Deep Learning for Drought Prediction: Leveraging Neural Networks and Satellite Data for Climate Resilience

Abstract

In this study, we investigate deep learning models for forecasting drought in Australia by predicting the Standardized Precipitation-Evapotranspiration Index (SPEI) from ERA5 climate data. We compare ConvLSTM, CNN, Variational Autoencoder, and the Transformer-based Earthformer architecture. While most models struggled to generalize and underestimated extreme droughts, Earthformer showed promise in capturing spatiotemporal patterns. However, performance dropped in a two-stage setup. Our results highlight both the potential and current limitations of deep learning for drought prediction, emphasizing the need for better temporal generalization and model integration.

1. Introduction

Droughts are among the most devastating climate-related disasters, affecting ecosystems, agriculture, water supply, and the economy. Their unpredictability and widespread impacts make them a major global concern, particularly in regions where rainfall is limited and erratic (Zhao et al., 2023).

Although drought can occur anywhere, Australia is especially vulnerable. The continent is accustomed to prolonged dry spells that have led to significant agricultural losses and water scarcity in recent years. According to the Australian Bureau of Meteorology, much of Australia receives little rainfall that is not only low in volume, but also highly variable. During the past decade, the country has experienced persistent and widespread drought conditions, making them a defining feature of the recent climate landscape of Australia (NASA Earth Observatory, n.d.).

As climate variability increases, the ability to accurately predict droughts with sufficient lead time has become increasingly important for enabling early warning systems, informed policymaking, and long-term climate resilience (Li et al., 2024). However, drought forecasting remains a challenging task. Traditional statistical and physics-based models often struggle to capture the nonlinear, high-dimensional, and spatiotemporally correlated nature of climate systems. These approaches are typically limited in their ability to integrate diverse datasets—such as temperature, precipitation, evaporation, and humidity—in a way that reflects the complex interactions driving drought events. As a result, many existing methods fall short in terms of both spatial resolution and predictive accuracy (Zhao et al., 2023).

To address this challenge, we propose a deep learning-based framework for data-driven drought forecasting using reanalysis and satellite-derived climate data. Our approach aims to leverage the strengths of modern spatiotemporal neural networks, which have demonstrated strong performance in meteorological prediction tasks. The goal is to build a model that can accurately forecast the Standardized Precipitation-Evapotranspiration Index (SPEI), a widely used drought index (Vicente-Serrano et al., n.d.), by learning patterns in historical climate data and predicting how drought conditions evolve over time.

Specifically, we will evaluate four deep learning architectures: ConvLSTM (baseline), a CNN-based model, a Temporal Variational Autoencoder (VAE), and an attention-based Transformer. Each has demonstrated success in recent environmental forecasting literature, and their comparative evaluation will allow us to identify the most effective approach for drought prediction in Australia.

Our methodology will predicts the future SPEI drought index from ERA5 climate variables. Instead of separating the forecasting and drought prediction steps, the model takes in spatial-temporal blocks of daily-aggregated climate data—such as temperature, precipitation, and evaporation—and learns the spatial and temporal relationships within a single end-to-end framework.

We will use high-resolution datasets that are well-suited for this task:

- **ERA5**: Hourly reanalysis dataset covering atmospheric and surface climate variables (Copernicus Climate Change Service, n.d.)
- **SPEI Database**: Provides historical drought indices with a 0.5 degrees spatial resolution and a monthly time resolution (Vicente-Serrano et al., n.d.)

The geographic region selected for drought prediction spans from latitude –20.5° to –31.0° and longitude 120.0° to

148.0°, covering a significant portion of central Australia. This area was carefully chosen to avoid coastal boundaries and large bodies of water, ensuring that model training is focused on land-based climate dynamics relevant to agricultural and ecological drought. The region is representative of arid and semi-arid zones, making it a suitable testbed for SPEI-based drought forecasting. Figure 1 highlights the exact area used for model input and evaluation.

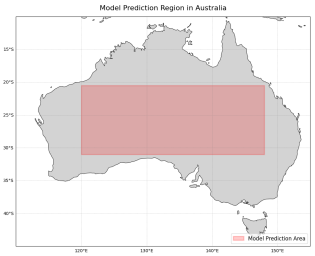

Figure 1. Geographic Region for Model Prediction

Model performance is evaluated using a suite of metrics that capture different aspects of predictive accuracy, including error magnitude, variance explained, and relative forecasting skill. These evaluation criteria are discussed in detail in Section 3.

Our work is guided by recent literature highlighting the potential of deep learning for drought prediction, which is further detailed in Section 2.

2. Related Work

Traditional statistical models often struggle with the complexity of spatiotemporal environmental data, especially when multiple interacting climate variables are involved (Zhao et al., 2023). As a result, deep learning methods have gained significant traction in recent years due to their ability to model nonlinear relationships and high-dimensional temporal dependencies.

Convolutional LSTM (ConvLSTM) networks, proposed by Shi et al. (Shi et al., 2015), have been widely used in precipitation forecasting. ConvLSTMs combine convolutional layers (to extract spatial features) with LSTM units (to capture temporal dependencies), making them suitable for environmental data. This architecture has become a standard baseline in many Earth science prediction tasks.

CNN-based models have also shown strong potential in drought prediction. For example, Li et al. (Li et al., 2024) applied a CNN to forecast drought severity in the Fenhe River Basin using meteorological inputs and the SPEI index. Their work highlights the value of spatial feature extraction in short-term drought forecasting.

Transformer-based architectures have recently emerged as powerful tools for Earth system modeling. EarthFormer (Gao et al., 2022) introduced a cuboid attention mechanism to efficiently capture spatiotemporal dependencies in climate data, achieving state-of-the-art results in various geoscientific forecasting tasks.

Variational Autoencoders (VAEs) have also been applied to environmental time series forecasting. Liu et al. (Liu et al., 2023) proposed a hybrid VAE that learns both local and global temporal structures, showing improved ability to capture seasonal trends and long-term climate variability.

A comprehensive review by Zhao et al. (Zhao et al., 2023) emphasizes the advantages of deep learning approaches—including CNNs, RNNs, and hybrid models—over traditional methods in handling spatial heterogeneity and complex variable interactions in drought modeling.

Our work builds directly on this literature by implementing and comparing multiple deep learning architectures—ConvLSTM, CNN, VAE, and Transformer—for predicting drought severity in Australia. Unlike most prior studies that focus on a single architecture, we adopt a comparative approach using high-resolution, publicly available datasets (ERA5 (Copernicus Climate Change Service, n.d.) and SPEI (Vicente-Serrano et al., n.d.)), positioning our work at the intersection of deep learning innovation and applied climate risk forecasting.

3. Methodology

The target variable for this study was derived from the Standardized Precipitation Evapotranspiration Index (SPEI), a drought index that accounts for both precipitation and potential evapotranspiration to quantify drought severity over different time scales. The SPEI data was obtained from the *SPEI Global Drought Monitor*, developed by the Spanish National Research Council (CSIC) and hosted by the University of Valencia.

The input data for this study was derived from the ERA5 reanalysis dataset, a product of the Copernicus Climate Change Service (C3S) providing a comprehensive record of global atmospheric, land, and oceanic variables. We utilized the Climate Data Store (CDS) API to download hourly data for the specified variables, time period and region. To ensure a consistent temporal representation across all months, even those with varying lengths, a padding procedure was applied. This involved extending shorter months to a uniform length by padding the data with a sentinel value. The downloaded hourly data was then organized into a multidimensional array, incorporating month, day, hour, and spatial dimensions. A fixed value (-9999) was used to flag missing data introduced by the padding.

To reduce the temporal resolution of the data, an aggregation process was employed. This involved applying a set of statistical functions (e.g., mean, maximum, minimum, sum) across the hourly dimension to derive daily summary values. This aggregation step reduces the computational demands of subsequent processing and focuses the analysis on daily-scale variations.

Spatial blocking was implemented to divide the full spatial domain into smaller, overlapping blocks. Both the ERA5 input data and the target SPEI data were partitioned into blocks of uniform spatial dimensions. To handle edge cases where spatial blocks might be smaller than the defined dimensions, a padding procedure was applied. This involved adding a fixed value (-9999) to the boundaries of these smaller blocks to achieve a uniform size across all blocks.

The preparation of the dataset involved coordinating the ERA5 data with the SPEI data. The ERA5 data was downloaded, and the SPEI data, derived from a separate source, was spatially re-gridded to align with the ERA5 spatial grid. The data was then organized into input-output pairs, where the input features consisted of the temporally aggregated ERA5 variables for a given month, and the corresponding target consisted of the SPEI data for the following month.

For the modelling, we follow two parallel approaches to forecast our measure of drought, the SPEI index. First, a two-step approach involves modelling the determinants of the SPEI using a transformer architecture that is able to forecast multidimensional climate variables. A simpler MLP architecture is used to model the target as a function of this forecasted data, replicating the mathematical formula that underlies the calculation of the SPEI. Second, a single step approach involves directly modelling the index as a function of these variables, where we compare different architecture's abilities to directly forecast the target.

A range of evaluation metrics is employed to assess distinct dimensions of predictive performance. Mean Absolute Error (MAE) provides a straightforward measure of the average absolute difference between predicted and actual values, offering an intuitive sense of how far off predictions are on average. Root Mean Squared Error (RMSE) similarly assesses prediction error but penalizes larger deviations more heavily, making it especially useful when large forecasting errors—such as underestimating extreme drought—carry greater consequences. To assess overall explanatory power, the coefficient of determination (R^2) is used, quantifying the proportion of variance in the target variable that is captured by the model, with values closer to 1 indicating better performance. Lastly, the Nash-Sutcliffe Efficiency (NSE), a metric widely used in hydrological modeling, compares the model's predictive skill to that of a naive baseline using the historical mean; values below zero indicate that the model performs worse than this simple benchmark.

For training, we use a custom masked mean squared error (MSE) loss function to handle missing or invalid data values, which are represented by `-9999.0`. This ensures that the model only learns from valid observations by excluding masked regions from the loss calculation, improving stability and robustness when training on climate datasets with spatial or temporal gaps. The masked_mse function serves as the evaluation metric, computing the mean absolute error (MSE) while also excluding -9999 values from the target variable.

4. Model Development

4.1. Convolutional Long Short-Term Memory (ConvLSTM)

The Convolutional Long Short-Term Memory (ConvLSTM) model forecasts the Standardized Precipitation Evapotranspiration Index (SPEI) across spatial grids in Australia. The model ingests a month of daily ERA5 reanalysis variables structured as a 5D tensor of shape `(31, 20, 20, C)`, where 31 is the number of daily time steps, 20×20 is the spatial resolution, and C represents the number of selected input channels (e.g., temperature, evaporation, cloud cover). Channels were filtered by retaining only those with positive Pearson correlation to SPEI to reduce noise and improve generalization.

The architecture is inspired by Shi et al. (2015)'s Convolutional LSTM Network for spatiotemporal sequence forecasting (Shi et al., 2015), and modern drought forecasting models such as those proposed by Zhao et al (Zhao et al., 2023). It leverages the ConvLSTM2D layer, which combines convolutional operations with LSTM dynamics to capture both spatial and temporal dependencies simultaneously (Keras, 2025). The network jointly captures spatial and temporal dependencies using stacked convolutional recurrent layers:

- Three stacked ConvLSTM2D layers with tanh activations and `return_sequences=True`, enabling the model to learn spatial and temporal dependencies jointly.
- BatchNormalization and Dropout (20%) follow each ConvLSTM block to stabilize learning and reduce overfitting.
- A Conv3D bottleneck layer reduces channel depth, followed by LeakyReLU activation.
- A final Conv3D layer with linear activation outputs continuous SPEI forecasts across the spatiotemporal volume..

The model is compiled using the Adam optimizer (learning rate: 0.0003) and trained with a masked Mean Squared Error (MSE) loss.

To enhance convergence and prevent overfitting:

- Early stopping was applied, monitoring validation loss with a patience of 20 epochs.
- Learning rate reduction on plateau was used, halving the learning rate if validation loss did not improve for 5 consecutive epochs.

Regression Model Performance and Limitations

Extensive experimentation was conducted to improve the performance of the regression-based ConvLSTM model. Multiple architectural configurations were explored, including varying the number of ConvLSTM2D layers, adjusting filter sizes, modifying kernel dimensions, and introducing residual connections and bottlenecks via Conv3D layers. Regularization techniques such as dropout, spatial dropout, and L2 weight penalties were tested to combat overfitting. The model was trained with different learning rates and optimizers, with early stopping and learning rate scheduling used to stabilize convergence. Additionally, various data preprocessing strategies were evaluated, including normalization methods, clipping of extreme values, and feature selection based on correlation with the target variable. Despite these efforts, the model consistently regressed toward the climatological mean and struggled to capture the magnitude of extreme droughts—especially during test set evaluation—highlighting limitations in its ability to generalize across time.

Despite careful design and preprocessing, the regression model exhibited limitations in predictive accuracy—particularly on unseen test data. The model achieved:

- Validation Pearson correlation of 0.443
- Validation RMSE of 0.854
- Validation R^2 and NSE of 0.190

These results suggest the model captured modest temporal-spatial patterns. However, the test set correlation fell to 0.108, and R^2 dropped to -2.22, revealing that the model generalized poorly beyond the training year. Predicted SPEI values regressed heavily toward the mean, failing to capture extreme drought events—a critical shortcoming for operational early warning systems.

Binary Classification Approach

In addition to the regression model, a binary classification approach was introduced to better capture extreme drought events. This model reframed the problem as predicting whether a spatial grid cell was in drought (\texttt{SPEI ¡ -1}), aiming to improve the model's sensitivity to severe conditions that the regression framework struggled to identify. The architecture retained its ConvLSTM core, but the output layer was modified to use sigmoid activation, and the model was trained with binary cross-entropy loss. Several refinements were explored, including adjustments to depth, dropout, learning rate schedules, and threshold tuning.

Classification Model Performance and Limitations

Despite the class imbalance (drought is rare), the binary model achieved high precision on both validation and test sets:

- Validation AUC: 0.7760
- Validation Recall: 0.0014
- Test AUC: 0.5788
- Test Precision: 72.03%
- Test Recall: 7.12%

The binary model was conservative: it predicted droughts infrequently, but was correct when it did. However, the low recall indicates that many droughts went undetected. While such conservatism avoids false alarms, it risks missing critical events—a key drawback for real-time applications.

While the ConvLSTM regression model learned temporal-spatial dependencies, it struggled with generalization and underperformed in predicting extreme events. The binary classification model, although more focused and precise, also failed to deliver strong recall.

The final models represent the best balance achieved between interpretability and performance within this experimental framework. Future work could include incorporating class imbalance techniques (e.g., focal loss, SMOTE), using ensemble models, or exploring ordinal classification to capture varying drought severity levels. Nonetheless, this demonstrates the potential and challenges of applying deep learning to drought forecasting using raw climate reanalysis inputs.

4.2. Convolutional Neural Network (CNN)

CNNs have found widespread use in drought and climate prediction problems, with their ability to process satellite imagery making them a valuable tool in many applications. Standard approaches typically employ 2D CNN architectures that use a single satellite image as input to predict corresponding drought indices, as demonstrated in previous studies (e.g., (Chaudhari et al., 2021)). In contrast, our approach leverages the temporal evolution of satellite-based features over the preceding month to predict the drought index for the subsequent month. To capture this temporal context, we extend the architecture to a 3D CNN by introducing an explicit temporal dimension.

The input features X used for training have the shape $(31, 20, 20, 12)$, where 31 corresponds to the number of days in a month, 20×20 represents the spatial resolution of each grid block, and 12 denotes the number of feature channels. Since SPEI is computed on a monthly basis, the target variable y does not vary across days. Consequently, the temporal dimension of the prediction is collapsed before the final output layer, reducing its shape from $(31, 20, 20, 1)$

to $(20, 20, 1)$. This dimensionality reduction ensures that the model outputs a single drought index value per spatial location for each monthly sample.

The core of our model comprises four 3D convolutional (Conv3D) blocks. Each Conv3D layer employs a $(3, 3, 3)$ kernel to extract features from the input data across the temporal, height, and width dimensions. This enables the model to learn spatio-temporal patterns, capturing how values evolve over time within local spatial regions. The number of filters in these blocks progressively increases $(32, 64, 64, 128)$, allowing the model to capture increasingly complex features. Early layers may detect simple patterns, while later layers combine these to recognize more intricate structures. To maintain spatial dimensions throughout the convolutional layers, we use padding. Each Conv3D layer is followed by a ReLU activation function, introducing non-linearity and enabling the model to learn non-linear relationships. L2 regularization is applied to the Conv3D layers to mitigate overfitting by penalizing large weights and promoting simpler patterns. Batch normalization is also applied after each Conv3D layer to stabilize training, normalizing layer activations and reducing sensitivity to input data scaling. Furthermore, Spatial 3D Dropout is employed to further prevent overfitting. Unlike traditional dropout, this technique drops entire feature maps, which is more effective in convolutional layers by encouraging the learning of more robust features.

The model is trained using the Adam optimizer with gradient clipping to stabilize training by limiting the magnitude of gradients. To prevent overfitting, we utilize an Early Stopping callback, which monitors the validation loss and terminates training when it ceases to improve for a specified number of epochs (patience=10), restoring the model's weights to the epoch with the best validation performance. The model is trained with a batch size of 32.

Performance and Limitations

The model, trained over 42 epochs, demonstrated a poor initial fit with a high training loss of 4.6932 and validation loss of 4.1585 in the first epoch. Early training (epochs 2-5) showed significant fluctuations in validation performance, with valuation loss peaking at 6.3571 in epoch 2, likely due to unstable dynamics or suboptimal initial weights.

From epoch 6 onward, both training and validation loss decreased, indicating improved generalisation. Notably, valuation loss dropped to 2.1051 (valuation masked MSE of 0.7491) by epoch 9, and further to 1.2729 by epoch 15. The best validation performance was observed around epochs 18-19 and 38-39, with valuation loss reaching 1.0575 and 0.7635, and corresponding valuation masked MSE values as low as 0.6084, suggesting the model captured key data patterns.

However, some overfitting occurred, as validation loss occasionally increased (e.g., to 1.5197 at epoch 35), despite a general decline in training loss. Overall, the model reduced prediction error and captured relevant patterns, but the relatively high masked MAE and validation MSE indicate room for improvement. While the model converged and outperformed other CNN approaches, the task remains challenging, potentially due to pattern complexity or CNN limitations.

On the test set, the MAE was 1.36, with a RMSE of 1.59, indicating substantial prediction errors and occasional large deviations. The R^2 and NSE were both -2.50, confirming the model's failure to explain data variance and its inferior performance compared to a simple mean prediction.

These results suggest that the CNN architecture may be ill-suited for capturing the long-range temporal dependencies crucial for SPEI-based drought prediction. The limited training data (2000-2001) may not represent the climate variability of the 2002 test year, and the model may suffer from underfitting.

The development process involved extensive experimentation with various CNN architectures and training configurations, including adjustments to network depth, kernel sizes, pooling strategies, dropout rates, optimization algorithms, learning rates, and regularization techniques. Despite these efforts, performance differences among the tested CNN architectures were marginal. This suggests that the observed limitations were not primarily architectural, but rather stemmed from the complexity of the underlying data, the distributional shift between training and test periods, or the CNN's inherent inability to capture long-range temporal dependencies.

A classification-based modeling approach was also explored as an alternative to regression, aiming to detect drought conditions as a binary outcome. However, this approach did not yield significant improvements in predictive performance and was therefore excluded from the final analysis.

4.3. Temporal Variational Auto Encoders

Drought prediction represents a complex spatiotemporal challenge requiring models that can capture both spatial heterogeneity and temporal evolution of meteorological conditions. Variational Autoencoders (VAEs) offer a particularly suitable framework for drought modeling due to the ability to learn probabilistic latent representations while quantifying prediction uncertainty—a critical consideration in climate science applications.

VAEs provide distinct advantages for drought modeling compared to deterministic approaches. The probabilistic framework enables modeling drought states as distributions rather than point estimates, inherently representing uncer-

tainty in drought evolution. The learned latent space disentangles complex drought factors, potentially separating short-term meteorological drivers from long-term hydrological memory effects. Additionally, VAEs enable generation of plausible drought scenarios, facilitating risk analysis under different climate conditions. This probabilistic foundation makes VAEs uniquely positioned to capture the inherent stochasticity of drought processes.

The implemented architecture employs several innovative design choices addressing drought prediction challenges. TimeDistributed convolution preserves the 5D tensor structure (batch, time, height, width, channels), maintaining spatial coherence through encoding. Residual connections preserve extreme drought signals that can be lost in deep networks, critical for maintaining the full dynamic range of drought conditions. Progressive spatial compression (20×20 → 5×5) captures multi-scale interactions between local moisture deficits and regional atmospheric patterns.

A key innovation is the dual-LSTM configuration where the first layer preserves intra-seasonal dynamics while the second extracts inter-annual drought progression patterns. Recurrent dropout (0.3) prevents overfitting to local temporal artifacts. The decoder mirrors this structure through transposed convolutions and upsampling operations.

Training methodology incorporates specialized techniques for drought data: reduced KL weight (0.05) prioritizes reconstruction fidelity over latent space regularity, preserving sharp transitions characteristic of drought onset. Context-specific regularization applies spatial dropout (0.4) to combat ERA5 data redundancy and LSTM dropout to break strong auto-correlations in sequences. The decoder's final layer employs tanh activation scaled by 3, enforcing climatological SPEI bounds of [-3,3].

Performance and Limitations

Feature importance analysis through systematic perturbation revealed the dew-point temperature and total column water vapor as the strongest positive predictors, aligning with the meteorological understanding. The model successfully captured the known relationships between skin temperature, precipitation, and severity of drought.

The enhanced model achieved modest improvements, reducing test MAE by 8.1% (1.40 → 1.29) and MSE by 16.9% (2.60 → 2.16). However, persistent negative R^2 values reveal fundamental challenges. The substantial gap between training (MAE ≈ 0.78) and test performance (MAE ≈ 1.29) suggests non-stationarity in drought processes that standard train-test splits fail to capture. Both models underestimated extreme drought events (SPEI ¡ -1.4), and PCA visualizations of the latent space revealed only partial organization by severity of drought.

A complementary binary classification approach (SPEI ¡ -1) exhibited a substantial shift in class distribution between training (32.7% drought) and test (87.9% drought) periods, quantifying the temporal distribution shift problem. High precision (93%) but low recall (39%) in drought detection highlights a critical limitation—the model avoids false alarms but misses many drought events.

This extensive experimentation highlights that standard machine learning assumptions of stable distributions between training and testing may be fundamentally problematic for drought prediction. The VAE framework offers valuable probabilistic representations and uncertainty quantification, but may require more sophisticated priors to fully capture drought dynamics. Complementary approaches (continuous severity estimation and binary classification) provide different operational insights, suggesting ensemble methods may be more robust for drought early warning systems.

4.4. Earthformer (Transformer)

The Earthformer model is a specialised variant of the SWIN transformer, proposed by Gao et al (Gao et al., 2022), designed to capture spatial and temporal trends in multidimensional data. The architecture notably introduces a cuboid decomposition technique to model high dimensional data, decomposing the climate data into smaller spatio-temporal blocks. Once decomposed, a mixture of local and global attention layers allow the transformer to model both local relationships within each block and any global trends that may occur between these blocks. While the authors suggest a variety of modelling strategies, we implement an axial cuboid decomposition strategy, which involves modelling the relationships along one dimension of each layer (spatial and temporal). An encoder stacks multiple cuboid decomposition layers with interleaved convolution layers to compress the data. A decoder is layered with both masked self attention and cross attention cuboid layers following the canonical transformer architecture, where the compressed features extracted by the encoder are fed into the cross attention layers of the decoder. Finally, the output is upsampled and combined with the global trend features extracted by the encoder and decoder using dense layers and reshaped to fit the target variable.

- Encoder with two sets of three stacked axial blocks (one for time, width, and height each). Two convolution layers downsample the data by a combined factor of four.
- Decoder with one masked self attention stacked axial block (3 cuboid self attention layers) and two cross attention axial blocks.
- A Dense layer to combine learnt local and global features.
- Final Conv3D and upsampling layer to forecast the

climate inputs into the next month.

Performance and Limitations :

The model is compiled with an Adam optimisation algorithm, utilising a learning rate scheduler to fine-tune convergence. Careful experimentation with a variety of dropout blocks placed after local and global attention layers reveal a plateau in the validation MAE between 0.42 and 0.43 standard deviation units. The models test error metrics are:

- MAE of 0.43
- RMSE of 0.63
- R^2 and NSE of 0.56
- Trend accuracy of 73%

The Earthformer based model converges immediately from an initial MAE of 0.7 to an average loss of 0.4 on subsequent epochs. Both the validation and test error metrics indicate that the Earthformer architecture is able to generate a reasonably accurate forecast of the climate data. However, the MAE threshold of 0.4 and a maximum R^2 of 0.56 potentially highlight limitations in the data, where further increases in model performance are likely to depend on the inclusion of additional climate variables to increase the models explanatory power. The forecasted climate variables are tested on a simple MLP model targetting the SPEI index. The model reaches a training MAE and MSE of 0.14 and 0.04 respectively, easily replicating the mathematical function that determines the index. Conversely, the test metrics diverge from these observations:

- MAE of 1.0544
- RMSE of 1.2799
- R^2 and NSE of -0.90
- Trend accuracy of 13%

While both models perform reasonably well on their individual tasks, they reveal a key limitation of the two step modelling approach. That of error compounding, where precision limitations in the first model are likely to feed into the test error of the prediction task in the second step. The index prediction results and an R^2 of 0.56 suggest that the explanatory power of the forecasting model may be inadequate for downstream prediction tasks. Additional variables must be included with extensive feature engineering to improve the models performance in downstream tasks that require a high level of precision. Finally, the predicted SPEI distribution highlights a glaring limitation inherent in climate data analysis. While the SPEI index varies by 2 units around a mean of 0, the predicted index was highly negatively skewed suggesting a temporal drift in the test predictions that do not accurately capture out of sample observations. This phenomenon is likely the result of an annual climate trend that is not caught by the current forecasting transformer model, where annual increases in temperature are likely shifting the variance of the predictive distribution. Consequently, we suggest expanding the training dataset to multiple years and changing the model architecture to account for longer temporal trends as a possible further area of study when applying a two step process to forecast the SPEI index.

5. Comparative Analysis

5.1. Quantitative Evaluation

We implemented and evaluated four distinct deep learning architectures to forecast drought severity in Australia based on SPEI values. This section summarizes the comparative performance of these models and interprets their respective strengths and limitations.

Table 1. Test Set Performance Summary

Model	MAE	RMSE	R^2	NSE
CNN	1.36	1.59	-2.50	-2.50
ConvLSTM	1.42	1.66	-2.22	-2.22
VAE	1.29	1.47	-1.52	-1.52
Earthformer	0.42	0.62	0.56	0.56
2-Step MLP	1.05	1.28	-0.90	-0.90

As shown in Table 1, the Transformer-based Earthformer model demonstrated the highest predictive performance for climate variable forecasting, achieving an MAE of 0.42 and an R^2 of 0.56. However, the two-step prediction pipeline—forecasting climate variables with Earthformer and then estimating SPEI using an MLP—resulted in substantial error propagation, with the final SPEI prediction exhibiting an MAE of 1.05 and clear bias.

The ConvLSTM regression model captured modest temporal-spatial patterns but regressed heavily toward the climatological mean during testing. The binary classification variant slightly improved on detecting drought conditions, achieving a test precision of 72%, but suffered from very low recall (7.12%).

The CNN model, despite architectural regularization, significantly underperformed on the test set. With negative R^2 and NSE values, it failed to generalize to unseen drought conditions, particularly due to limited temporal context and data sparsity.

The VAE model reduced test MAE by 8.1% compared to CNN and offered probabilistic outputs, making it well-suited for uncertainty-aware applications. However, it also exhibited weak generalization, with negative R^2 and poor recall in extreme drought classification.

5.2. Interpretation and Discussion

The Transformer-based Earthformer architecture most effectively learned the underlying structure of multidimensional climate data and yielded the best forecasting performance. Its ability to model both local and global temporal-spatial relationships via axial cuboid attention contributed to superior trend prediction. However, its use in a two-step framework led to compounding errors, undermining its practical value for drought classification.

CNNs proved inadequate for the task, struggling with long-term temporal dependencies and limited training data. ConvLSTM networks, while conceptually suited for spatiotemporal data, failed to generalize under climate variability and were overly conservative in regression settings. The binary classification variant was more precise but lacked sensitivity.

The VAE architecture offers the promise of uncertainty quantification and multi-scale feature learning. However, distributional shift and class imbalance posed critical challenges. Both the CNN and VAE models showed strong performance on training data but failed on temporally disjoint test periods—suggesting non-stationarity as a core challenge in climate modeling.

5.3. Conclusion of Comparative Analysis

In summary, Earthformer is the most promising architecture for spatiotemporal climate forecasting, particularly in modeling climate drivers of drought. However, integrating prediction into operational drought monitoring systems requires addressing compounding error in multi-stage models. While ConvLSTM and VAE provide valuable alternatives, their performance suffers under non-stationary and limited datasets.

6. Conclusion

This study investigated the application of four deep learning architectures for forecasting drought conditions in Australia using SPEI. Despite the growing success of deep learning models in other domains, accurately predicting droughts remains a particularly difficult challenge. All models encountered substantial performance issues, especially when generalizing from the training years (1999–2001) to the test year (2002).

While some models, such as the Earthformer, performed reasonably well on the validation data, the majority of models struggled significantly on the test set. CNN, ConvLSTM, and VAE models all showed high prediction errors and negative values for key evaluation metrics such as R^2 and NSE. These results indicate that the models were unable to capture the underlying variability in the data and, in some cases, performed worse than simple baseline models. The Earthformer model demonstrated the best overall performance in forecasting climate variables, and its integration into a two-stage pipeline to predict SPEI appeared promising. However, the final predictions of the drought index were subject to compounding errors from the initial forecasting stage, resulting in suboptimal performance on the target task.

Several factors likely contributed to the limited success of the models. One key limitation was the restricted size and temporal coverage of the dataset. Training on just three years of data limited the models' ability to capture long-term climate variability and trends, while the 2002 test set may have exhibited different climatological conditions than the training period. This temporal distribution shift likely impaired the models' generalization ability. Another important factor was the relatively short temporal context used in modeling: each model was given one month of climate variables to predict the drought index for the following month. This forecasting setup may have been insufficient to capture the cumulative and delayed effects of climate drivers that lead to drought. Furthermore, the one-month forecasting horizon limits the practical utility of the models for early warning and policy planning. Additionally, all models had difficulty detecting and predicting extreme drought events, tending instead to regress towards the climatological mean.

Despite these challenges, the Earthformer architecture stands out as a promising direction for future research. Its ability to capture complex spatiotemporal relationships in high-dimensional climate data is evident, and refining this two-stage approach could yield more reliable predictions. Future work should focus on expanding the training and testing periods to include more years, thereby exposing the models to a wider range of climate variability. Increasing the temporal window of input data could also help models better capture the precursors of drought events. Improving the integration of the forecasting and prediction stages, for example, through joint training or uncertainty-aware mechanisms, may reduce the impact of compounding errors. Furthermore, incorporating additional variables and broader atmospheric indicators could enhance the models' predictive power and robustness.

In summary, this study highlights both the potential and the limitations of deep learning methods for drought forecasting. While current approaches face notable challenges in generalization, data limitations, and temporal sensitivity, the findings provide valuable insights into how future models might be improved. With further development, deep learning can become a vital tool in building more accurate and timely drought early warning systems.

A. Statement of Individual Contributions

This project was completed collaboratively by a group of four members. Each member was responsible for the development, implementation, and evaluation of one deep learning model for drought prediction. The individual technical contributions are as follows:

- **Candidate 41117** was responsible for designing, implementing, and evaluating the CNN model. This included preprocessing the input data, building the model architecture, training the model, and analysing the performance results.
- **Candidate 44112** developed the ConvLSTM model. Their contributions included designing the model architecture, preparing and preprocessing the dataset, implementing advanced data strategies, conducting hyperparameter tuning and experimentation, and evaluating model performance..
- **Candidate 42371** worked on the Temporal VAE model, involving end-to-end design and implementation of VAE architectures, advanced data handling, extensive hyperparameter tuning and experimentation, architectural enhancements, and performance evaluation using both quantitative metrics and latent space analysis.
- **Candidate 46673** was in charge of the Earthformer based transformer model. This included preparing the temporal input structure, modifying the model to suit spatial-temporal data, and interpreting prediction outcomes.

All members contributed equally to the final report, literature review, and group discussions throughout the project lifecycle.

B. GitHub Repository

The full codebase is available at this GitHub repository: https://github.com/lse-st456/2025-projects-team-stats

References

Chaudhari, S., Sardar, V., Rahul, D., Chandan, M., Shivakale, M. S., and Harini, K. Performance analysis of cnn, alexnet and vggnet models for drought prediction using satellite images. In *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1–6, 2021. doi: 10.1109/ASIANCON51346.2021.9545068.

Copernicus Climate Change Service. Climate reanalysis (era5). https://climate.copernicus.eu/climate-reanalysis, n.d. Accessed: 2025-04-11.

Gao, Z., Xu, Y., Xu, Z., Qi, Q., Zhu, Y., Chen, Z., Xiong, C., Liu, Y., Zhang, W., Yu, B., et al. Earthformer: A spatiotemporal transformer for earth system forecasting, 2022. URL https://arxiv.org/abs/2207.05833.

Keras. Convlstm2d layer. https://keras.io/api/layers/recurrent_layers/conv_lstm2d/, 2025. Accessed: 2025-04-28.

Li, T., Liu, C., Zhang, Z., Li, X., and Wei, Y. Basin-scale daily drought prediction using convolutional neural networks in fenhe river basin, china. *Atmosphere*, 15(2):155, 2024. doi: 10.3390/atmos15020155. URL https://www.mdpi.com/2073-4433/15/2/155.

Liu, J., Wang, Y., Yu, J., Zhang, Y., and Zhang, B. Hybrid variational autoencoder for time series forecasting, 2023. URL https://arxiv.org/pdf/2303.07048.

NASA Earth Observatory. World of change: Australia's vegetation. https://earthobservatory.nasa.gov/world-of-change/AustraliaNDVI, n.d. Accessed: 2025-04-11.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-C. Convolutional lstm network: A machine learning approach for precipitation nowcasting, 2015. URL https://arxiv.org/abs/1506.04214.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I. Spei global drought monitoring: Database. https://spei.csic.es/database.html, n.d. Accessed: 2025-04-11.

Zhao, P., Qiao, Y., Hu, L., and Ma, Z. Drought forecasting using deep learning: A review and future perspectives. *Sustainability*, 15(7):6160, 2023. doi: 10.3390/su15076160. URL https://www.mdpi.com/2071-1050/15/7/6160.