

Five conditioned stages execute the target series together with exogenous covariates, categorical/context and explicit mask for block occurrences of the missingness. One, a Causal-Regime Tokeniser (CaRT) learns differentiable regime boundaries directly from the data, without any external change-point heuristic, learning also compact context tokens which are causally-invariant. Then a Wavelet-State-Space Dual-Path Predictor (WSDP) generates two forecasts per horizon: one from a wavelet decomposed latent structure with multi-scale periodic and transient structure and one from a Wiener-Hopf path forecast based on a latent structure modeled by a neural state-space structure with smooth latent structure. These two paths are built to fail in different ways on purpose. Finally, an evidence-weighted fusion module (EviFuse) fuses them treating them with weights based on scores from recent residual behaviour and regime context, and not fixed weights as is in other detectors; this means that the fused forecast will favour the more reliable path. A stress-testing cross-regime

validation suite (CF-XValid) is developed during the system development phase to test the system with an imposed range of specified shocks, rather than under ordinary held-out splits. The drift-aware continual-adaptation module (DriftShield-ACL) monitors the drift in the distribution of tokens and residuals and adapts the predictor by invoking the light update option, reconfigure prediction intervals online using Conformal Prediction and retrain prediction intervals only when needed. No change was made to this pipeline for the current study—in the three domains below, no component was retrained with a different architecture or with a different set of hyperparameters. The same sets of trained instances as described in the companion paper are used here.

#### 4. Datasets and Experimental Setup

Three datasets were chosen that share characteristics of non-stationarity, contextual dependence and changing dynamic with well documented cases, but are structurally distinct from the framework's original non-stationarity, contextual dependency and evolving dynamics, and from each other: financial market series, a meteorological series, and a physiological signal series.

**S&P 500 daily closing prices (5-year period).** The daily closing-price data which were tested for the framework, were taken across macroeconomic-announcement- as well as volatility-driven regime shifts found in financial markets, also across bull markets and correction time periods.

**NOAA hourly temperature (3-year period).** The hour-by-hour temperature data was once used to assess the forecasting capabilities across the four seasons and deviations from regular temperature (storm, heatwave).

**WESAD physiological stress data.** To predict a continuous stress index, physiologic time series – such as heart-rate variability, electrodermal activity and motion – were used, incorporating non-stationarity brought about by circadian rhythms, behaviour and movement between rest and induced stress.

The sets of datasets are then compared to the three base predictors developed for the framework's core validation: NSPLformer [8] (a progressive-learning transformer for non-stationary series), a Two-Phase Taylor series/Reinforcement Learning hybrid [9] and a Kernel Based Non-Markovian predictor [10] which were chosen for their methodological diversity for direct comparability between the two studies. The evaluation metrics are aligned with the time-honored reporting practices of each domain—mean absolute error, root mean square error, mean absolute percentage error, continuous ranked probability skill score, and empirical coverage 90%—and a 90% empirical coverage is added to the financial and meteorological series, where available, along with an F1-score metric for the stress-classification component of the WESAD series.

Each dataset is divided into a 70:30 train–test split consistent with that used to validate the key elements of the framework, to prevent information from the test period from entering into the train set for any domain. Splits are not pooled (final), but per-series and are applied per ticker for S&P 500, per station for NOAA, per subject for WESAD which is to say that a subject/station present in train is never evaluated on unseen data from a different subject/station in the future.

For the S&P index series the first 70% of each five year/day closing window  $\approx 3.5$  years, is used for training and the last 30%  $\approx 1.5$  years, is used for testing with a deliberate inclusion of at least one volatility regime not experienced during training. Input values are based upon a 336-day trailing time period, with forecast periods of 5, 10, and 20 trading days representing short, medium, and long-range, respectively, within a domain that is dominated by weeks, and not months. All variables are log-normalized prior to entry in the CaRT regime-tokenization stage to ease variance; exogenous variables are daily trading volume and realised volatility over the previous 20 days area.

In the case of the NOAA hourly series, the first 70% of the three-year window ( $\approx 25.6$  months) is used for training and the last 30% ( $\approx 11$  months) is used for testing, and the last 30% ( $\approx 11$  months) time period is also a whole seasonal cycle, the split time window was chosen in this way. All three input (24/48/96-hour) forecast lengths are computed on a 336-hour (14-day) trailing window, as are all three benchmark lengths for infrastructure. Exogenous covariates are humidity, wind speed and pressure readings with an hourly resolution.

In the 70:30 splits for WESAD, each one of the 70%/30% splits is by subject, not by calendar time, so that each subject's baseline, amusement, and stress protocol were used for the first 70% of their continuous recording of physiology, and the last 30% was used for the test set, without mixing up the segments in the time series. Blood volume pulse, electrodermal activity, respiration, and temperature signals are all sampled at the native rate of the dataset but resampled to a common 4 Hz sampling rate to make up the 60 second input window and 30 second forecast horizon for a continuous stress index, matched window for F1-scored stress/non-stress classification. Including absolute physiological signals varies greatly between individuals, thus prior to the tokenisation by CaRT, signals are z-normalised per subject.

Preprocessing then proceeds as described in the companion framework paper [11] for the infrastructure benchmarks, but there is no change in the architecture of any of the five pipeline stages from the other models.

## 5. Results

**Table 1. Stock Price Prediction Performance (S&P 500 Daily Closing Prices, 5-Year Period)**

| Metric       | NSPLformer | Taylor + RL | Kernel Non-Markovian | Proposed Model |
|--------------|------------|-------------|----------------------|----------------|
| MAE          | 1.842      | 2.103       | 1.956                | <b>1.612</b>   |
| RMSE         | 2.734      | 3.082       | 2.881                | <b>2.389</b>   |
| MAPE (%)     | 2.41       | 2.86        | 2.58                 | <b>2.07</b>    |
| CRPS         | 1.192      | 1.364       | 1.281                | <b>1.034</b>   |
| 90% Coverage | 83.9%      | 79.5%       | 82.1%                | <b>89.4%</b>   |

Market environments evolve rapidly — macro-economic announcement, geopolitical development, sudden period of volatility clustering, and so on, the market environment changes. The NSPLformer is quite adaptive, though exhibiting spurts in volatility following the sudden changes. The Taylor + RL hybrid shows poor RMS error when the signal has a high frequency of fluctuations and is delivered in a nonlinear form. The Taylor + RL hybrid becomes worse when the signals have high frequency fluctuations and come in nonlinear form. It is also possible to apply kernel-based prediction that has a nice long memory, but is over-smooth about price shocks in the process. The proposed framework yields the lowest MAE and RMSE here, consistent with the cross-transparency CaRT offers in the regime-change-regimes being reasonably well carried over into financial regime dynamics. It's more than just that these lower CRPS or that this roughly 6-point increase in coverage over NSPLformer; conformal recalibration is capable of dealing with financial-market drift, not just the drifts inherent to the grid or sensors over which it was developed and tested.

**Table 2. Weather Forecast Performance (NOAA Hourly Temperature, 3-Year Period)**

| Metric       | NSPLformer | Taylor + RL | Kernel Non-Markovian | Proposed Model |
|--------------|------------|-------------|----------------------|----------------|
| MAE (°C)     | 1.87       | 2.14        | 1.98                 | <b>1.63</b>    |
| RMSE (°C)    | 2.65       | 2.94        | 2.77                 | <b>2.31</b>    |
| CRPS         | 1.28       | 1.47        | 1.35                 | <b>1.09</b>    |
| 90% Coverage | 84.6%      | 80.8%       | 82.9%                | <b>89.2%</b>   |

The temperature series includes deviations that tend to be mostly periodic and seasonal in nature and are driven by storm and heat waves. The trend of seasonal changes is captured close to NSPLformer, but diminishes slightly with sudden weather changes. Local-approximation is prone to instability and bias at temperature extremes with the Taylor + RL method. The seasonal pattern is forecast well, but is slow to respond when a quick-moving front pass. The regime conditioned wavelet decomposition and dynamic state modeling of the proposed model reduce the MAE, RMSE values and maintain the calibration, respectively. The lower CRPS and the improvement in the coverage give additional indications that adequate calibration occurred when the storm reached the service area at a particular stage of the transition: when the storm struck the area. The lower CRPS and improved coverage suggest some specific change in overall calibration at storm onset: the storms were calibrated correctly at the point they reached the area.

**Table 3. Stress Level Prediction Performance (WESAD Physiological Stress Dataset)**

| Metric             | NSPLformer | Taylor + RL | Kernel Non-Markovian | Proposed Model |
|--------------------|------------|-------------|----------------------|----------------|
| MAE (Stress Index) | 0.318      | 0.356       | 0.337                | <b>0.276</b>   |
| RMSE               | 0.447      | 0.503       | 0.472                | <b>0.389</b>   |
| F1-Score           | 0.82       | 0.76        | 0.79                 | <b>0.87</b>    |
| CRPS               | 0.261      | 0.297       | 0.278                | <b>0.229</b>   |
| 90% Coverage       | 85.1%      | 81.7%       | 83.4%                | <b>90.3%</b>   |

The non-stationarity of WESAD is caused by circadian cycles, behaviour variability, and quick transits from rest to stress. NSPLformer does well in steady states and is slow to move in response to stress. The Taylor + RL method becomes unstable when the physiological change is very fast. The prediction of a kernel can't differentiate between ordinary measurement noise and context driven state transition. This proposed model has improved results on all metrics including F1 score when compared to each baseline. The coverage and the calibration issue is also important for a practical reason, as calibrated forecasts can be a genuine risk if there is an actual stress transition: too much uncertainty and nothing more is missed.

The results of the evaluation of the framework, which exceeds all of the three baseline points across the three domains both on a point basis and as assessed by a probabilistic error measurement demonstrates that. The theory of non-stationarity of these three domains has little to nothing in common in terms of what factors actually explain their non-stationarity (economics-based outlets, physics-based processes in the atmosphere, biology-based processes in human physiology). It's this consistency rather than three separate cases of fit in a specific domain that make this evidence for domain transfer and not simply domain specific fit.

## 6. Discussion

Calibration is more prominent than any other signal across all three domains: not just accuracy at individual points, but calibration. For each domain, the proposed framework is estimated to have 90% empirical coverage in the neighborhood of a point of the nominal target. Each base-line undercovers, typically 5–10 percentage points and doesn't change significantly from one domain to another, even those with different causal non-stationarity. I don't know how to explain that as a coincidence that corresponds to the structure of one particular domain. It integrates much better with the baseline design philosophy of continuously monitoring the drift state and re-calibrating the drift interval (DriftShield-ACL) as opposed to re-calibrating the interval once during the training, and then forget about it. The same holds true for worst-case behaviour: because the biggest gaps in these tables show up exactly where each domain's regime shift is sharpest (e.g., in financial volatility, storm onset, quick stress transitions), worst-case error is covered well by reliability-aware fusion, which did not perform well for the average case during the original compound-stress testing on infrastructure data.

However, the extent of the improvement in the various domains is not equivalent and this difference should not be overlooked but noted. The largest coverage gain is tailored to financial series, with approximately 6 points, suggesting an abrupt and discrete change, similar to the changes in the 'level' and 'demand surge' settings that the framework's counterfactual validation suite was designed for. The WESAD gains, though better than those obtained by the other three methods tested, are not as large as other signals, since physiological signals are less sharply defined in terms of regime boundaries and have more gradual circadian-based non-stationarity as compared to a sudden shock. Given what is in front of us, this is a plausible explanation, which isn't a proven explanation as it will require further datasets for each domain than are included in this current study of work, and is thus an excerpt of the future study to be settled upon it for proof, not a final conclusion.

This paper does not propose: These results apply to transfer to 3 domains, as measured on these aggregate metrics. They do not make the claim that the segmentation of CaRT is individually required in each newly defined domain (although they claim that, but it would need an ablation study for each domain, not simply being taken over from the companion paper's infrastructure-domain ablations). Nor does it indicate any computational expense or latency in a deployment with totally different live requirements, such as high frequency trading compared to weather updates on an hourly basis, for example. This paper does not propose: These results apply to transfer to 3 domains, as measured on these aggregate metrics. They do not make the claim that the segmentation of CaRT is individually required in each newly defined domain (although they claim that, but it would need an ablation study for each domain, not simply being taken over from the companion paper's infrastructure-domain ablations). Nor does it indicate any computational expense or latency in a deployment with totally different live requirements, such as high frequency trading compared to weather updates on an hourly basis, for example.

## 7. Limitations

Experiments with transfer are conducted on just one dataset for each of three domains. If you're looking for a more general claim of domain-general transfer, then you need more data within each, say, several equity indices; several weather stations; some data beyond the WESAD; to stick out the transfer from a fit that's just plain working on this specific set of data. The results were no different with the default tuning – which is the whole point of this study – but also reflects a lower bound on the potential of the framework given some adaptation based on domain knowledge – a fair comparison when testing the generalization, but not in how a practitioner would want to test the best model in a specific domain. The area of finance represents one specific 5-year period and performance for other combinations of macroeconomic policy — be it a prolonged period of unstable volatility or a lengthy period of little liquidity — can't be predicted. Just as in the companion framework paper [11] also, the widespread ability to adapt constantly to changes in domains is constrained to light-weight adapter updates; for faster-moving domains, particularly financial markets, over long live-deployment cycles, this retrospective evaluation remains unable to answer if it's still sufficient. This would require a live or simulated streaming study to be performed instead.

## 8. Conclusion

In this paper, we investigate whether a five-stage, regime-aware forecasting framework developed and tested on infrastructure-style data scales up to domains characterized by a different type of non-stationarity, unaltered. In all three categories (financial, meteorological and physiological time series), it did; lowest MAE and RMSE was obtained vs. three contemporary baselines in all of them, F1 scores for stress classification (physiological) were the best, and probabilistic calibration (CRPS or gap to 90%-baseline) was best, with the confidence level being at or within 0–1 point around the 90%-confidence target compared to the scoring of 5–10-point undercoverage of 90%-baseline. The signature is striking because it involves three sources or classes of non-stationarity that are basically unrelated — that it's exhibiting the same signature across all three of these domains is highly suggestive that the architecture underpinning this recursive learning mechanism is a valid general framework for non-stationary time series, rather than an architecture suited around a particular infrastructure domain. This generalization property is important in isolation from the accuracy that may result in any domain if one is implementing one forecasting system in more than one stream that is statistically different.

## 9. Future Scope

From there a number of natural implications come. Additional testing on other datasets at each domain (several equity markets under different volatility regimes, multiple weather stations, and physiological datasets other than WESAD) would support the domain-transfer argument more than just a single dataset for each domain. A per domain destruction of CaRT, WSDP and EviFuse would identify what pieces create the transfer in the first place and what piece(s) turns out to be neutral when not embedded in the walls of the infrastructure domain it was designed to remain in. Consideration of the computational cost and latency for various areas with vastly different real-time requirements, as well as testing DriftShield-ACL's adaptation to different dynamics such as fast-moving financial data intraday, would resolve deployment questions that couldn't be addressed in this retrospective study.

## References

- [1] Y. Liu, H. Wu, J. Wang, and M. Long, "Non-Stationary Transformers: Exploring the Stationarity in Time Series Forecasting," *Advances in Neural Information Processing Systems*, vol. 35, 2022.
- [2] Y. Zhang et al., "Deep Learning Models for Price Forecasting of Financial Time Series: A Review of Recent Advancements: 2020-2022," *WIREs Data Mining and Knowledge Discovery*, 2024. <https://doi.org/10.1002/widm.1519>
- [3] P. Ranjan, R. Itani, and A. Faccia, "An Interpretable 1D-CNN Framework for Stock Price Forecasting: A Comparative Study with LSTM and ARIMA," *FinTech*, vol. 4, no. 4, p. 63, 2025. <https://doi.org/10.3390/fintech4040063>
- [4] R. Lam, A. Sanchez-Gonzalez, R. Keisler, et al., "Learning Skillful Medium-Range Global Weather Forecasting," *Science*, vol. 382, no. 6677, pp. 1416-1421, 2023. <https://doi.org/10.1126/science.adi2336>
- [5] W. Li and K. L. E. Law, "Deep Learning Models for Time Series Forecasting: A Review," *IEEE Access*, vol. 12, pp. 92306-92327, 2024. <https://doi.org/10.1109/ACCESS.2024.3422528>
- [6] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proc. 20th ACM Int. Conf. Multimodal Interaction (ICMI)*, 2018.
- [7] T. Islam and P. Washington, "Personalized Prediction of Recurrent Stress Events Using Self-Supervised Learning on Multimodal Time-Series Data," *arXiv preprint arXiv:2307.03337*, 2023.
- [8] S. Jiaying, L. Yanhui, and Z. Yuying, "NSPLformer: Exploration of Non-Stationary Progressively Learning Model for Time Series Prediction," *Scientific Reports*, 2025. <https://doi.org/10.1038/s41598-025-13680-2>
- [9] D. Dey, L. Ghosh, D. Bhattacharya, et al., "A 2-Phase Prediction of a Non-Stationary Time-Series by Taylor Series and Reinforcement Learning," *Applied Soft Computing*, 2023. <https://doi.org/10.1016/j.asoc.2023.110565>
- [10] F. Gilani, D. Giannakis, and J. Harlim, "Kernel-Based Prediction of Non-Markovian Time Series," *Physica D: Nonlinear Phenomena*, 2021. <https://doi.org/10.1016/j.physd.2020.132829>
- [11] Pradnya P. Shelar and M. Bhatnagar, "Regime-Aware Evidence Fusion for Calibrated Forecasting of Non-Stationary Time Series," *Journal of Intelligent Decision Making and Information Science*, Vol. 3 No. 13s, 2026.
- [12] Taqvi, S. (2025). Data-driven agricultural development in Assam through machine learning. *Journal of Artificial Intelligence in Agriculture and Applied Research*, 1(1), 34–48.