# Forecasting of Air Quality with Machine Learning

**DANIEL SAMWEL MAKALA**

*Ministry of Energy*

*Tanzania*

*https://www.linkedin.com/in/daniel-makala-67581230/*

# Traditional vs Machine Learning Approaches

## Traditional Statistical Methods

- Linear regression
- ARIMA models
- Limited pattern recognition

## Machine Learning Models

- Capture complex relationships
- Learn from historical data
- Adapt to changing patterns

# Why Air Quality Prediction?

### Increasing urbanization

Higher population density in cities

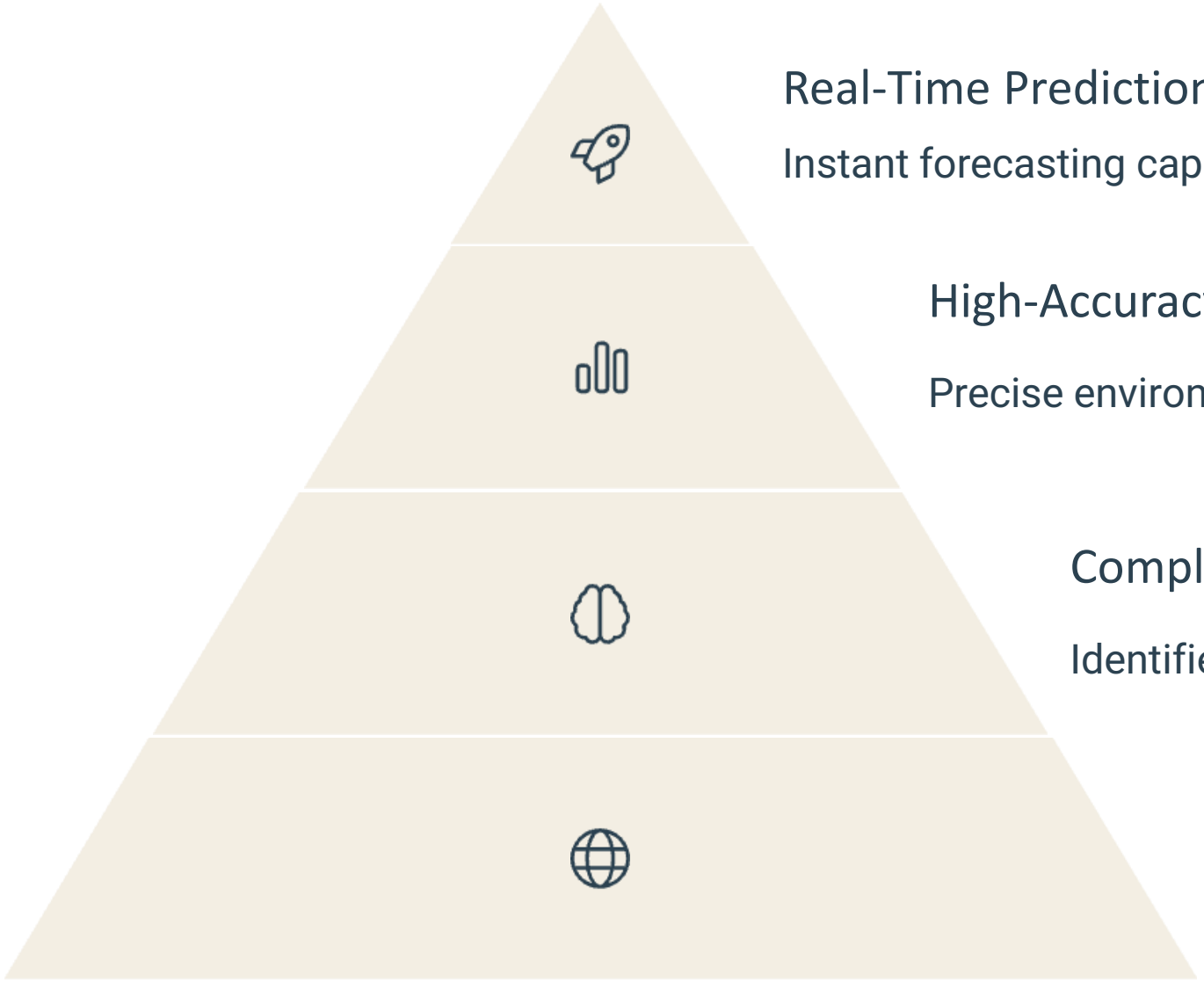
### Industrial activities

Manufacturing and power generation emissions

### Population growth

More people means increased pollution sources

# Why Machine Learning Matters



### Real-Time Prediction

Instant forecasting capabilities

### High-Accuracy Modeling

Precise environmental projections

### Complex Pattern Recognition

Identifies subtle environmental relationships

### Geographic Scalability

Adaptable across diverse regions

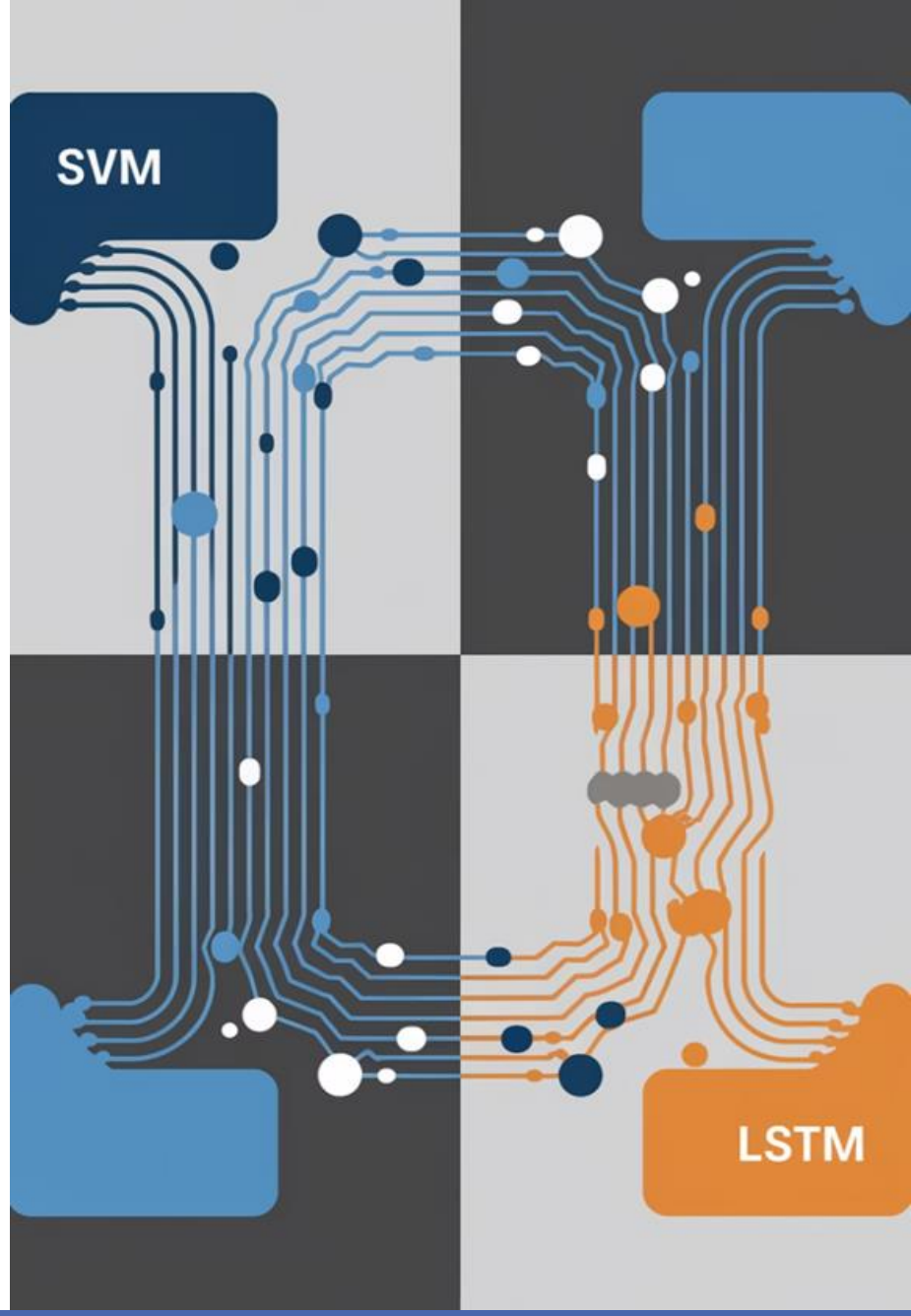
# Study Focus: SVM vs LSTM

## Support Vector Machine (SVM)

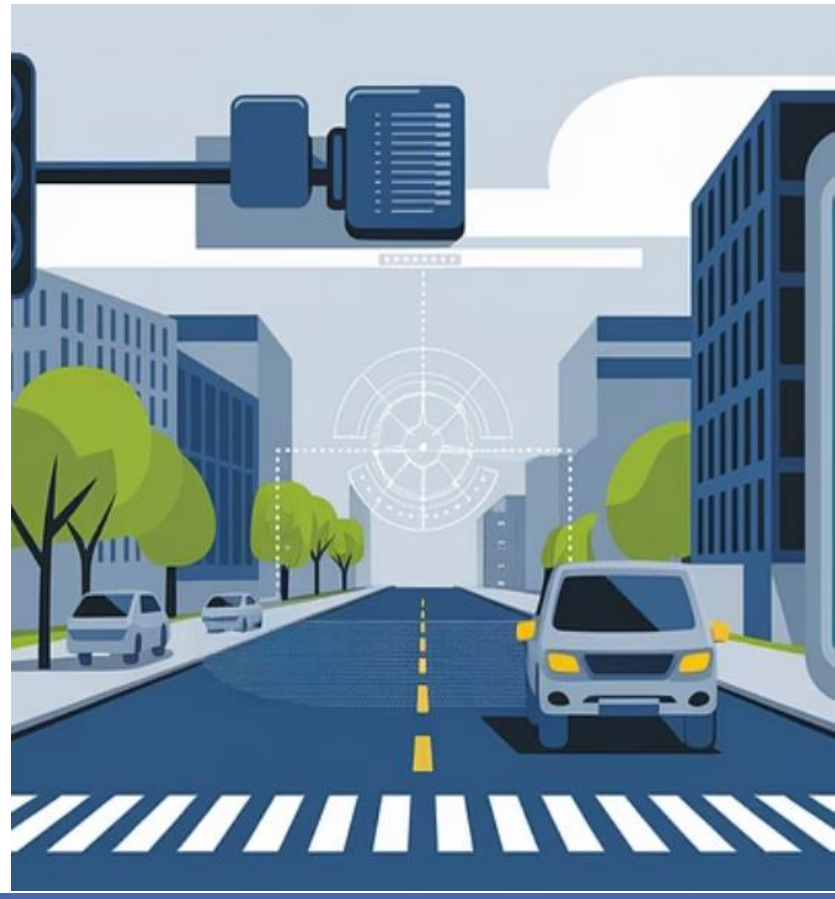
- Classification algorithm
- Identifies optimal boundaries
- Less specialized for time series

## Long Short-Term Memory Memory (LSTM)

- Recurrent neural network
- Designed for sequence data
- Remembers long-term dependencies

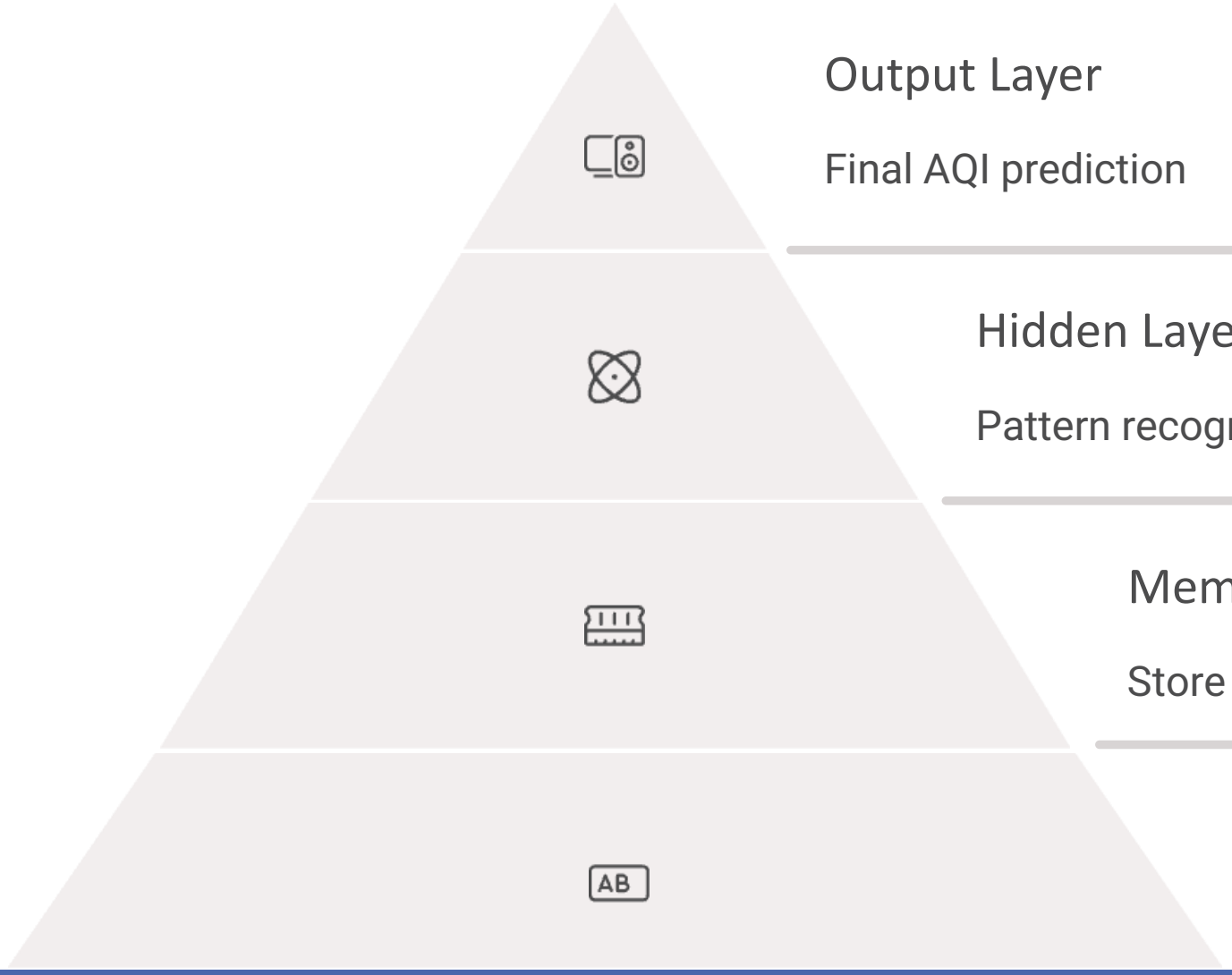


# Data Collection Site: Dali, Taiwan



Hourly AQI data captured using sophisticated monitoring network across Dali - Taiwan

# LSTM Architecture




Output Layer

Final AQI prediction

Hidden Layers

Pattern recognition

Memory Cells

Store temporal patterns

Input Layer

Historical AQI data

# SVM Architecture

## Input Data

Historical AQI measurements

## Kernel Transform

Maps data to higher dimension

## Hyperplane

Finds optimal decision boundary

## Prediction

Forecasts future AQI values

# Model Performance Metrics

| Metric | LSTM | SVM |
| --- | --- | --- |
| RMSE | 3.7 | 43.24 |
| MAE | 2.4 | 32.4 |


# RMSE Comparison

3.7

LSTM RMSE

Lower is better

4.13

SVM RMSE

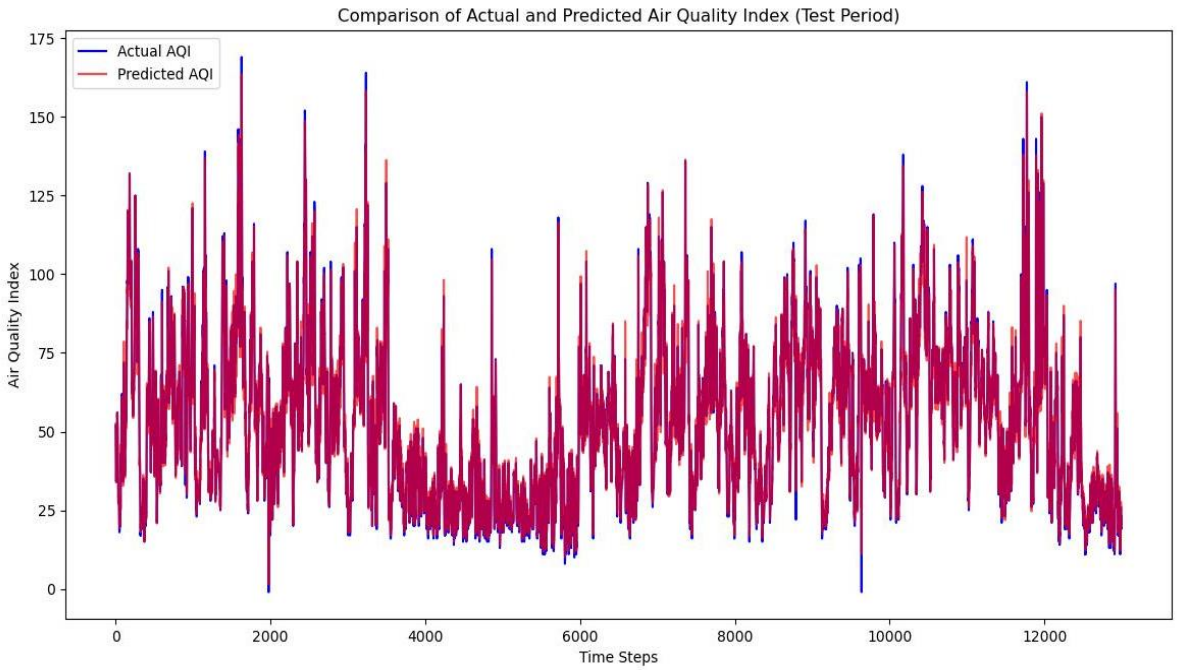
Over 10x higher error

11.62%

Improvement

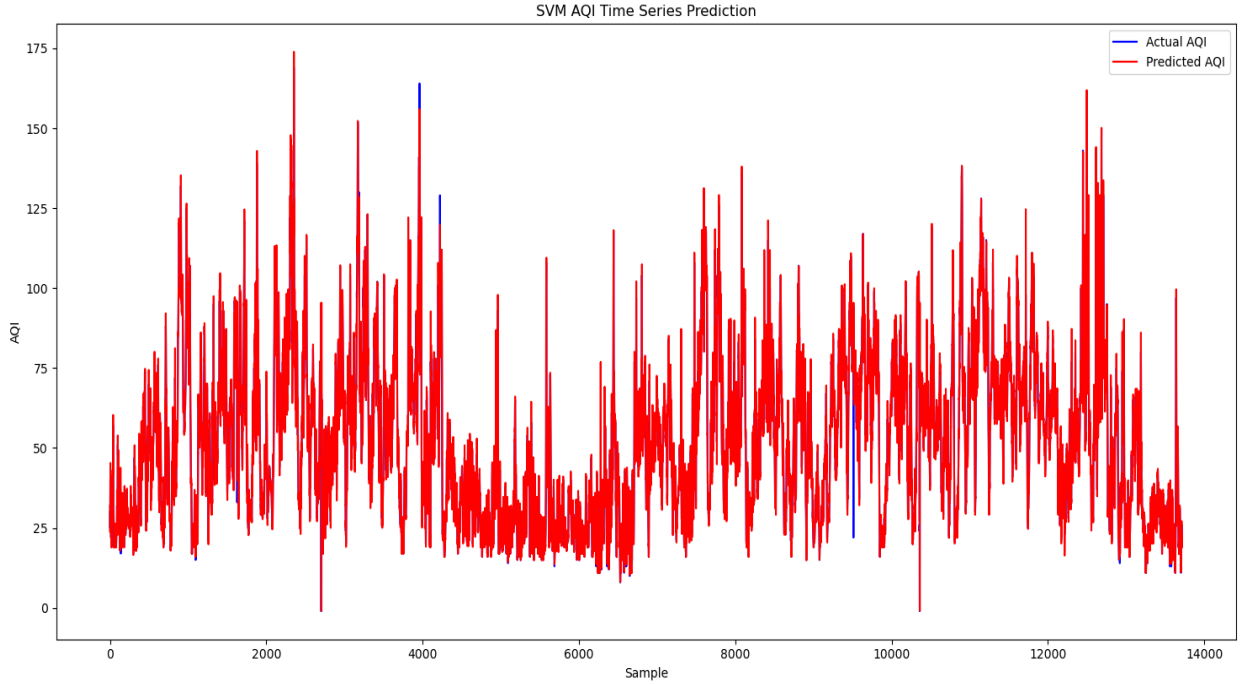
LSTM over SVM

# Visualization of Model Predictions




LSTM Predictions

R-square of 0.98




SVM Predictions

R-square of 0.97

# Why LSTM Excels for Time Series

### Memory Architecture

Dedicated cells store relevant historical information

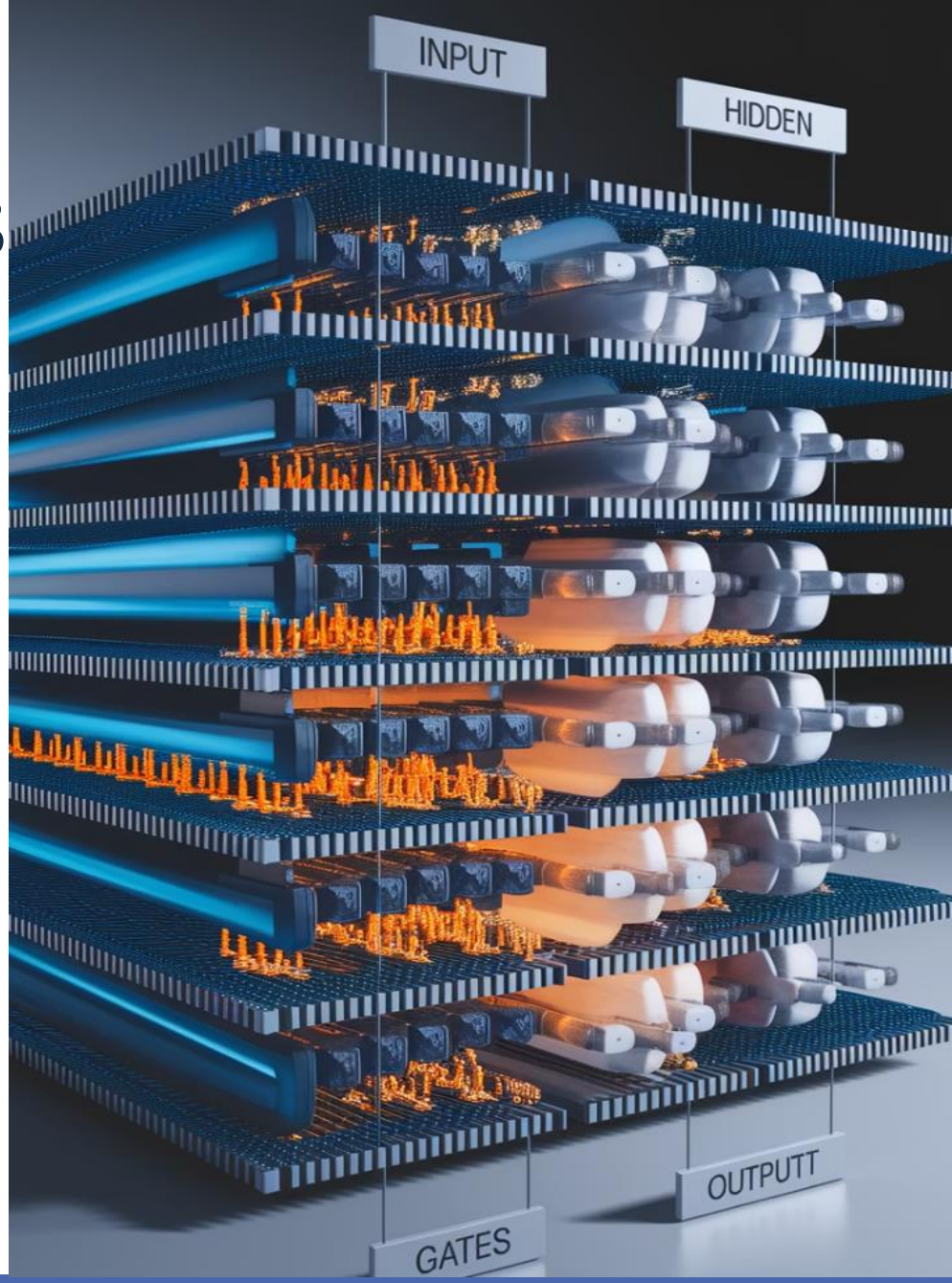
### Selective Learning

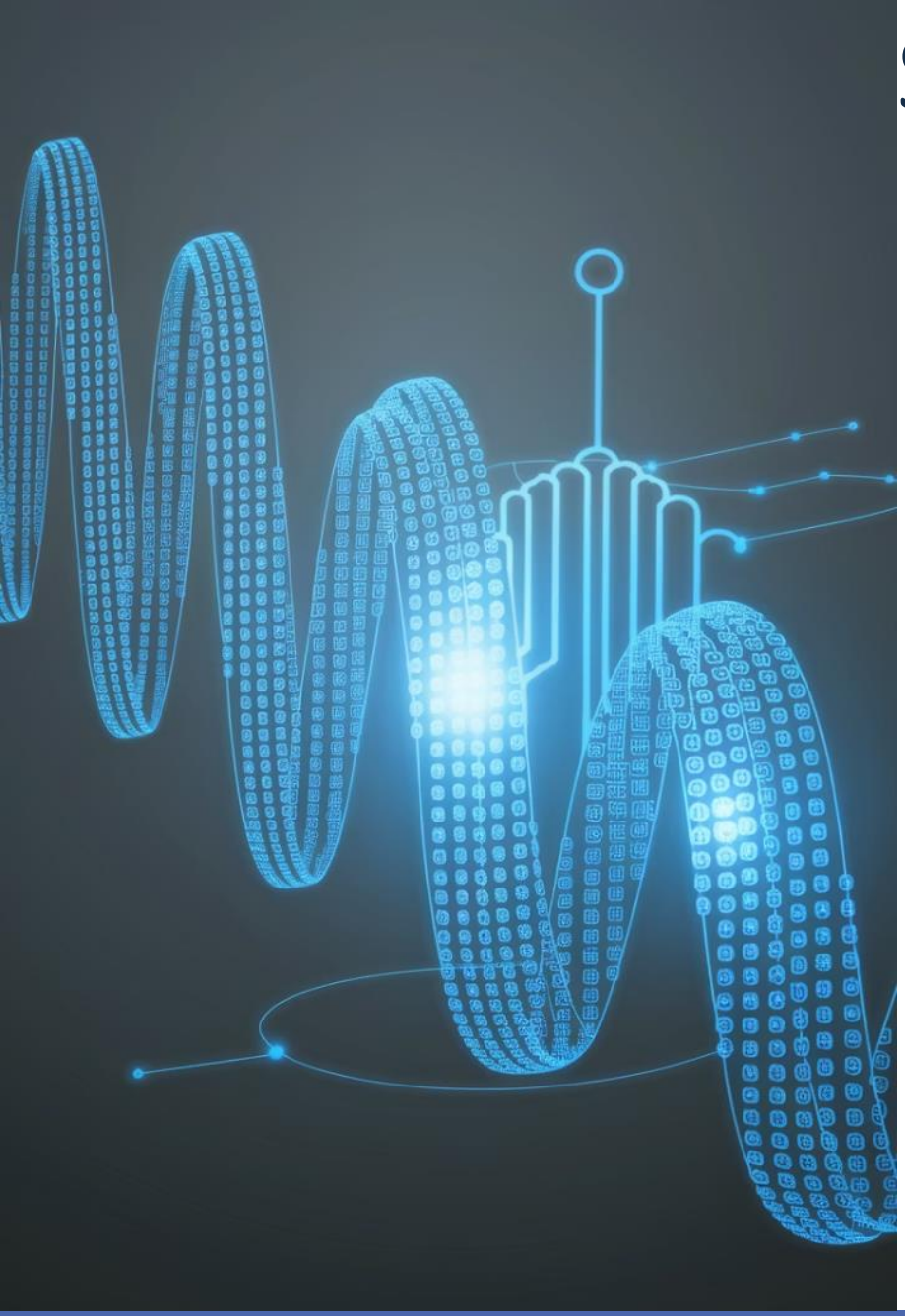
### Pattern Recognition

Identifies complex temporal relationships in AQI data

### Adaptive Learning

Updates weights based on prediction accuracy





# SVM Limitations for AQI Prediction

### Temporal Independence

Treats each data point separately

### Dimensionality Issues

Struggles with high-dimensional time series

### Pattern Limitations

Cannot capture sequential dependencies

### Kernel Selection

Results highly dependent on chosen kernel

# Applications in Environmental Monitoring

### Early Warning Systems

Predict dangerous pollution events

### Urban Planning Planning

Optimize city development for air quality

### Public Health

Reduce exposure to harmful conditions

### Policy Making

Data-driven environmental regulations

### Fraud Detection

Fraud in Accounting

Anomality detection

### Resource Mgt

Crude oil prediction

Water quality prediction

# Future Research Directions

## Hybrid Models

Combining LSTM with other techniques

## Multi-Sensor Integration

Incorporating diverse data sources

## Transfer Learning

Applying models across different regions

## Real-Time Prediction

Moving from hourly to minute-by-minute forecasts


# Key Takeaways

## LSTM Superiority

11.62% lower RMSE than SVM for AQI prediction

## Temporal Patterns

Critical for accurate air quality forecasting

## Practical Applications

Early warning systems, urban planning, public health

## Future Direction

AI will transform environmental science

# Let’s continue the conversation!

**Message me your questions or comments in the IAIA25 app.**

#iaia25