

Leveraging LSTM and Dominating Sets for Precise NPK Nutrient Forecasting in Agriculture

Ika Hesti Agustin*, Qonita Ilmi Awal, Alfian Futuhul Hadi, Dafik, Indah Lutfiyatul Mursyidah, and M. Venkatachalam

Abstract—This research focuses on discovering new theorems regarding dominating sets in graphs, specifically on $P_n \triangleright F_2$ and mP_n , and their application in designing planting patterns for forecasting NPK nutrient needs in crops. The dominating set is a concept in graph theory that plays a crucial role in optimizing various systems, including agriculture. By utilizing planting patterns optimized through dominating set theory, this study proposes a *Graph-Assisted Long Short-Term Memory (LSTM)* model—an adaptation inspired by the Graph-Based LSTM (GLSTM) framework. Unlike conventional GLSTM models that integrate graph adjacency structures directly into LSTM layers, our method applies graph theory as a preprocessing step to determine optimal sensor placement. The selected dominator nodes then serve as inputs to a standard LSTM model, enabling efficient multivariate time-series forecasting of NPK requirements. The results show that this integration not only reduces the number of sensors needed for data collection but also improves the accuracy of nutrient forecasting, with the best Mean Squared Error (MSE) value of 0.00053. This approach offers a novel hybrid framework with the potential to enhance agricultural efficiency and productivity through more optimal resource utilization.

Index Terms—Dominating set, LSTM, precision agriculture, time series forecasting.

I. INTRODUCTION

Dominating set is one of the important topics in graph theory. Let G be a connected graph of order n with vertex set $V(G)$ and edge set $E(G)$ [1], [2], [3]. A set of vertices D is called a dominating set for a graph G if every vertex is either in D or adjacent to a vertex in D [4], [5]. Formally, the vertex set $D \subseteq V(G)$ is a dominating set of graph G if for every vertex $x \in V(G) \setminus D$, there exists at least one vertex $u \in D$ such that $x \sim u$, where the notation $x \sim u$ means that vertex x is adjacent to vertex u [6].

Manuscript received September 3, 2024; revised July 26, 2025. This work was supported by LP2M Universitas Jember and PUI-PT Combinatorics and Graphs (CGANT) at Universitas Jember, whose invaluable support and collaboration in our 2025 research initiatives have been instrumental. Their steadfast encouragement and resources have played a crucial role in driving innovation, advancing our research, and enhancing our academic partnership.

Ika Hesti Agustin is a Lecturer in the Department of Mathematics, University of Jember and PUI-PT Combinatorics and Graph, CGANT-University of Jember, Indonesia (corresponding author, e-mail: ikahesti.fmipa@unej.ac.id).

Qonita Ilmi Awal is a Graduate student in the Master's Program in Mathematics at the University of Jember, Indonesia (e-mail: qonitailmi18@gmail.com).

Alfian Futuhul Hadi is a Lecturer in the Department of Mathematics, University of Jember, Indonesia (e-mail: afhadi@unej.ac.id).

Dafik is a Professor in the Department of Mathematics, University of Jember and PUI-PT Combinatorics and Graph, CGANT-University of Jember, Indonesia (e-mail: d.dafik@unej.ac.id).

Indah Lutfiyatul Mursyidah is a Researcher in the PUI-PT Combinatorics and Graph, CGANT-University of Jember, Indonesia (e-mail: mursyidahindah6@gmail.com).

M. Venkatachalam is a Lecturer in the Department of Mathematics, Kongunadu Arts and Science College, Tamil Nadu, India (e-mail: venkatmaths@kongunaducollege.ac.in).

The minimum cardinality of a dominating set in graph G is called the domination number and is denoted by $\gamma(G)$ [7], [8], [9]. Ore et al. defined the dominating set as a subset D of $V(G)$ for a graph G . Other definitions of dominating sets have also been proposed, such as those by Chartrand, Lesniak, and Haynes et al. Among various definitions, the most commonly accepted is that a set of vertices $D \subseteq V$ in a graph $G = (V, E)$ is a dominating set if every vertex $v \in V \setminus D$ is adjacent to at least one vertex in D .

The domination number $\gamma(G)$ is the minimum number of vertices in a dominating set of G [10], [11]. Many researchers have studied this topic. For example, Booth and Johnson [12] investigated dominating sets in chordal graphs, while Bertossi [13] studied them in split and bipartite graphs. Roifah and Dafik [14] explored dominating sets in special types of graphs, such as complete, cycle, and path graphs. Further studies can be found in [15], [16], [17], [18], [19].

Dominating sets have practical applications in various fields, including agriculture. One such application is in cropping patterns, which involve the efficient arrangement of crops to meet specific objectives. Factors such as planting rates, crop spacing, and crop management are central to cropping pattern design [20]. For optimal plant growth and productivity, fertilizers such as NPK (Nitrogen, Phosphorus, and Potassium) must be used efficiently [21]. This not only promotes healthy plant development but also mitigates environmental issues such as soil and water pollution. Therefore, advanced and accurate forecasting techniques are essential for determining appropriate NPK requirements. This study employs deep learning techniques, specifically Long Short-Term Memory (LSTM), a branch of machine learning that excels in sequential data prediction [22].

LSTM is an advanced variant of the Recurrent Neural Network (RNN), designed to overcome issues like the vanishing gradient problem and loss of historical information during training [23]. It can capture both short-term and long-term dependencies in data, making it well-suited for time series modeling. The LSTM architecture includes three core gates—forget gate, input gate, and output gate—that regulate the flow of information. These gates allow the network to discard irrelevant information and retain essential patterns, thus enhancing accuracy and learning efficiency. In contrast to traditional RNNs that rely solely on functions like tanh, LSTM offers a more robust architecture capable of capturing long-term relationships [24].

The versatility of LSTM models has been demonstrated in numerous domains, achieving state-of-the-art performance in various predictive tasks. For example, hybrid models combining LSTM and ARIMA have improved prediction accuracy for stock prices [24]. In agriculture, LSTM has shown promise in managing water resources and forecasting

nutrient requirements for crops, especially for NPK [25]. In engineering, LSTM has been utilized to estimate the axial compression capacity of CCFST columns [26]. Moreover, Attention-GRU models have introduced attention mechanisms for tasks like predicting NO_x emissions in thermal power plants [27]. Additional studies highlight the importance of parameter tuning in LSTM, especially when working with small datasets, to achieve optimal model performance [28].

The novelty of this study lies in the integration of dominating set theory for optimal sensor placement with a multivariate LSTM model for nutrient prediction. Unlike prior works on Graph-based LSTM (GLSTM) which integrate graph adjacency structures into neural network computations [29], our approach utilizes dominating set theory as a preprocessing step to select representative input points for training a conventional LSTM. This hybrid setup is herein referred to as a *Graph-Assisted LSTM*. Based on the stated motivations, the proposed research questions are:

- 1) What are the dominating set and domination number for the graphs $P_n \supseteq \mathcal{F}_2$ and mP_n ? These graphs are used in the planting design to determine optimal locations for placing NPK sensors.
- 2) How can LSTM be used for NPK forecasting? The aim is to identify the optimal time to reapply NPK so that crops do not suffer from nutrient deficiencies.

II. METHOD

This research employs both analytical and experimental methods. The analytical component uses a deductive mathematical approach to establish the theoretical foundation, while the experimental component involves simulations using computer programming [30], [31]. Nitrogen level forecasting is conducted using the Long Short-Term Memory (LSTM) method. Additionally, the planting process is modeled using dominating set theory to determine optimal sensor placement.

A. Dominating Set

The following definitions related to dominating set theory are used to construct theorems for determining the domination number.

Definition 1: A set $D \subseteq V(G)$ is called a *dominating set* of graph G if every vertex in $V(G) \setminus D$ is adjacent to at least one vertex in D . The minimum cardinality of such a set is called the *domination number*, denoted by $\gamma(G)$. If a dominating set has cardinality $\gamma(G)$, it is called a γ -set [32], [33].

The following theorem provides a lower bound that supports the derivation of new domination numbers in the context of planting design graphs.

Theorem 1: Let G be any graph. Then,

$$\left\lceil \frac{p}{1 + \Delta(G)} \right\rceil \leq \gamma(G) \leq p - \Delta(G),$$

where p is the number of vertices and $\Delta(G)$ is the maximum degree of the graph.

B. Long Short-Term Memory (LSTM)

This study also examines the application of dominating set theory in conjunction with time series forecasting using Long Short-Term Memory (LSTM). The architecture used in this study is illustrated in Fig. 1.

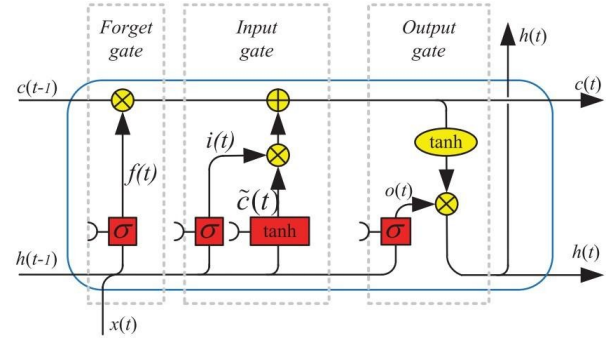


Fig. 1: LSTM model architecture.

Several steps are required in this architecture to optimize performance. The mathematical formulation of the LSTM algorithm is outlined below.

1) *Forget Gate:* The first step in the LSTM unit is to determine which information should be removed from the memory cell. This is achieved using the forget gate, which applies a sigmoid activation function:

$$f_t = \sigma \left(W_f \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_f \right)$$

where f_t is the forget gate vector at time t , σ is the sigmoid activation function, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias term.

2) *Input Gate:* Next, the input gate determines what new information will be stored in the cell state. It consists of two components: a sigmoid function and a candidate value generator.

$$i_t = \sigma \left(W_i \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_i \right)$$

$$\tilde{C}_t = \tanh \left(W_C \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_C \right)$$

where i_t is the input gate vector, $\tilde{C}_t$ is the candidate cell state, W_i and W_C are weight matrices, and b_i and b_C are bias terms.

3) *Update Cell State:* The updated cell state C_t is computed by combining the effects of the forget gate and the input gate:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

where C_{t-1} is the previous cell state.

4) *Output Gate:* Finally, the output gate determines the next hidden state:

$$o_t = \sigma \left(W_o \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_o \right)$$

$$h_t = o_t * \tanh(C_t)$$

where o_t is the output gate vector, h_t is the hidden state output, W_o is the weight matrix, and b_o is the bias.

C. Data and Setup

The dataset used in this study was obtained from Kaggle and consists of second-order time-series data for three variables: nitrogen (N), phosphorus (P), and potassium (K). These variables were measured on 25 plants over 400 days, with measurements recorded every 12 hours. Data normalization was performed using the min-max method to ensure uniformity.

The dataset was divided into three subsets: training data (70%), validation data (15%), and testing data (15%). The LSTM model was trained on the training set and validated to determine optimal hyperparameters. The final model configuration includes two LSTM layers with approximately 100 neurons, a learning rate of 0.001, a batch size of 64, a dropout rate of 0.3, an Adam optimizer, a time step sequence of 10, and a mean squared error (MSE) as the loss function. All computations and simulations were conducted using Google Colab.

III. RESULT AND DISCUSSION

A. Implementation of Dominating Set Theory

Dominating set theory is applied in this study to determine optimal sensor placement within a farmland. This strategic placement enables efficient and accurate data collection while minimizing the number of required sensors. In this section, we present proofs of the domination number for the graph $P_n \triangleright \mathcal{F}_2$ and the graph mP_n . These theorems are used to establish the minimum number of sensors needed for nutrient monitoring.

Theorem 2: Let G be a graph isomorphic to $P_n \triangleright \mathcal{F}_2$. Then, the domination number of $P_n \triangleright \mathcal{F}_2$ is $\gamma(G) = n - 1$.

Proof: Let $P_n \triangleright \mathcal{F}_2$ be a connected graph with vertex set $V(P_n \triangleright \mathcal{F}_2) = \{u_i, x_j, y_k, z_l : 1 \leq i \leq n, 1 \leq j, k, l \leq n - 1\}$, and edge set $E(P_n \triangleright \mathcal{F}_2) = \{u_i u_{i+1}, u_i x_i, u_{i+1} x_i, x_i y_i, x_i z_i, y_i z_i : 1 \leq i \leq n - 1\}$.

The graph contains $4n - 3$ vertices and $6(n - 1)$ edges. The maximum degree in $P_n \triangleright \mathcal{F}_2$ is 4. To construct a minimal dominating set, we examine the neighborhood structure of the graph. By selecting one vertex from each $\mathcal{F}_2$ subgraph, we ensure that every non-dominating vertex is adjacent to at least one vertex in the dominating set. Thus, $\gamma(P_n \triangleright \mathcal{F}_2) \geq n - 1$.

Conversely, since the highest degree occurs at vertices $\{u_i : 2 \leq i \leq n - 1\}$ and $\{x_i : 1 \leq i \leq n - 1\}$, we select $\{x_i : 1 \leq i \leq n - 1\}$ as the dominating set D . It is easy to verify that all vertices u_i, y_k , and z_l are adjacent to at least one vertex in D , implying $\gamma(P_n \triangleright \mathcal{F}_2) \leq n - 1$.

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright \mathcal{F}_2) = n - 1$. The example of dominating set of $P_4 \triangleright \mathcal{F}_2$ can be seen in Fig. 2.

Theorem 3: Let G be a disjoint union of path graphs denoted by mP_n . Then, the domination number of mP_n is $\gamma(G) = m \lceil \frac{n}{3} \rceil$.

Proof: Let mP_n be a disjoint union of m path graphs, each of length n . The vertex set is defined as $V(mP_n) = \{v_{i,j} : 1 \leq i \leq n, 1 \leq j \leq m\}$, and the edge set is $E(mP_n) = \{v_{i,j} v_{i+1,j} : 1 \leq i \leq n - 1, 1 \leq j \leq m\}$. The graph has nm vertices and $m(n - 1)$ edges.

Since the paths are disjoint, the domination number of mP_n is simply the sum of the domination numbers of each

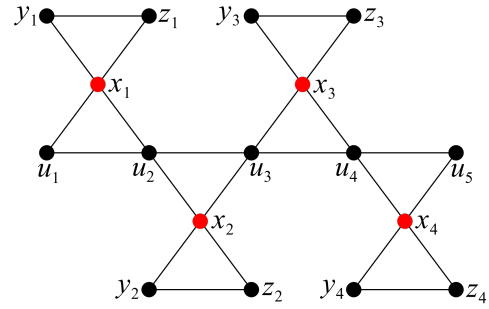


Fig. 2: $\gamma(P_4 \triangleright \mathcal{F}_2)$

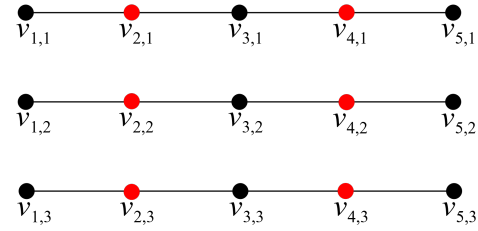


Fig. 3: (a) $3P_4$, (b) $\gamma(3P_4) = 6$

path. The domination number for a single path P_n is $\lceil \frac{n}{3} \rceil$, thus the domination number for m disjoint paths is $m \lceil \frac{n}{3} \rceil$. The example of dominating set of $3P_4$ can be seen in Fig. 3.

Theorem 4: Let G be a graph isomorphic to $P_n \triangleright S_m$. Then, $\gamma(G) = n$.

Proof: Let $P_n \triangleright S_m$ be a connected graph with vertex set $V(P_n \triangleright S_m) = \{x_i; 1 \leq i \leq n\} \cup \{x_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\}$ and edge set $E(P_n \triangleright S_m) = \{x_i x_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\} \cup \{x_i x_{i+1}; 1 \leq i \leq n\}$.

The graph contains $mn + n$ vertices and $mn + n - 1$ edges. The maximum degree in $P_n \triangleright S_m$ is $n + 2$. To construct a minimal dominating set, we examine the neighborhood structure of the graph. By selecting one vertex from each S_m subgraph, we ensure that every non-dominating vertex is adjacent to at least one vertex in the dominating set. Thus, $\gamma(P_n \triangleright S_m) \geq n$.

Conversely, since the highest degree occurs at vertices $\{x_i; 1 \leq i \leq n\}$, we select $\{x_i; 1 \leq i \leq n\}$ as the dominating set D . It is easy to verify that all vertices $x_{i,j}$ are adjacent to one vertex in D , implying $\gamma(P_n \triangleright S_m) \leq n$.

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright S_m) = n$ for $n \geq 2$ and $m \geq 3$. The example of dominating set of $P_n \triangleright S_m$ can be seen in Fig. 4.

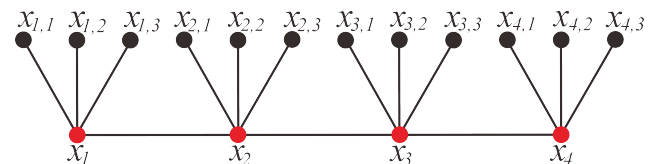


Fig. 4: $\gamma(P_4 \triangleright S_3) = 4$

Theorem 5: Let G be a graph isomorphic to $P_n \triangleright Fr_m$. Then, $\gamma(G) = n$.

Proof: Let $P_n \triangleright Fr_m$ be a connected graph with vertex set $V(P_n \triangleright Fr_m) = \{x_i; 1 \leq i \leq n\} \cup \{y_{i,j}, z_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\}$ and edge set $E(P_n \triangleright Fr_m) = \{x_i y_{i,j}, x_i z_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\} \cup \{y_{i,j} z_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\}$.

$$i \leq n, 1 \leq j \leq m\} \cup \{x_i x_{i+1}; 1 \leq i \leq n\}.$$

The graph contains $2mn + n$ vertices and $3mn + n - 1$ edges. The maximum degree in $P_n \triangleright Fr_m$ is $2n + 2$. To construct a minimal dominating set, we examine the neighborhood structure of the graph. By selecting one vertex from each Fr_m subgraph, we ensure that every non-dominating vertex is adjacent to at least one vertex in the dominating set. Thus, $\gamma(P_n \triangleright Fr_m) \geq n$.

Conversely, since the highest degree occurs at vertices $\{x_i; 1 \leq i \leq n\}$, we select $\{x_i; 1 \leq i \leq n\}$ as the dominating set D . It is easy to verify that all vertices $y_{i,j}$ and $z_{i,j}$ are adjacent to one vertex in D , implying $\gamma(P_n \triangleright Fr_m) \leq n$.

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright Fr_m) = n$ for $n \geq 2$ and $m \geq 2$. The example of dominating set of $P_n \triangleright Fr_m$ can be seen in Fig. 5.

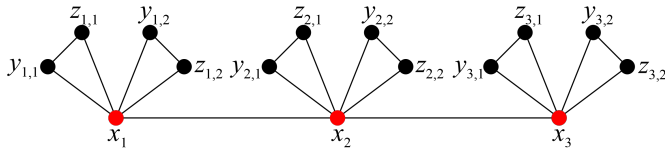


Fig. 5: $\gamma(P_3 \triangleright Fr_2) = 3$

Theorem 6: Let G be a graph isomorphic to $P_n \triangleright C_m$.

$$\text{Then, } \gamma(G) = \begin{cases} n, & \text{for } m = 3 \\ 2\lceil \frac{n}{3} \rceil + \lfloor \frac{n}{2} \rfloor, & \text{for } m = 4 \\ n\lceil \frac{m}{3} \rceil, & \text{for } m \geq 5. \end{cases}$$

Proof: Let $P_n \triangleright C_m$ be a connected graph with vertex set $V(P_n \triangleright C_m) = \{x_{i,j}; 1 \leq i \leq n, 1 \leq j \leq m\}$ and edge set $E(P_n \triangleright C_m) = \{x_{i,j} x_{i,j+1}; 1 \leq i \leq n, 1 \leq j \leq m-1\} \cup \{x_{i,1} x_{i,m}; 1 \leq i \leq n\} \cup \{x_{i,1} x_{i+1,1}; 1 \leq i \leq n-1\}$. The graph contains mn vertices and $mn + n - 1$ edges.

Untuk membuktikan teorema ini, kita akan membagi pembuktian menjadi beberapa kasus yaitu kasus saat $m = 3, m = 4$, dan $m \geq 5$.

Case 1. $m = 3$.

To construct a minimal dominating set, we examine the neighborhood structure of the graph. By selecting one vertex from each C_3 subgraph, we ensure that every non-dominating vertex is adjacent to at least one vertex in the dominating set. Thus, $\gamma(P_n \triangleright C_3) \geq n$.

Conversely, since the highest degree occurs at vertices $\{x_{i,1}; 1 \leq i \leq n\}$, we select $\{x_{i,1}; 1 \leq i \leq n\}$ as the dominating set D . It is easy to verify that all vertices $x_{i,2}$ and $x_{i,3}$ are adjacent to one vertex in D , implying $\gamma(P_n \triangleright C_3) \leq n$.

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright C_m) = n$ for $n \geq 2$ and $m = 3$.

Case 2. $m = 4$.

To construct a minimal dominating set, we examine the neighborhood structure of the graph. By selecting two vertices from each odd-indexed C_4 subgraph and one vertex from each even-indexed C_4 subgraph, we ensure that every non-dominating vertex is adjacent to at least one vertex in the dominating set. Thus,

$$\gamma(P_n \triangleright C_4) \geq 2\lceil \frac{n}{3} \rceil + \lfloor \frac{n}{2} \rfloor.$$

Next, we will prove the upper bound of $\gamma(P_n \triangleright C_4)$ by defining a dominating set D . We define $D = \{x_{i,1}, x_{i,3}; i \equiv$

$1 \pmod{2}, 1 \leq i \leq n\} \cup \{x_{i,3}; i \equiv 0 \pmod{2}, 2 \leq i \leq n\}$. It is easy to verify that all vertices $V(P_n \triangleright C_4) - D$ are adjacent to one vertex in D , implying

$$\gamma(P_n \triangleright C_4) \leq 2\lceil \frac{n}{3} \rceil + \lfloor \frac{n}{2} \rfloor.$$

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright C_m) = \gamma(P_n \triangleright C_4) \leq 2\lceil \frac{n}{3} \rceil + \lfloor \frac{n}{2} \rfloor$ for $n \geq 2$ and $m = 4$. The example of dominating set of $P_n \triangleright C_4$ can be seen in Fig. 6.

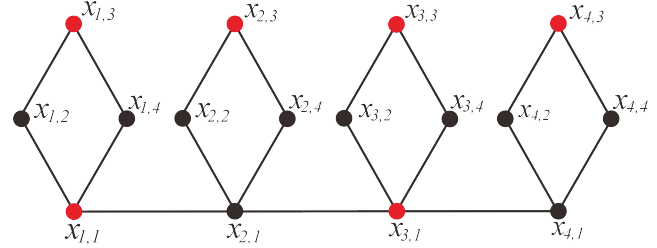


Fig. 6: $\gamma(P_4 \triangleright C_4) = 6$

Case 3. $m \geq 5$.

To construct a minimal dominating set, we examine the neighborhood structure of the graph. We know that $P_n \triangleright C_m$ is a graph obtained by taking one copy of P_n and $|V(P_n)|$ copies of C_m and grafting the i -th copy of C_m at the vertex x_1 to the x_1 -th vertex of P_n . $\gamma(C_n) = \lceil \frac{n}{3} \rceil$. So we have $\gamma(P_n \triangleright C_m) \geq n\lceil \frac{m}{3} \rceil$ for $n \geq 2$ and $m \geq 5$.

Next, we will prove the upper bound of $\gamma(P_n \triangleright C_m)$ by defining a dominating set D . We define $D = \{x_{i,j}; 1 \leq i \leq n, j \equiv 1 \pmod{3}, 1 \leq j \leq m\}$. It is easy to verify that all vertices $V(P_n \triangleright C_m) - D$ are adjacent to one vertex in D , implying

$$\gamma(P_n \triangleright C_m) \leq n\lceil \frac{m}{3} \rceil.$$

Since both bounds are equal, we conclude that $\gamma(P_n \triangleright C_m) = n\lceil \frac{m}{3} \rceil$ for $n \geq 2$ and $m \geq 5$. ■

Precision agriculture has become an increasingly important topic, as it enables farmers to make informed decisions regarding agricultural strategies. In this study, the proposed model leverages a structured dataset to recommend optimal cropping patterns that minimize the number of NPK sensors required for monitoring. An example of horizontal farming implementation is depicted in Fig. 7, where the planting pattern is modeled using the graph $P_4 \triangleright F_2$.

In this graph, each vertex represents a crop. By applying dominating set theory, seven dominator vertices are identified, each representing a crop that will be equipped with a sensor. Forecasting of N, P, and K nutrient levels is then conducted specifically for these dominator vertices.

B. Implementation of LSTM

Assuming that each dominator on the farm has been identified, the following input data can be used

$$x_t = \begin{bmatrix} 0.5 \\ 0.1 \end{bmatrix}$$

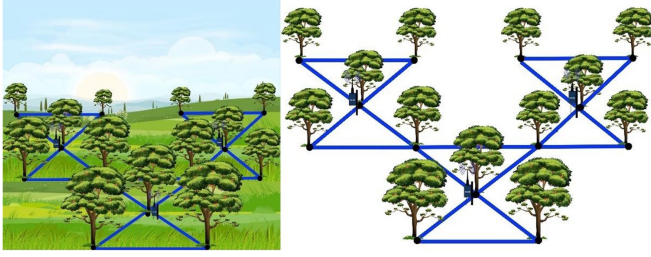


Fig. 7: The Illustration of Horizontal Farming using Dominating Set

1) *Weight and Bias Initialization*: Assume we have two units in the LSTM layer. So, we will have the following weight and bias matrix:

$$W_f = \begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix}, \quad b_f = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}, \quad W_c = \begin{bmatrix} 0.3 & 0.4 \\ 0.5 & 0.6 \\ 0.7 & 0.8 \end{bmatrix},$$

$$b_f = \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix}, \quad W_f = \begin{bmatrix} 0.4 & 0.5 \\ 0.6 & 0.7 \\ 0.8 & 0.9 \end{bmatrix}, \quad b_f = \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix}$$

Initialize the state of the hidden cells and outputs:

$$C_{t-1} = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}, \quad h_{t-1} = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$$

2) *Steps of LSTM*:

a. Calculate forget gate f_t :

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ &= \sigma \left[\begin{bmatrix} 0.1 & 0.2 \\ 0.3 & 0.4 \\ 0.5 & 0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix} \right] \\ &= \sigma \left[\begin{bmatrix} 0.12 \\ 0.26 \\ 0.4 \end{bmatrix} \right] = \begin{bmatrix} 0.5300 \\ 0.5650 \\ 0.5987 \end{bmatrix} \end{aligned}$$

b. Calculate the input gate i_t :

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ &= \sigma \left[\begin{bmatrix} 0.2 & 0.3 \\ 0.4 & 0.5 \\ 0.6 & 0.7 \end{bmatrix} \cdot \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.2 \\ 0.2 \end{bmatrix} \right] \\ &= \sigma \left[\begin{bmatrix} 0.25 \\ 0.39 \\ 0.53 \end{bmatrix} \right] = \begin{bmatrix} 0.5622 \\ 0.5960 \\ 0.6293 \end{bmatrix} \end{aligned}$$

c. Calculate the new candidate value C_t :

$$\begin{aligned} C_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ &= \tanh \left[\begin{bmatrix} 0.3 & 0.4 \\ 0.5 & 0.6 \\ 0.7 & 0.8 \end{bmatrix} \cdot \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.3 \\ 0.3 \end{bmatrix} \right] \\ &= \tanh \left[\begin{bmatrix} 0.35 \\ 0.56 \\ 0.77 \end{bmatrix} \right] = \begin{bmatrix} 0.3366 \\ 0.5086 \\ 0.6293 \end{bmatrix} \end{aligned}$$

d. Update cell state C_t :

$$\begin{aligned} C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\ &= \begin{bmatrix} 0.5300 \\ 0.5650 \\ 0.5987 \end{bmatrix} * \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.5622 \\ 0.5960 \\ 0.6293 \end{bmatrix} * \begin{bmatrix} 0.3366 \\ 0.5086 \\ 0.6479 \end{bmatrix} \\ &= \begin{bmatrix} 0.1892 \\ 0.3033 \\ 0.4076 \end{bmatrix} \end{aligned}$$

e. Calculate output gate o_t :

$$\begin{aligned} o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ &= \sigma \left[\begin{bmatrix} 0.4 & 0.5 \\ 0.6 & 0.7 \\ 0.8 & 0.9 \end{bmatrix} \cdot \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} + \begin{bmatrix} 0.4 \\ 0.4 \end{bmatrix} \right] \\ &= \sigma \left[\begin{bmatrix} 0.45 \\ 0.71 \\ 0.97 \end{bmatrix} \right] = \begin{bmatrix} 0.6106 \\ 0.6700 \\ 0.7252 \end{bmatrix} \end{aligned}$$

f. Calculate output h_t :

$$\begin{aligned} h_t &= o_t * \tanh(C_t) \\ &= \begin{bmatrix} 0.6106 \\ 0.6700 \\ 0.7252 \end{bmatrix} * \tanh \begin{bmatrix} 0.1892 \\ 0.3033 \\ 0.4076 \end{bmatrix} \\ &= \begin{bmatrix} 0.6106 & 0.1871 & 0.1142 \\ 0.6700 & 0.2948 & 0.1974 \\ 0.7252 & 0.3866 & 0.2805 \end{bmatrix} \end{aligned}$$

In this study, a Graph-Assisted Long Short-Term Memory (LSTM) model is utilized to forecast future values in a multivariate time series, specifically nutrient levels in agricultural soil. The "graph-assisted" aspect refers to the use of dominating set theory as a preprocessing step to determine optimal sensor placements. Only the sensor locations identified as dominator nodes are used as input sources for the LSTM model, thereby reducing data dimensionality and optimizing input relevance.

The forecasting performance of the LSTM depends on several key processes. The first step involves analyzing the characteristics or patterns within the time series data, such as trends, seasonality, and noise, all of which significantly influence the model's predictive capability. In this study, the nutrient time series (nitrogen, phosphorus, and potassium) exhibit seasonal patterns, as the concentration levels in the soil tend to decrease approximately every ten days.

After the pattern analysis, the dataset is divided into three subsets: training, validation, and testing. A larger proportion is allocated to the training set to enable the model to learn more effectively from the available data, thus improving its generalization and forecasting accuracy. The integration of dominating set theory ensures that the input fed into the LSTM model is both representative and optimized, aligning with the objectives of precision agriculture and resource efficiency.

Before implementing the LSTM model, it is essential to assess the correlation between the features to be predicted. Correlation is a statistical metric used to quantify the strength and direction of the relationship between two or more variables. The correlation results for nitrogen

TABLE I: Correlation Value of N, P, K Features

	N	P	K
N	1.000	0.985	0.989
P	0.985	1.000	0.987
K	0.989	0.987	1.000

(N), phosphorus (P), and potassium (K) are presented in Table I. As shown, the correlation between N and P is approximately 0.985, indicating a very strong positive relationship. Similarly, the correlation between N and K is about 0.989, while that between P and K is approximately 0.987, both also representing strong positive associations. These results suggest that increases in one nutrient level are likely to coincide with increases in the others. Overall, the high correlation coefficients (close to 1) confirm that the three features tend to move together over time. Following this analysis, the prediction process using the LSTM model is conducted.

Furthermore, the construction of sequential data for LSTM input is commonly referred to as a *time step*, which represents the number of previous time intervals used to predict the next value in a time series. In this study, the term *time step* is equivalent to the *window size*, which defines the length of historical data observed for forecasting. A window size of 5 is utilized, meaning the model makes predictions based on the five previous time steps.

The proposed model consists of three main components: an input layer, an LSTM layer, and two dense layers. The input layer serves as the entry point and specifies the shape of the input data. In this case, the input shape is (5, 3), indicating a time window of 5 steps with 3 features (N, P, and K).

The LSTM layer, as the core component of the model, processes the sequential data and captures long-term dependencies. This layer contains 20 LSTM units, each responsible for learning temporal patterns from the input sequences.

The model includes two dense layers: a hidden dense layer and an output dense layer. The hidden dense layer comprises 10 fully connected units activated by the ReLU (Rectified Linear Unit) function, which introduces non-linearity into the model. The output dense layer has 3 units, each corresponding to a predicted value for N, P, and K. This layer uses a linear activation function, suitable for continuous-valued outputs typically required in forecasting tasks.

In summary, the model receives input data with a shape of (5, 3), processes it through the LSTM and dense layers, and generates numerical predictions with an output dimension of 3.

The training process is carried out using the training and validation datasets to evaluate model performance. Key hyperparameters used include activation functions (ReLU and linear), number of epochs, learning rate, and optimizer. In this study, variations in the number of epochs and LSTM units were tested to identify the optimal configuration.

After training and validation, the model was evaluated using the testing dataset. Among the tested configurations, the best result was achieved with 150 epochs and 20 LSTM units. The model was tested on 50 data points representing

15% of the total dataset allocated for testing. The results indicate that the model effectively captured the underlying patterns, with predicted values closely aligning with the actual data, demonstrating high forecasting accuracy.

Multivariate forecasting of nitrogen (N), phosphorus (P), and potassium (K) content was performed using the LSTM model, configured with a window size of 5, 20 neurons, and 150 training epochs. This configuration yielded a low training error of 0.00053, indicating high model accuracy. A window size of 5 corresponds to forecasting nutrient values five days ahead. The prediction results for the dominator node X2 closely followed the actual data trend, which exhibited a consistent downward pattern. This alignment is illustrated by the four predicted data points shown in Fig. 8.

To evaluate the effectiveness of the proposed model, a comprehensive performance comparison was conducted against several benchmark approaches, including LSTM-ARIMA [24], conventional LSTM [18], CNN-LSTM [20], and Attention-GRU [27]. The comparison results are presented in Table II. Across all performance metrics—Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Accuracy, and the coefficient of determination (R^2)—the Graph-Assisted LSTM consistently outperformed the other methods. In particular, it achieved the lowest RMSE (4.5938) and MAE (3.8235), as well as the highest accuracy (0.8335) and R^2 value (0.7253). These results demonstrate the superiority of our approach in forecasting and monitoring NPK nutrient requirements in crops, validating the effectiveness of incorporating dominating set theory for optimized sensor input selection. A visualization comparing the proposed method with other models is presented in Fig. 9.

Advanced models such as Gated Recurrent Unit (GRU) and Temporal Fusion Transformer (TFT) emerge as close competitors to LSTM, exhibiting relatively low RMSE values and demonstrating the ability to capture long-term dependencies in sequential data. However, both models still fall short of the final performance achieved by the proposed *Graph-Assisted LSTM*, which consistently yields the lowest RMSE and MAE values, along with the highest accuracy and R^2 across all evaluated metrics. This improvement is attributed to the integration of dominating set theory, which effectively selects the most representative and informative sensor locations—reducing noise and redundancy in the input space—before training the LSTM. In contrast, traditional statistical models such as Auto-Regressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR) display significantly higher error values, highlighting their limited capacity to handle nonlinearities and complex temporal dynamics inherent in multivariate agricultural datasets.

Additionally, their convergence behavior is slower and less stable, as observed from the training loss curves, which contrasts with the smoother and more consistent declines demonstrated by the advanced deep learning methods. The graphical comparison further reinforces the robustness of the Graph-Assisted LSTM model, as its RMSE steadily decreases across epochs, reflecting its strong generalization and learning efficiency. These results underscore the importance of combining graph-theoretic optimization with deep learning, positioning the proposed model as a superior

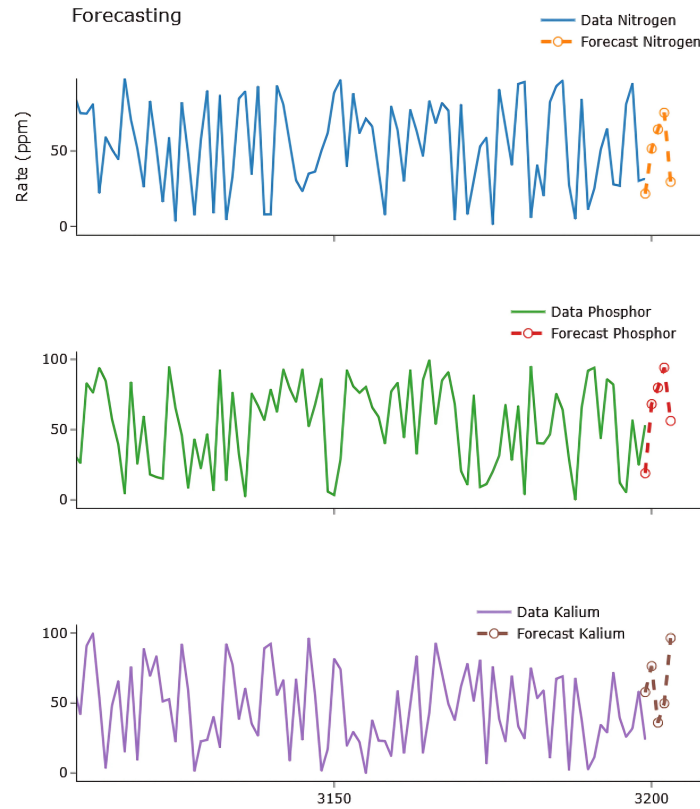


Fig. 8: NPK Forecasting Result for 5 Days Ahead on Dominator X2

TABLE II: The Comparison of Graph-Assisted LSTM with Other Methods

Metric	LSTM [18]	CNN-LSTM [20]	LSTM-ARIMA [24]	Attention-GRU [27]	Graph-Assisted LSTM (Our Work)
RMSE	8.2030	6.7632	5.4038	5.0251	4.5938
MAE	6.0362	5.4523	4.2103	3.9856	3.8235
Accuracy	0.4836	0.6745	0.7832	0.8123	0.8335
R^2	0.1623	0.5732	0.6891	0.7123	0.7253

solution for forecasting nutrient requirements in precision agriculture.

The integration of dominating set theory and Long Short-Term Memory (LSTM) in this research is formulated as a *Graph-Assisted LSTM* framework, presenting a novel and practical approach to optimizing precision agriculture systems. Rather than modifying the internal structure of the LSTM, this approach leverages graph-theoretic principles—specifically, dominating set theory—to preprocess spatial data and select a minimal yet representative subset of sensor nodes. Based on the established theorems and graph configurations ($P_n \triangleright F_2$ and mP_n), the minimal number of dominator nodes is identified, leading to an efficient placement strategy. This optimization is not only mathematically valid but also impactful in reducing hardware costs and energy consumption in field implementations.

Furthermore, the forecasting accuracy achieved by the Graph-Assisted LSTM model demonstrates strong predictive capabilities, achieving a minimum Mean Squared Error (MSE) of 0.00053. When compared to other state-of-the-art models such as ARIMA, SVR, GRU, and TFT, the Graph-Assisted LSTM outperformed all in terms of RMSE, MAE, Accuracy, and R^2 . These quantitative metrics are

supported by visual comparisons in Fig. 9 and the summary in Table II, reinforcing the statistical and practical significance of the model.

In real-world applications, this hybrid integration has direct implications for sustainable agriculture. By minimizing sensor deployment while improving prediction precision, the proposed method contributes to better nutrient management, reduced environmental impact, and increased crop productivity. Additionally, the strong correlation among N, P, and K variables supports the robustness of the multivariate forecasting model in learning temporal dependencies effectively, especially when informed by graph-theoretic input optimization.

IV. CONCLUSIONS

In this study, we investigated the dominating sets of the graphs $P_n \triangleright F_2$, mP_n , $P_n \triangleright S_m$, $P_n \triangleright Fr_m$, and $P_n \triangleright C_m$, and successfully determined the exact values of their domination numbers $\gamma(G)$. It is important to note that determining the domination number is generally a computationally challenging task, particularly for graphs of unbounded order, as it belongs to the class of NP-hard problems.

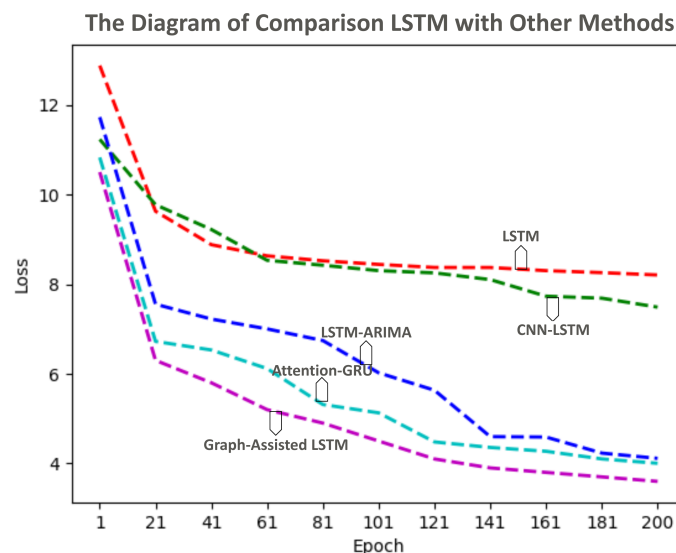


Fig. 9: The Diagram of Comparison LSTM with Other Methods

Building upon these theoretical foundations, we proposed a novel *Graph-Assisted Long Short-Term Memory (LSTM)* framework that integrates graph-theoretic optimization with deep learning for forecasting soil nutrient levels, specifically nitrogen (N), phosphorus (P), and potassium (K). In this framework, dominating set theory is employed to identify the most efficient sensor placements, thereby reducing input dimensionality and enabling a more focused and informative data representation for time-series modeling. The selected dominator nodes serve as inputs to a multivariate LSTM model trained to predict nutrient levels over time.

Experimental results demonstrated the superior performance of the Graph-Assisted LSTM, with the best configuration (150 training epochs and 20 LSTM units) achieving a minimum mean squared error (MSE) of 0.00053. This confirms the model's ability to capture temporal dependencies effectively while maintaining high prediction accuracy and resource efficiency. The integration of graph-theoretic input selection and LSTM forecasting not only enhances computational performance but also supports sustainable precision agriculture by minimizing sensor usage without compromising predictive quality.

For future work, we propose several open research directions: (i) identifying dominator structures in other graph classes and characterizing their properties; (ii) integrating adaptive graph models that evolve based on field dynamics; and (iii) extending the current framework toward multivariate time-series forecasting using hybrid architectures that combine LSTM with Graph Neural Networks (GNN) to further improve performance in complex spatio-temporal domains.

REFERENCES

- [1] Alfarisi, R., Prihandini, R.M., Adawiyah, R., Albirri, E.R., and Agustin, I.H., "Graceful chromatic number of unicyclic graphs," *Journal of Physics: Conference Series*, vol. 1306, no. 1, pp. 012039, 2019.
- [2] Dafik, Agustin, I.H., Surahmat, Alfarisi, R., and Sy, S., "On non-isolated resolving number of special graphs and their operations," *Far East Journal of Mathematical Sciences*, vol. 102, pp. 2473–2492, 2017.
- [3] Septory, B.J., Utoyo, M.I., Sulistiyono, B., and Agustin, I.H., "On rainbow antimagic coloring of special graphs," *Journal of Physics: Conference Series*, vol. 1836, no. 1, pp. 012016, 2021.
- [4] Agustin, I.H., and Dafik, "On the domination number of some families of special graphs," in *Prosiding Seminar Matematika dan Pendidikan Matematika Universitas Jember*, vol. 1, no. 1, pp. 139–143, 2014.
- [5] Booth, K.S., and Johnson, J.H., "Dominating sets in chordal graphs," *SIAM Journal on Computing*, vol. 11, no. 1, pp. 191–199, 1982.
- [6] Bouamama, S., and Blum, C., "An improved greedy heuristic for the minimum positive influence dominating set problem in social networks," *Algorithms*, vol. 14, no. 3, 79, 2021.
- [7] Allan, R.B., and Laskar, R., "On domination and independent domination numbers of a graph," *Discrete Mathematics*, vol. 23, no. 2, pp. 73–76, 1978.
- [8] Dafik, Agustin, I.H., and Prihandoko, A.C., "On the resolving strong domination number of corona and Cartesian product of graphs," *Palestine Journal of Mathematics*, vol. 10, Special Issue II, pp. 169–177, 2021.
- [9] Roifah, M., and Dafik, "Kajian himpunan dominasi pada graf khusus dan operasinya," in *Prosiding Seminar Matematika dan Pendidikan Matematika*, vol. 1, no. 5, pp. 191–196, 2014.
- [10] Boutin, D.L., "Determining sets, resolving sets, and the exchange property," *Graphs and Combinatorics*, vol. 25, no. 6, pp. 789–806, 2009.
- [11] Chartrand, G., and Lesniak, L., *Graphs and Digraphs*, Wadsworth & Brooks, 1986.
- [12] Booth, K.S., and Johnson, J.H., "Dominating sets in chordal graphs," *SIAM Journal on Computing*, vol. 11, no. 1, pp. 191–199, 1982.
- [13] Bertossi, A.A., "Dominating sets for split and bipartite graphs," *Information Processing Letters*, vol. 19, no. 1, pp. 37–40, 1984.
- [14] Vaidya, S.K., and Kothari, N.J., *Distance-k Domination in Some Cycle Related Graphs*, Miskolc University Press, 2018.
- [15] Anto, K.A., and Vetha, J.V.A., "Center concepts on distance-k dominating sets," *International Journal of Computer Applications Technology and Research*, 2017.
- [16] Drewel, G.I., and Al-Bahadili, R.J., "Air pollution prediction using LSTM deep learning and metaheuristics algorithms," *Measurement: Sensors*, vol. 24, 2022.
- [17] Gonçalves, D., Pinlou, A., Rao, R., and Thomassé, S., "The domination number of grids," *SIAM Journal on Discrete Mathematics*, 2011.
- [18] Lin, Z., Shi, Y., Chen, B., Liu, S., Ge, Y., Jien, M., Yang, L., and Lin, Z., "Early warning method for power supply service quality based on three-way decision theory and LSTM neural network," *Energy Reports*, 2021.
- [19] Wardani, D.A.R., Dafik, Agustin, I.H., Marsidi, and Kurniawati, E.Y., "The relation of $\gamma(G \circ H)$ and $\gamma(G * H)$," *IOP Conference Series: Journal of Physics*, 2017.
- [20] Hou, J., Wang, Y., Hou, B., Zhou, J., and Tian, Q., "Spatial simulation and prediction of air temperature based on CNN-LSTM," *Applied Artificial Intelligence*, vol. 37, no. 1, pp. 2166235, 2023.
- [21] Belavadi, S.V., Rajagopal, S.R., Ranjani, and Mohan, R., "Air quality forecasting using LSTM RNN and wireless sensor networks,"

- in *International Conference on Ambient Systems, Networks and Technologies*, 2020.
- [22] Drewel, G.I., and Al-Bahadili, R.J., "Air pollution prediction using LSTM deep learning and metaheuristics algorithms," *Measurement: Sensors*, vol. 24, 2022.
 - [23] Lu, Z., Lv, W., Cao, Y., Xie, Z., Peng, H., and Du, B., "LSTM variants meet graph neural networks for road speed prediction," *Neurocomputing*, 2020.
 - [24] Ma, C., Wu, J., Hu, H., Chen, Y.N., and Li, J.Y., "Predicting stock prices using hybrid LSTM and ARIMA model," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 3, pp. 424–432, 2024.
 - [25] Sarjerao, J.S., and Sudhagar, G., "Water resource optimization by using a hyper parameter tuned LSTM of a smart agriculture," in *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, pp. 1–11, 2024.
 - [26] Cheng, Y., Tao, H., Lu, X., Xu, Z., Fu, H., and Zhu, C., "Prediction method of axial compression capacity of CCFST columns based on deep learning," *IAENG International Journal of Computer Science*, vol. 51, no. 3, pp. 243–251, 2024.
 - [27] Gao, X., Xue, C., Jiang, W., and Liu, B., "Attention-GRU based intelligent prediction of NOx emissions for the thermal power plants," *IAENG International Journal of Computer Science*, vol. 51, no. 8, pp. 1171–1181, 2024.
 - [28] B  rot, O.S., Canot, H., Durand, P., Hassoune-Rhabbour, B., Acheritobehere, H., Laforet, C., and Nassiet, V., "Choice of parameters of an LSTM network, based on a small experimental dataset," *Engineering Letters*, vol. 32, no. 1, pp. 59–71, 2024.
 - [29] Kumar, S., Mallik, A., and Panda, B.S., "Influence maximization in social networks using transfer learning via graph-based LSTM," *Expert Systems with Applications*, vol. 212, pp. 118770, 2023.
 - [30] Dafik, Mursyidah, I.L., Agustin, I.H., Baihaki, R.I., Febrinanto, F.G., Husain, S., Binti, S.K., and Sunder, R., "On rainbow vertex antimagic coloring and its application on STGNN time series forecasting on subsidized diesel consumption," *IAENG International Journal of Applied Mathematics*, vol. 54, no. 5, pp. 984–996, 2024.
 - [31] Kristiana, A.I., Rachmasari, E., Dafik, Agustin, I.H., Mursyidah, I.L., and Alfarisi, R., "On b-coloring analysis of graphs: An application to spatial-temporal graph neural networks for multi-step time series forecasting of soil moisture and pH in companion farming," *European Journal of Pure and Applied Mathematics*, vol. 17, no. 4, pp. 3356–3369, 2024.
 - [32] Agustin, I.H., Hasan, M., Adawiyah, R., Alfarisi, R., and Wardani, D.A.R., "On the locating edge domination number of comb product of graphs," *Journal of Physics: Conference Series*, vol. 1022, no. 1, pp. 012003, 2018.
 - [33] Gembong, A.W., and Agustin, I.H., "Bound of distance domination number of graph and edge comb product graph," *Journal of Physics: Conference Series*, vol. 855, no. 1, pp. 012014, 2017.