DETAILED LITERATURE REVIEW OF RESEARCH PAPERS RELATED TO KCC AND PEST PREDICTION

PAPER 1

Deep Learning-based Query Count Forecasting using Farmers' Helpline Data

Godara, Samarth, and Durga Toshniwal. "Deep Learning-based query-count forecasting using farmers' helpline data." Computers and Electronics in Agriculture 196 (2022): 106875

Title: Deep Learning-based Query Count Forecasting using Farmers' Helpline Data

Data Source: Kisan Call Center (Government of India)

Core Focus:

📞 Estimating & forecasting the number of farmer calls over time

Why this paper matters

- Call volume reflects real agricultural issues
- Helps in early detection of crop stress
- Supports policy planning & manpower allocation

Problem

- Traditional agricultural surveys are:
  - Slow
  - Costly
  - Not real-time
- Massive Kisan Call Center data is underutilized

Objective of the Paper

- Use historical call data to:
  - Forecast future query call counts
  - Identify seasonal and regional patterns
- Focused on rice crop across major Indian states

Dataset

- Kisan Call Center call logs
- Time period: 2013 - 2020
- Millions of call records
- States analyzed: Major rice-producing states

Key Challenge

- Data is transactional (one row = one call)
- Forecasting needs time-series format
- Major effort needed in data transformation

Methodology (High-Level Pipeline)

1. Collect raw KCC call logs
2. Clean and merge state-wise data
3. Convert call records → daily call-count time series
4. Train forecasting models
5. Compare machine learning vs deep learning models

Models Used

Models Compared

- Machine Learning
  - Support Vector Regression (SVR)
  - Multi-Layer Perceptron (MLP)
- Deep Learning
  - Long Short-Term Memory (LSTM)
  - Gated Recurrent Unit (GRU)

Forecasting Setup

- Input: Past daily call counts
- Output: Future call counts
- Forecast horizons:
  - Short-term to long-term (days to weeks)

Key Results & Takeaways

Results

1. LSTM performed best across all states
2. Deep learning models:
    - Capture seasonality better
    - Handle long-term trends effectively
3. Clear seasonal peaks aligned with rice cropping cycles

Main Takeaway

📌 Farmer call volume can be treated as a time-series signal

📌 Deep learning enables reliable forecasting of agricultural stress trends

PAPER 2

KisanQRS: A Deep Learning-based Automated Query–Response System

Title: KisanQRS: A Deep Learning-based Automated Query-Response System

- Domain: Agriculture + NLP + Deep Learning
- Data Source: Kisan Call Center (Government of India)
- Core Focus:
- 🤖 Automating farmer query understanding and answer retrieval

Problem

- Millions of farmer queries handled manually
- Issues with current system:
- High response time
- Inconsistent advice
- Poor scalability
- Traditional keyword-based systems:
- Ignore semantic meaning
- Perform poorly on similar but differently worded queries

Motivation

- Need for a scalable, intelligent advisory system
- Must understand intent, not just keywords

Main Objectives

- Automatically group similar farmer queries
- Map a new query to the most relevant group
- Retrieve top relevant answers automatically
- 📌 This is not simple QA
- 📌 It is a cluster-based intelligent query-response system

Methodology (High-Level Pipeline)

1. Collect farmer queries and answers (KCC data)
2. Clean and preprocess text data
3. Cluster semantically similar queries (offline)
4. Train a deep learning model to map new queries to clusters
5. Retrieve and rank the most relevant answers

Models Used

Key Techniques

- Sentence embeddings for semantic understanding
- Hybrid similarity (semantic + lexical)
- Deep learning model for query mapping

Best Performing Model

- LSTM-based model
- Achieved high accuracy in predicting correct query clusters

Key Results & Takeaways

Results

1. High accuracy in mapping queries to correct clusters
2. Fast response time
3. Effective retrieval of relevant answers

Main Takeaway

📌 Deep learning enables semantic understanding of farmer queries

📌 Clustering improves scalability and efficiency

📌 System can reduce dependency on human call-center agents

PAPER 3

AgriMine: Spatio-temporal analysis of farmers’ helpline data

Title: AgriMine: Spatio-temporal analysis of farmers' helpline data

- Dataset: Kisan Call Center (KCC), Government of India
- Time Period: 2013-2021

What this paper does

🌍 Where farmers ask for help

⏱️ When farmers face specific agricultural problems

📈 Forecasts short-term query demand

Problem

- Traditional agricultural surveys:
  - Expensive
  - Slow
  - Not real-time
- Farmer literacy limits use of:
  - Social media
  - Online forums
- Massive helpline data exists but is under-analyzed

Motivation

- Use farmers’ call behavior as a proxy for real agricultural problems
- Support policy, extension, and marketing decisions

Main Objectives

1. Perform spatial analysis
    - Identify regions actively seeking help
2. Perform temporal analysis
    - Identify seasonal phases of problems
3. Introduce step-series
    - Detect exact start & end of problem phases
4. Forecast topic-wise query demand
    - Short-term (7 days)

👉 **This paper is analytics + forecasting, not QA or NLP**

Methodology (High-Level Pipeline)

1. Collect KCC call-log data
2. Clean & preprocess records
3. Spatial analysis
    - Convert addresses → geo-coordinates
    - Generate region-wise maps
4. Temporal analysis
    - Convert calls → daily query time-series
    - Smooth & analyze trends
5. Forecast future query demand

Models Used

Temporal Insights

- Time-series plots of query volume
- Step-series representation
  - Divides time into:
    - No calls
    - Low frequency
    - Medium frequency
    - High frequency
- Automatically extracts:
  - Uprising phase
  - Peak phase
  - Decline phase

Forecasting Models Compared

- FFT
- N-BEATS
- RNN, LSTM, GRU
- TCN (best performing)

Key Results & Takeaways

Key Results

- TCN outperformed all models
- Accurate 7-day forecasting of query demand
- Spatial maps reveal:
  - Region-specific crop interest
- Temporal analysis aligns with:
  - Crop growth stages
  - Seasonal farming activities

Main Takeaway

📌 Deep learning enables semantic understanding of farmer queries

📌 Clustering improves scalability and efficiency

📌 System can reduce dependency on human call-center agents

PAPER 4

Pest and Disease Prediction and Management for Sugarcane Using a Hybrid Autoregressive Integrated Moving Average—A Long Short-Term Memory Model

Focus:

🌾 Predicting sugarcane pest and disease outbreaks using time-series modeling

Key Idea:

- Combine statistical forecasting and deep learning to improve prediction accuracy.

Why Sugarcane?

- Major global crop
- Pest & disease outbreaks cause 20%+ yield loss
- Early prediction = better management

Problem

- Pest and disease occurrence:
    - Highly seasonal
    - Strongly influenced by weather
    - Shows nonlinear behavior
- Traditional models:
    - ARIMA → handles trends but fails on nonlinear patterns
    - LSTM → captures nonlinearity but may miss long-term trends

Motivation

📌 Need a robust hybrid model

📌 Must capture both linear trends and nonlinear fluctuations

Main Objectives

1. Predict future pest and disease incidence in sugarcane
2. Compare:
   - ARIMA
   - LSTM
   - Hybrid ARIMA–LSTM
3. Support early warning and pest management decisions

👉 This is a forecasting & decision-support paper, not detection or image-based disease diagnosis.

Methodology (High-Level Pipeline)

1. Data cleaning & normalization
2. ARIMA models linear trends
3. LSTM models remaining nonlinear patterns
4. Combine both predictions

Models Used

How the Hybrid Model Works

- ARIMA:
  - Captures long-term trends & seasonality
- LSTM:
  - Learns complex nonlinear behavior
- Final prediction = ARIMA output + LSTM correction

Why Hybrid is Better

📌 ARIMA alone → misses nonlinear spikes

📌 LSTM alone → unstable for long trends

📌 Hybrid → balanced and stable

Key Results & Takeaways

Results

- Hybrid ARIMA-LSTM:
  - Lowest prediction error
  - More accurate than ARIMA or LSTM alone
- Performs well during:
  - Stable periods
  - High-fluctuation pest outbreaks

Key Takeaways

📌 Weather data is critical for pest prediction

📌 Hybrid models outperform single models

📌 Useful for early intervention & precision agriculture

PAPER 5

Hybrid Linear Time Series Approach for Long-Term Forecasting of Crop Yield

Focus:

🌾 Long-term forecasting of crop yield for agricultural planning

Crop Studied: Rice

Region: Aligarh district, Uttar Pradesh (India)

Key Contribution:

- **Introduces a hybrid ARIMA + ANN framework to extend short-term forecasts into long-term forecasts (up to 2025).**

Problem

- Crop yield forecasting is crucial for:
  - Food security
  - Policy planning (Vision 2030, 2025, etc.)
- Traditional ARIMA models:
  - Work well only for short-term forecasts
  - Fail for long-term forecasting due to:
    - Mean convergence
    - Inability to capture nonlinear behavior

Motivation

📌Need a statistically sound long-term forecasting method

📌 Must handle both linear trends and nonlinear residual patterns

Main Objectives

1. Improve short-term yield forecasting accuracy
2. Use improved short-term forecasts as baseline
3. Extend forecasts reliably to long-term horizons (up to 2025)
4. Compare:
    - ARIMA alone
    - Hybrid ARIMA-ANN approach

👉 This paper focuses on yield forecasting, not pest or disease prediction.

Methodology (High-Level Pipeline)

1. Apply ARIMA to model linear yield trends
2. Extract residuals from ARIMA model
3. Test residuals for nonlinearity
4. Model residuals using Artificial Neural Network (ANN)
5. Correct ARIMA forecasts using ANN outputs
6. Use corrected forecasts as baseline for long-term prediction

Models Used

- ARIMA (2,1,0) → selected using AIC/BIC
- ANN → used to model nonlinear residuals

Why ANN?

📌 Captures nonlinear relationships

📌 Does not require predefined model structure

Key Results & Takeaways

Results

- ARIMA alone:
    - High forecasting error
- Hybrid ARIMA–ANN:
    - Significant error reduction
    - Much better fit to actual yield values
- Hybrid approach enabled:
    - Forecasting up to 2025
    -

Key Takeaways

📌 Weather data is critical for pest prediction

📌 Hybrid models outperform single models

📌 Useful for early intervention & precision agriculture

PAPER 6

Temporal Graph Convolutional Networks for Predicting Disease Outbreaks in Public Health Surveillance

Focus:

🦠 Predicting disease outbreaks using spatio-temporal deep learning

Core Idea:

Diseases spread over time and space, so the model must capture both dimensions together.

Region:

Aligarh district, Uttar Pradesh (India)

Key Contribution:

- **Introduces Temporal Graph Convolutional Networks (TGCNs) for public-health surveillance.**

Problem

- Disease outbreaks depend on:
    - Temporal trends (weekly/monthly cases)
    - Spatial interactions (movement between regions)
- Traditional models:
    - Time-series models → ignore spatial relations
    - ML models → struggle with dynamic transmission networks

Motivation

📌📌 Need a model that jointly captures

- when diseases spread
- where diseases spread

Main Objectives

1. Develop a new outbreak prediction model using TGCNs
2. Model spatio-temporal disease transmission networks
3. Compare TGCN with existing deep-learning methods
4. Improve early warning capability in public-health systems

👉 This paper focuses on disease outbreak forecasting, not diagnosis.

Methodology (High-Level Pipeline)

1. Overall TGCN Framework
2. Collect disease surveillance data (WHO, CDC)
3. Construct graph structure
4. Nodes → regions/cities
5. Edges → geographic & mobility connections
6. Apply:
7. Graph Convolution → learn spatial dependencies
8. Temporal Convolution → learn time-based trends
9. Fuse spatial + temporal features
10. Predict future outbreak events

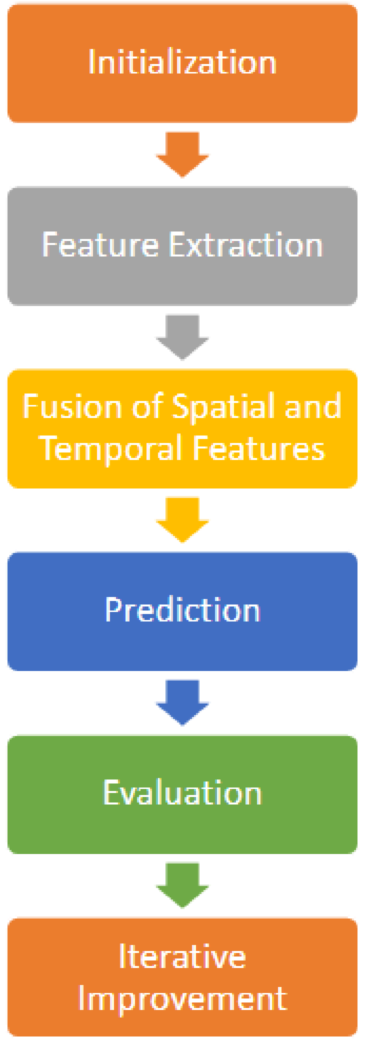

Figure 1. Proposed TGCNs method

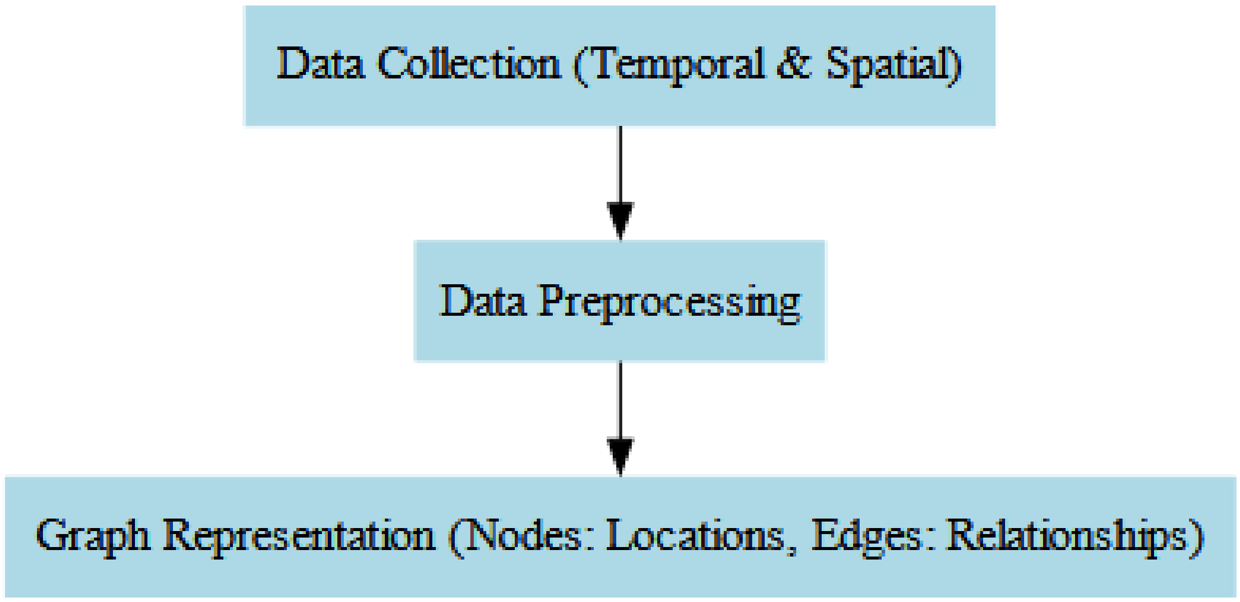

Figure 2. Graph representation

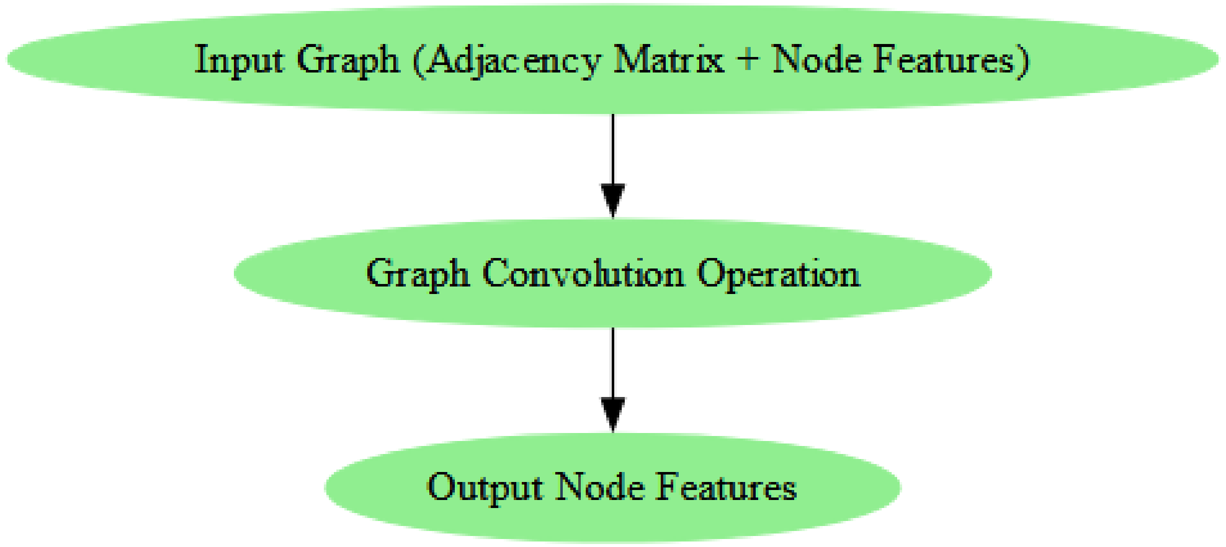

Figure 3. Graph convolution operation

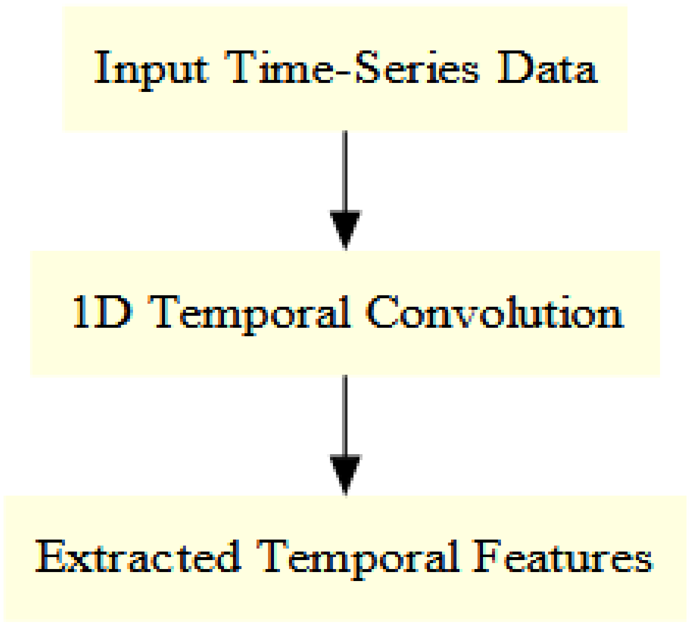

Figure 4. Temporal convolutional layers

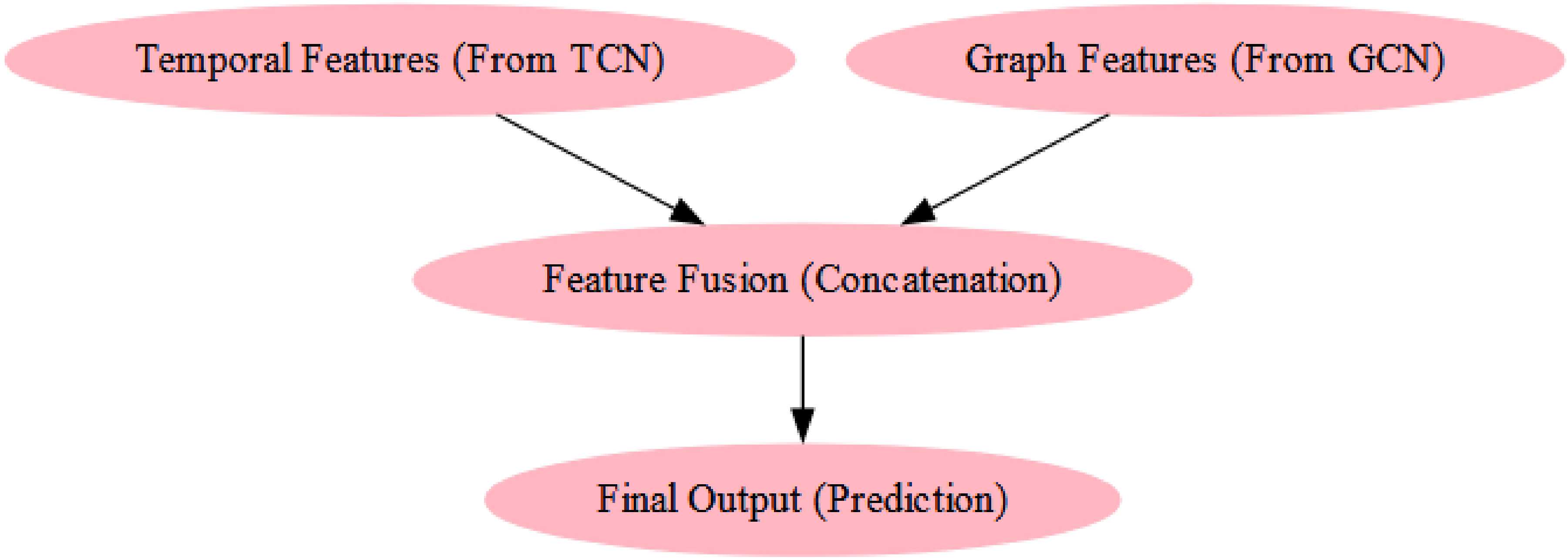

Figure 5. Fusion of spatial and temporal features

Dataset & Model Setup

Dataset

- Diseases studied:
  - Influenza
  - Dengue
  - Measles
- Data sources:
  - WHO
  - CDC
- Features include:
  - Infection rates
  - Population mobility
  - Temperature & humidity

Models Compared

- RNN
- ResNet-based model
- Proposed TGCN

Key Results & Takeaways

Results

- TGCN outperformed:
  - RNN
  - ResNet
- Improvement of 10–15% accuracy
- Better precision, recall, and F1-score
- Efficient execution time despite complex modeling

Key Takeaways

- Spatio-temporal modeling is critical for outbreak prediction
- Graph-based deep learning captures real transmission dynamics
- TGCNs enable early warning and preventive action

THANK YOU