Int. J. AI Adv. Comput. Vol.1, Iss.4 (2026). pp-145-152

Contents lists available at AcademiansEdu

International Journal of AI and Advanced Computing
ISSN: 3141-3161
journal homepage: https://academiansedu.org/ijaac/

Research Article

Early Paddy Yield Forecasting in Sri Lanka Using SARIMA and LSTM for Farmer and Collector Decision Support

R. M. Piyumi Prabodhya Jayasooriya[1]*, Saravanapavan Nasiketha[2]
[1]*Faculty of Computing, NSBM Green University, Homagama, Sri Lanka*
[2]*Faculty of Computing, NSBM Green University, Homagama, Sri Lanka*

ARTICLE INFO

Article history:
Received: 12 August 2026
Received in revised form: 18 September 2026
Accepted: 20 October 2026
Available online: 31 October 2026

Keywords:
paddy yield forecasting; SARIMA; LSTM; decision-support system; Sri Lanka; usability

DOI:
10.36099/ijaac.v1i4.003

ABSTRACT

Early paddy-yield information can support cultivation and procurement planning in Sri Lanka, where farmers and collectors make decisions before final harvest outcomes are known. This study developed and evaluated a web-based decision-support system (DSS) that integrates district-level paddy-yield forecasting with separate farmer and collector workflows. Historical district-seasonal records from 2004–2023 covered 21 retained districts and 797 observations. District-specific Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) models were compared using an archived final-four-observation holdout. Mean Absolute Percentage Error (MAPE), averaged across districts, was 22.06% for SARIMA and 17.38% for LSTM, with LSTM recording lower MAPE in 12 of 21 districts; a paired district comparison reported *p* = 0.0333. However, chronology, scaling, and unequal forecast-information horizons limit this archived comparison, so the result is not interpreted as validated LSTM superiority. A problem-validation survey of 122 participants (103 farmers and 19 collectors) identified reported planning uncertainty and strong perceived need for earlier forecasts. The DSS was implemented using FastAPI and React/Vite with English and Sinhala presentation. A post-demonstration evaluation with 34 intended users produced a mean System Usability Scale score of 73.53. A shared planning-uncertainty item did not differ significantly from the independent baseline sample. The contribution is therefore an integrated and bounded forecasting-DSS prototype with preliminary stakeholder evidence and explicitly stated validation limitations.

1. Introduction

Rice is central to Sri Lankan food security and rural livelihoods (Rice Research and Development Institute, n.d.). The Department of Census and Statistics (DCS) publishes seasonal paddy statistics for Maha and Yala cultivation (Department of Census and Statistics, n.d.). These records are valuable for monitoring past production, but farmers must commit land, labor, and inputs before harvest, while collectors must anticipate procurement, transport, and storage requirements before actual supply is known. Early yield information can therefore have decision value even when it is presented as an estimate rather than as a guaranteed farm-level outcome.

Paddy yields vary across climatic zones and water regimes, with rainfall and temperature among relevant environmental influences (Chandrasiri et al., 2023). At the same time, consistent district-level historical yield records provide a practical basis for examining whether earlier observations can support time-series forecasting. District resolution offers a compromise between a national average that masks local variation and village- or farm-level prediction that the available historical data cannot support. In this study the forecasting target is average yield in kilograms per hectare, not total production, market price, or an individual farmer's harvest.

Prior Sri Lankan research includes ARIMA and seasonal forecasting of paddy production (Sivapathasundaram & Bogahawatte, 2012; Munasingha & Napagoda, 2021), neural paddy-yield estimation (Nandani & Vidanapathirana, 2024),

© 2025 The Authors. Published by AcademiansEdu. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International License (CC BY 4.0) (https://creativecommons.org/licenses/by/4.0/).

and more recent satellite-, weather-, and historical-data approaches for Sri Lankan agro-zones (Senevirathne & Premasiri, 2025). Broader crop-yield research also documents increasing use of machine learning (van Klompenburg et al., 2020). Agricultural decision-support system (DSS) studies further show that prediction and decision support are distinct tasks because model outputs must be translated into information that users can interpret for planning (Zhai et al., 2020; Mohamed et al., 2023; Anbananthen et al., 2021). These streams motivate a more integrated Information Systems question: how can a district-level forecasting comparison be translated into role-oriented information for farmers and collectors and then evaluated with intended users?

This paper addresses that question through four connected components. First, district-specific Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) models are compared on the same retained district-seasonal yield dataset. Second, a problem-validation survey examines planning uncertainty and perceived need for early forecasting among farmers and collectors. Third, the resulting forecasts are incorporated into a bilingual web DSS with separate farmer and collector workflows. Fourth, the working prototype is evaluated after demonstration using the System Usability Scale (SUS) and exploratory acceptance measures. The contribution is not a claim of a nationally deployed forecasting service or proven economic improvement; rather, it is a traceable forecasting-to-decision-support artifact with explicit evidence boundaries.

The paper makes three contributions. The first is a district-wise comparison that preserves differences between locations rather than collapsing the historical data into one national series. The second is a role-oriented system design in which farmers and collectors receive different interpretations of the same underlying district information. The third is an evaluation chain that keeps forecasting error, reported stakeholder need, software verification, and post-demonstration usability analytically separate. This separation is particularly important in agricultural decision support because a usable interface cannot compensate for weak model validation, and a numerically accurate model may still fail to communicate information that users can interpret.

2. Related Work and Research Gap

Statistical forecasting has a clear precedent in Sri Lankan paddy research. Sivapathasundaram and Bogahawatte (2012) used ARIMA for national paddy production forecasting, while Munasingha and Napagoda (2021) analyzed production trends and Maha/Yala forecasting. These studies establish the relevance of time-series methods, but their target is aggregate production rather than district average yield. Climate-oriented work further shows that paddy performance differs across climatic zones and that environmental variables can add explanatory information (Chandrasiri et al., 2023). Nandani and Vidanapathirana (2024) applied an artificial neural network to Sri Lankan paddy-yield forecasting with climate-related inputs, while Senevirathne and Premasiri (2025) more recently reported a satellite-, weather-, and historical-data approach for Sri Lankan agro-zones. Such studies support nonlinear and multivariate modeling, but their targets and input structures are not directly comparable with the univariate district-yield setting used here.

Agricultural DSS literature presents a complementary perspective. Zhai et al. (2020) identify information delivery, integration, and usability as continuing challenges for Agriculture 4.0 decision support. Mohamed et al. (2023) demonstrate a Sri Lankan mobile decision-support platform aimed at agricultural sustainability, and Anbananthen et al. (2021) report a web-oriented intelligent crop-yield DSS using hybrid machine learning. These systems show that prediction and decision support are related but distinct tasks: a numerical model result must be contextualized for the decision maker. Sri Lanka also introduced broader national digital-agriculture foundations, including CROPIX, in 2026 (Food and Agriculture Organization of the United Nations, 2026). The present prototype neither integrates with nor replaces such national platforms; its contribution is the narrower forecasting-to-role-oriented-DSS evidence chain evaluated in this study.

Forecast evaluation also requires explicit temporal discipline. General forecasting guidance recommends assessing later observations using only information that would have been available at the forecast origin (Hyndman & Athanasopoulos, 2021). Bergmeir et al. (2018) discuss the conditions under which cross-validation for autoregressive time series is valid and reinforce the need to avoid information leakage. Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) remain in the target unit, whereas Mean Absolute Percentage Error (MAPE) expresses error proportionally and can be misleading near zero (Hyndman & Koehler, 2006). These considerations are directly relevant to the present work because its district series are short, seasonal, and evaluated through an archived holdout rather than a completed rolling-origin design. Table I summarizes selected prior work and its distinction from the present study.

Table I
Selected Related Work and Distinction from the Present Study

Study	Relevance and Distinction
Sivapathasundaram and Bogahawatte (2012)	ARIMA for national paddy production; differs from district average-yield forecasting.

Munasingha and Napagoda (2021)	Maha/Yala production forecasting; not the same district-level yield target.
Nandani and Vidanapathirana (2024)	Neural paddy-yield forecasting with climate inputs; different architecture and predictors.
Zhai et al. (2020)	Agricultural DSS survey; supports translation of analytics into decision support.
Mohamed et al. (2023)	Sri Lankan agricultural mobile DSS; advice-oriented rather than the same SARIMA/LSTM workflow.
Anbananthen et al. (2021)	Web-based intelligent crop-yield DSS; different data, models, and user-evaluation setting.

The research gap is therefore narrower than claiming that paddy forecasting or digital agriculture is absent in Sri Lanka. The gap addressed here is the integration of (i) a documented SARIMA/LSTM comparison on one district-seasonal yield basis, (ii) primary validation of farmer and collector planning needs, (iii) role-oriented forecast presentation with explicit assumptions, and (iv) post-demonstration usability and acceptance assessment. This bounded framing avoids comparing performance figures produced from incompatible targets, datasets, and evaluation designs.

3. Methodology

3.1 Research Design

The study followed a predominantly quantitative design informed by Design Science Research Methodology (DSRM), which is appropriate when an Information Systems study develops and evaluates an artifact rather than stopping at algorithm comparison (Peffers et al., 2007; Hevner et al., 2004). The research combined historical forecasting, a problem-validation survey, iterative DSS development, representative functional verification, and a post-demonstration user evaluation. Evidence from these components was interpreted separately before being synthesized; for example, favorable usability ratings were not treated as proof of forecasting accuracy or economic impact.

3.2 Historical Dataset and Forecasting Target

The supplied modeling file was a cleaned compilation attributed in the project materials to DCS seasonal paddy publications; DCS Paddy Statistics is cited here only as the institutional source context (Department of Census and Statistics, n.d.). The file contained 797 district-season observations from 2004–2023 across 21 retained districts. Twenty districts contained 38 seasonal records, while Nuwara Eliya contained 37 because the 2007 Yala row was absent. That season was left absent rather than replaced by an imputed yield value. Districts with incomplete earlier coverage and irrigation-area labels that were not comparable districts were excluded. Each retained row represented a district-season observation rather than a farm. The model target was Average Yield (kg/ha). Maha and Yala were intended to form one alternating sequence per district, so SARIMA used a seasonal period of two rather than fitting independent models for each season.

The comparison deliberately used historical yield as the common input basis. Scheme-specific extent fields were retained for later decision-support calculations, but rainfall, soil, market price, and future cultivated area were not model inputs. This kept the historical-yield input basis common to both model families while limiting the ability to anticipate a new drought, flood, pest event, or field-specific management change.

Data preparation standardized district labels, removed national or summary rows, converted numerical fields, and checked the retained district-year-season keys for duplicates. The retained target contained no missing Average Yield values. The 21-district scope reflects comparable historical availability rather than the total number of administrative districts in Sri Lanka. Consequently, every model result is interpreted at district-average level; a district forecast cannot be treated as a measurement of an individual holding with different irrigation, variety, soil, or management conditions.

3.3 SARIMA and LSTM Models

For each retained district, SARIMA candidates were selected with pmdarima auto_arima using seasonal = True, m = 2, stepwise search, and district-specific orders. The final four labeled seasonal observations formed the archived holdout, and SARIMA produced a four-step forecast from a single origin. Forecast accuracy was recorded using MAE, RMSE, and MAPE. These measures respectively express average absolute error, emphasize larger errors, and report proportional error; no single metric is interpreted as proof of operational reliability (Hyndman & Athanasopoulos, 2021; Hyndman & Koehler, 2006).

The LSTM used the four preceding seasonal yield observations, approximately two Maha–Yala cycles, to predict the next value. The implemented network contained one LSTM layer with 50 units and rectified linear unit (ReLU) activation, followed by a one-unit dense layer. Adam optimization, mean-squared-error loss, a maximum of 200 epochs, batch size four, and early stopping on training loss with patience 15 were used. The four-observation window was a practical design choice rather than the result of a reported exhaustive hyperparameter search. LSTM follows the gated recurrent architecture introduced by Hochreiter and Schmidhuber (1997).

3.4 Model Evaluation Boundary

District-wise SARIMA and LSTM MAPE values were paired for an archived paired *t*-test. The test used one MAPE summary per district (21 pairs), not the 797 historical rows as independent observations. Importantly, the archived comparison has three known limitations. First, an audit found that the year-season sort placed Maha before Yala within the same starting-year label, creating a chronology concern. Second, the LSTM MinMaxScaler had been fitted before separating the holdout, allowing information about the full value range to enter scaling. Third, SARIMA generated a fixed-origin four-step forecast while successive LSTM test windows could contain earlier observed holdout values, creating unequal forecast-information horizons. Forecasting guidance favors temporally ordered, leakage-aware evaluation and matched origins (Hyndman & Athanasopoulos, 2021; Bergmeir et al., 2018). The recorded comparison is therefore reported transparently rather than relabeled as a corrected rolling-origin experiment.

3.5 Stakeholder Studies

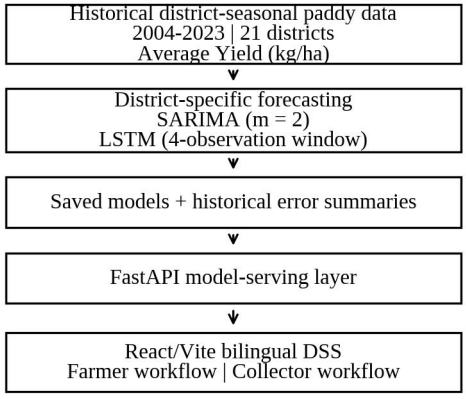
Study 1 used a structured Sinhala questionnaire to examine Experience-Based Planning (EBP), Planning Uncertainty (PU), Experienced Loss/Mismatch (ELM), and Felt Need for Forecasting (FN) on five-point scales. Convenience/purposive recruitment through agriculture-related contacts and organizations yielded 122 valid responses: 103 farmers and 19 collectors. Sixteen early collector submissions affected by a form-branching error were excluded before corrected recollection. The survey describes the respondent sample and is not used to estimate national prevalence.

Study 2 evaluated the working DSS after guided demonstrations in Kurunegala and Anuradhapura. The final sample comprised 34 intended users (25 farmers and nine collectors). It included the ten-item SUS (Brooke, 1996), Technology Acceptance Model (TAM)-informed perceived-usefulness and behavioral-intention items (Davis, 1989), and trust items. A first incomplete 22-response round omitted one SUS item and was discarded; 34 complete responses were recollected with the corrected instrument. The Study 1 and Study 2 participants were anonymous and independent, so a shared planning-uncertainty item was treated as an independent-group comparison rather than a matched before-and-after intervention. Participation in both studies was voluntary and anonymous; respondents were informed of the study purpose, and results were analyzed in aggregate.

For Study 1, internal consistency was reviewed before construct-level interpretation. H1 compared planning uncertainty with the neutral midpoint of three, H2 examined the association between experienced loss/mismatch and felt need for forecasting, and H3 compared farmer and collector responses using Mann–Whitney U tests. In Study 2, SUS was scored using the standard ten-item procedure; perceived usefulness, trust, and behavioral intention were summarized as exploratory acceptance indicators. H4 used only the planning-uncertainty item shared across the two independent surveys. Statistical significance was interpreted together with sampling and measurement limitations rather than as evidence of causal system impact.

4. DSS Design and Implementation

The DSS translated the forecasting experiment into two decision contexts over a shared service. Offline Python/Google Colab processing produced district-specific SARIMA and LSTM artifacts and historical error summaries. Saved statistical models, neural-network models, and district scalers were served through FastAPI, while a React/Vite interface presented English and Sinhala pages. Figure 1 summarizes the evidence-to-interface workflow.



Forecasting and role-oriented DSS workflow.

A general district request used the model family with the lower archived district MAPE as an operational selection rule. This rule supported the prototype but is not evidence that the same model would remain best after corrected prospective validation. The farmer route deliberately followed a saved district LSTM inference path. Reported previous harvest and land area were converted to kg/ha, transformed by the saved scaler, and inserted into the four-observation window; model weights were not retrained. The result is an illustrative adjustment of the next-season estimate, not farm-specific agronomic advice.

The collector route used the general district forecast and converted it to indicative production through the latest available net harvested extent: indicative production (MT) = forecast yield (kg/ha) × assumed harvested extent (ha) / 1,000. The entered collection requirement was then compared with this indicative amount to show a possible surplus or shortfall. Because future harvested extent and marketable supply were not forecast independently, the output is labeled as planning support rather than a guaranteed quantity available for procurement.

At the service boundary, the documented application programming interface (API) exposes a district list, a general district forecast, a farmer-specific route, and a collector-specific route. Positive numerical constraints reject missing, zero, and negative values before forecast calculation. The frontend receives structured, language-neutral values and handles English/Sinhala labels at presentation level. This separation allows one forecasting service to support different

decision contexts without training a separate model for every individual user.

The interface also displays Low/Medium/High historical-error labels using MAPE thresholds below 15%, 15–25%, and above 25%, respectively. These labels are communication heuristics rather than calibrated predictive probabilities or confidence intervals.

System verification focused on the implemented prototype rather than production deployment. Ten representative functional cases and six dedicated non-functional checks were recorded as passed. The functional cases covered valid and invalid districts, farmer harvest/area inputs, collector requirement inputs, and language switching; non-functional checks covered browser access, bilingual presentation, inference without retraining, input validation, and historical-error communication. These checks establish representative behavior under recorded test conditions; they do not establish production-grade scalability, security, or availability.

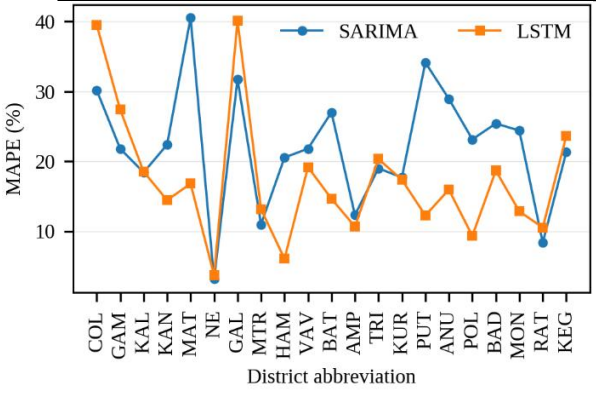
5. Results

5.1 Forecasting Results

The archived 21-district comparison recorded a mean district MAPE of 22.06% for SARIMA and 17.38% for LSTM, a difference of 4.68 percentage points. The paired district comparison reported *t* = 2.2863 and *p* = 0.0333. LSTM recorded lower MAPE in 12 districts, while SARIMA recorded lower MAPE in nine. This district-dependent pattern is more informative for the DSS than an aggregate average alone because the size and direction of the difference varied substantially by location. Table II summarizes the principal quantitative findings, and Figure 2 shows the district-wise MAPE pattern.

Table II
Key Quantitative Findings

Evidence	Measure	Recorded Result
Forecasting	Mean MAPE	SARIMA 22.06%; LSTM 17.38%
Forecasting	Districts lower MAPE	LSTM 12; SARIMA 9
Study 1 (N = 122)	PU / FN means	4.00 / 4.28
Study 1	ELM–FN association	*r* = 0.584, *p* < 0.001
Study 2 (N = 34)	SUS	73.53 (SD = 16.30)
Study 2	Behavioral intention	4.088 (SD = 0.866)
Shared PU item	Baseline-mean *t* / independent U	*t* = −1.124, *p* = 0.269; U = 1981.0, *p* = 0.648



Archived district-wise SARIMA and LSTM MAPE comparison.

Codes: COL Colombo; GAM Gampaha; KAL Kalutara; KAN Kandy; MAT Matale; NE Nuwara Eliya; GAL Galle; MTR Matara; HAM Hambantota; VAV Vavuniya; BAT Batticaloa; AMP Ampara; TRI Trincomalee; KUR Kurunegala; PUT Puttalam; ANU Anuradhapura; POL Polonnaruwa; BAD Badulla; MON Monaragala; RAT Ratnapura; KEG Kegalle.

The district results also caution against a single-model conclusion. SARIMA had lower MAPE in Colombo, Gampaha, Kalutara, Nuwara Eliya, Galle, Matara, Trincomalee, Ratnapura, and Kegalle; LSTM had lower MAPE in the remaining 12 retained districts. Differences were very small in Kalutara (18.41% versus 18.51%) and Kurunegala (17.69% versus 17.35%) but much larger in locations such as Puttalam (34.14% versus 12.24%). These results motivate district-specific presentation, while the evaluation limitations described in Section III prevent a definitive superiority claim.

The variation also affects how historical error is communicated to end users. A single national mean MAPE would hide districts in which both models were comparatively weak as well as districts in which their errors were very close. For that reason, the prototype associates the displayed district forecast with that district's recorded historical error rather than presenting one global reliability label. Even this district-specific indicator remains descriptive: it summarizes past errors from the archived evaluation and is not a calibrated probability that the next harvest will fall within a particular interval.

5.2 Study 1 Problem-Validation Results

Study 1 reported mean scores of 3.81 for EBP (SD = 0.61), 4.00 for PU (SD = 0.57), 4.16 for ELM (SD = 0.56), and 4.28 for FN (SD = 0.66). All four means were above the five-point neutral midpoint in the archived one-sample tests (*p* < 0.001). For the core planning-uncertainty result, PU was M = 4.00,

4.00, *t*(121) = 19.31, *p* < 0.001. ELM correlated positively with FN (*r* = 0.584, *p* < 0.001), while PU also correlated

with FN (*r* = 0.464, *p* < 0.001). These findings indicate that the surveyed participants reported both uncertainty and a perceived need for earlier forecasting information; they do not quantify actual financial loss caused by uncertainty.

Construct reliability supported different levels of confidence in the Study 1 scales. Cronbach's alpha was 0.754 for PU and 0.868 for FN, while EBP (0.643) and ELM (0.664) were more exploratory. A separate experience-based confidence item averaged 4.08 (*SD* = 0.69), showing that respondents could report confidence in their own experience while simultaneously reporting uncertainty about future yield. This distinction helped justify a DSS that supplements rather than replaces local experience.

Farmer–collector Mann–Whitney comparisons did not identify statistically significant differences for EBP, PU, ELM, or FN (all *p* > 0.05). This non-significance does not establish that the roles are equivalent. Their decision tasks remain different: farmers interpret yield in relation to their own cultivation context, whereas collectors consider procurement volume, transport, and storage at a broader scale.

5.3 Study 2 Usability and Acceptance Results

The final Study 2 sample produced a mean SUS score of 73.53 (*SD* = 16.30). Bangor et al. (2008) provide empirical context for interpreting SUS scores; accordingly, the observed score is treated as favorable immediate usability in this demonstration context rather than proof of sustained use. Perceived usefulness averaged 3.691 (*SD* = 0.807), trust 3.721 (*SD* = 0.780), and behavioral intention 4.088 (*SD* = 0.866). SUS internal consistency was high (alpha = 0.970), whereas the two-item perceived-usefulness and trust scales had weaker alpha values of 0.454 and 0.574; these short acceptance scales are therefore treated as exploratory.

The one shared planning-uncertainty item averaged 3.926 (*SD* = 0.694) in the independent Study 1 baseline and 3.706 (*SD* = 1.142) in Study 2. The archived one-sample comparison against the baseline mean reported *p* = 0.269, and an independent Mann–Whitney comparison reported *p* = 0.648. Neither result was significant at 0.05. Because participants were not paired or randomized, the descriptive decrease of 0.220 cannot be attributed to exposure to the DSS.

Exploratory Study 2 associations were positive: behavioral intention correlated with trust (*r* = 0.666, *p* < 0.001), perceived usefulness (*r* = 0.647, *p* < 0.001), and SUS (*r* = 0.536, *p* = 0.001). Perceived usefulness and trust were very strongly correlated (*r* = 0.965, *p* < 0.001), which, together with their weak two-item reliability, suggests construct overlap. These values are therefore presented as descriptive acceptance evidence rather than as a validated structural test of the Technology Acceptance Model.

6. Discussion and Limitations

The study provides three connected findings. First, the archived forecasting experiment shows that model behavior is district-dependent rather than uniformly favoring one family. The lower average LSTM MAPE and the 12-versus-nine district pattern are useful descriptive evidence, but the experiment does not satisfy the conditions required for a strong comparative model-selection claim. Second, Study 1 provides self-reported evidence of planning uncertainty and strong interest in early forecasts among the surveyed farmers and collectors. Third, the implemented prototype demonstrates how a shared district forecast can be translated into different farmer and collector workflows and received favorable immediate usability ratings after demonstration.

These findings also show why forecasting accuracy and DSS usefulness must not be collapsed into a single outcome. A lower historical MAPE does not establish that users will understand or act on a forecast, while a favorable SUS score does not validate the forecast itself. DSRM explicitly supports evaluating both the artifact and the evidence supporting its design (Peffers et al., 2007; Hevner et al., 2004). In the current prototype, farmer personalization is an LSTM input adjustment rather than a farm-calibrated model, and collector production is derived from an assumed historical extent rather than a forecast of future cultivated area. The user interface therefore presents assumptions and historical-error context alongside the numeric result.

From an information-systems perspective, the artifact contributes a traceable information flow from historical data through modeling and service logic to role-specific presentation. The same forecast is not presented as the same decision for every user: farmers receive a yield-oriented interpretation, whereas collectors receive an indicative supply comparison that requires an explicit extent assumption. This separation of model output, derived calculations, and user context is an information-quality design contribution because it reduces the risk of presenting a derived quantity as though it were directly predicted or observed.

The most important technical limitation is the archived model-comparison protocol. The season-ordering issue should be corrected so that Maha and Yala observations follow the intended chronology. LSTM scaling should be fitted only on the training data at each forecast origin. Both methods should then be compared using matched information sets, rolling or repeated forward origins, equal forecast horizons, and a simple seasonal-naive benchmark (Hyndman & Athanasopoulos, 2021; Bergmeir et al., 2018; Hyndman & Koehler, 2006). Until such replication is completed, the reported *p* = 0.0333 should be interpreted only within the archived procedure and not as proof that LSTM is intrinsically superior for Sri Lankan paddy forecasting.

6.1 Practical Implications

For farmers, the prototype is best understood as an early information aid that can be considered together with field knowledge, irrigation conditions, expected variety performance, and local advice; these contextual factors were not inputs to the implemented forecasting model. The system does not prescribe how many hectares to cultivate or how much input to purchase. For collectors, the forecast can support discussion of storage and procurement readiness, but the derived production estimate depends on an explicit historical harvested-extent assumption

and should not be interpreted as guaranteed marketable supply. Presenting these distinctions in the interface is part of the decision-support function because it helps users separate a model forecast from a subsequent planning calculation.

The district-dependent error pattern also has an operational implication. A deployment should not communicate one national reliability label when historical performance differs by district. District-specific error summaries can provide useful context, but they should be accompanied by data vintage, forecast season, model-selection rationale, and a statement that historical MAPE is not a prediction interval. If future versions incorporate weather, satellite, irrigation, or market variables, those inputs should be included only when their source, spatial alignment, forecast-time availability, and missing-data treatment are documented. Otherwise, adding more variables may create an appearance of sophistication without improving reproducibility or early-decision relevance.

6.2 Reproducibility and Evaluation Priorities

The next technical iteration should preserve a versioned data-preparation pipeline and a common evaluation protocol for every district. A reproducible run would explicitly order Maha and Yala, define the forecast origin and horizon, fit preprocessing only on training observations, compare SARIMA and LSTM against the same future targets, and retain a seasonal-naive benchmark. Repeated forward origins would show whether model rankings are stable or depend on one end-of-series holdout. The saved model used by the DSS should then be generated from the same validated pipeline rather than selected from an older comparison record. These steps would connect reported model performance more directly to the software artifact that users evaluate.

User evaluation should be strengthened in parallel rather than postponed until after technical refinement. A later field study could recruit a larger and more balanced farmer–collector sample, record whether participants can correctly interpret forecast units and assumptions, and use matched or longitudinal observations when examining changes in planning uncertainty. Where possible, evaluation should include realistic planning tasks rather than only attitude ratings. This would distinguish immediate interface acceptance from the more demanding question of whether the DSS changes decisions, reduces avoidable mismatch, or produces measurable operational value over a cultivation season.

User-evaluation limitations are equally important. Study 1 and Study 2 used non-probability samples, Study 2 contained only 34 participants from two demonstration areas, and the surveys did not track the same individuals. The evaluation measured reported usability, usefulness, trust, and intention following a guided demonstration; it did not observe independent multi-season use, changed cultivation or procurement decisions, income, or realized savings. The prototype also excludes live weather, soil, satellite, and market-price inputs and supports English and Sinhala rather than verified Tamil localization. Future work should therefore combine corrected forecasting validation with larger and, where feasible, matched or longitudinal field evaluation.

7. Conclusion

This research developed and evaluated a role-oriented web DSS for early district-level paddy-yield forecasting in Sri Lanka. A common historical yield basis was used to compare district-specific SARIMA and LSTM models, a problem-validation survey examined farmer and collector planning needs, and a FastAPI/React prototype translated forecast outputs into separate user workflows. The archived comparison recorded mean district MAPE values of 22.06% for SARIMA and 17.38% for LSTM, while Study 1 identified elevated reported planning uncertainty and strong forecast need among 122 respondents. Study 2 produced a mean SUS score of 73.53 among 34 intended users, indicating favorable immediate usability, but did not establish a significant or causal reduction in planning uncertainty.

The main contribution is therefore an integrated information-systems artifact and evidence chain rather than a claim of deployment-ready predictive superiority. Before wider operational use, the forecasting experiment should be repeated with corrected chronology, training-only scaling, matched rolling origins and horizons, and simple benchmarks; the DSS should then be evaluated through larger, less guided, and preferably longitudinal use. These steps would provide stronger evidence on both forecast reliability and the practical effect of the system on agricultural planning.

References

Anbananthen, K. S. M., et al. (2021). An intelligent decision support system for crop yield prediction using hybrid machine learning algorithms. *F1000Research, 10*, 1143. https://doi.org/10.12688/f1000research.73009.1

Bangor, A., Kortum, P., & Miller, J. (2008). An empirical evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction, 24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis, 120*, 70–83. https://doi.org/10.1016/j.csda.2017.11.003

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.

Chandrasiri, C. K., Tsusaka, T. W., Ho, T. D. N., Zulfiqar, F., & Datta, A. (2023). Impacts of climate change on paddy yields in different climatic zones of Sri Lanka: A panel data approach. *Asia-Pacific Journal of Regional Science, 7*(2), 455–489. https://doi.org/10.1007/s41685-022-00264-5

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Department of Census and Statistics. (n.d.). *Paddy statistics*. Agriculture and Environmental Statistics Division, Department

of Census and Statistics, Sri Lanka. https://www.statistics.gov.lk/Agriculture/StaticalInformation/PaddyStatistics

Food and Agriculture Organization of the United Nations. (2026). *Sri Lanka launches landmark digital foundations for a data-driven agriculture sector*. https://www.fao.org/srilanka/news/detail/sri-lanka-launches-landmark-digital-foundations-for-a-data-driven-agriculture-sector/en

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (3rd ed.). OTexts. https://otexts.com/fpp3/

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Mohamed, M. S. A., Wathugala, D. L., Indika, W. A., Madushika, M. K. S., & Ginige, A. (2023). Decision support system enabled digital mobile platform to assist farmers towards agricultural production sustainability in Sri Lanka. *Tropical Agricultural Research and Extension, 26*(1), 28–43. https://doi.org/10.4038/tare.v26i1.5641

Munasingha, M. A. P., & Napagoda, N. A. D. N. (2021). Trend analysis and forecasting for paddy production in Sri Lanka. *Applied Economics & Business, 5*(2), 1–10. https://doi.org/10.4038/aeb.v5i2.33

Nandani, E. J. K. P., & Vidanapathirana, T. T. S. (2024). Forecasting paddy yield in Sri Lanka using back-propagation learning in artificial neural network model. *Journal of the University of Ruhuna, 12*(2), 110–120. https://doi.org/10.4038/jur.v12i2.8032

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems, 24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Rice Research and Development Institute. (n.d.). *Rice the staple food: General information*. Department of Agriculture, Sri Lanka. https://doa.gov.lk/rrdi_rice_introduction-2/

Senevirathne, D. S., & Premasiri, A. C. (2025). AI driven paddy (Oryza sativa) yield forecasting using open satellite data, weather APIs and historical data for Sri Lankan agro zones. In *Fifth International Electronic Conference on Agronomy*. https://sciforum.net/paper/28336

Sivapathasundaram, V., & Bogahawatte, C. (2012). Forecasting of paddy production in Sri Lanka: A time series analysis using ARIMA model. *Tropical Agricultural Research, 24*(1), 21–30. https://doi.org/10.4038/tar.v24i1.7986

van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture, 177*, 105709. https://doi.org/10.1016/j.compag.2020.105709

Zhai, Z., Martinez, J. F., Beltran, V., & Martinez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture, 170*, 105256. https://doi.org/10.1016/j.compag.2020.105256