DroughtED: A dataset and methodology for drought forecasting spanning multiple climate zones

Christoph Minixhofer [1] **Mark Swan** [1] **Calum McMeekin** [1] **Pavlos Andreadis** [1]

Abstract

Climate change exacerbates the frequency, duration and extent of extreme weather events such as drought. Previous attempts to forecast drought conditions using machine learning have focused on regional models which have two major limitations for national drought management: (i) they are trained on localised climate data and (ii) their architectures prevent them from being applied to new heterogeneous regions. In this work, we present a new large-scale dataset for training machine learning models to forecast national drought conditions, named DroughtED. The dataset consists of globally available meteorological features widely used for drought prediction, paired with location meta-data which has not previously been utilised for drought forecasting. Here we also establish a baseline on DroughtED and present the first research to apply deep learning models - Long Short-Term Memory (LSTMs) and Transformers - to predict county-level drought conditions across the full extent of the United States. Our results indicate that DroughtED enables deep learning models to learn cross-region patterns in climate data that contribute to drought conditions and models trained on DroughtED compare favourably to state-of-the-art drought prediction models trained on individual regions.

1. Introduction

Droughts are natural events characterized by prolonged shortages of precipitation. Unforeseen droughts cause profound economic and social impacts, resulting in significant losses to agriculture, domestic water supply and natural wildlife (Svoboda et al., 2002). Since 1980, the United States (US) has experienced 26 major unexpected droughts, with an average cost of $9.6 billion incurred during each drought event (Smith, 2020). Under the current trajectory of global warming, the US is anticipated to experience chronic drought conditions with increased frequency and intensity over the coming decades (Elsner et al., 2010; Hayhoe et al., 2018). Effective monitoring and prediction models are therefore critical for informing drought management action and can help mitigate the detrimental economic impacts on individuals, communities and nature.

To the best of our knowledge, deep learning has never been applied to the task of nationwide drought forecasting. Previous attempts to predict future drought conditions using deep learning have focused on small homogeneous areas within a distinct geographic location (Agana & Homaifar, 2017; Dikshit et al., 2021; Mishra et al., 2007). Whilst this showcases the promising opportunity to apply deep learning for drought forecasting, these regional models and datasets remain restricted to an isolated area, thereby limiting the ability for environmental stakeholders to monitor drought conditions at larger scales. Regional models overfit to state-level observations in climate datasets and, by their construction, have no opportunity to learn meteorological patterns that are consistent across multiple heterogeneous regions. Furthermore, regional models have inherently siloed applications, therefore increasing the technical overhead for environmental organisations to transfer these models to new geographic areas.

The input features presented in this paper enable the prediction of national drought conditions using a single deep learning model. Forecasting drought across heterogeneous regions using exclusively meteorological observations is a challenging task. The distribution of climate patterns that contribute to drought are dependent on their geographic climate. For instance, in the US, states with vast regions of desert such as Nevada and semi-arid mountainous states such as Wyoming, both experience low-levels of annual rainfall, however, they do not share similar drought conditions (Easterling et al., 2017). Given this complexity, previous attempts to forecast drought in the US have focused exclusively on regional models, such as predicting drought conditions in Western US States or in the Colorado River Basin (Agana & Homaifar, 2017; Bolen et al., 2021).

[1]School of Informatics, The University of Edinburgh, Edinburgh, Scotland. Correspondence to: Christoph Minixhofer <christoph.minixhofer@gmail.com>.

Climate Change AI Workshop at ICML 2021, Copyright 2021 by the author(s).

To address these key issues with regional models, this paper contributes a new dataset for the task of large-scale drought forecasting, named DroughtED (Drought Earth Dataset). The dataset combines globally consistent spatial and temporal features related to drought that enable machine learning models to be trained for nationwide drought prediction and monitoring. Also, we include a location-identifier feature in the DroughtED dataset which, when used as an input feature to a deep learning model, enables drought predictions across diverse geographic regions. As a secondary contribution, we establish a baseline on DroughtED and present the first evaluation of two deep learning models - Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) and Transformer (Vaswani et al., 2017) models - for the task of drought forecasting across the US. Altogether, we hope that this work will facilitate future deep learning research in tackling the complex challenge of large-scale drought forecasting and will reduce the barrier to entry for new regions to employ machine learning systems for improved drought monitoring.

2. Related Work

Previous attempts to classify future drought conditions commonly forecast precipitation indices. For instance, SPI (Standardized Precipitation Index) (Guttman, 1999) and SPEI (Standardized Precipitation Evapotranspiration Index) (Serrano et al., 2010) values are frequently used (Agana & Homaifar, 2017; Dikshit et al., 2021; Fung et al., 2019; Khan et al., 2020; Poornima & Pushpalatha, 2019; Rhee et al., 2016). SPI/SPEI are derived from prior distributions of precipitation and temperature related to drought (see Appendix C). These indicators measure *meteorological drought*, which is defined as a lack of precipitation compared to a normal amount of precipitation estimated from previous data (Mishra & Singh, 2010). Prior research has argued that SPI and SPEI indicators do not capture the full extent of droughts because droughts in surface and ground waters are influenced by factors other than precipitation deficit (Agnew, 2000; Kubicz, 2018). For this reason, we have chosen not to include SPI/SPEI values in the proposed DroughtED dataset.

The leading data source of national drought monitoring and classification in the US is the US Drought Monitor (USDM). The USDM collates weekly meteorological and hydrological parameters and on-the-ground observations, prepared and interpreted by climate experts, to categorise current drought conditions across all 3,108 continental US counties. The USDM is used by environmental institutions to influence drought management protocols, from triggering disaster declarations to advising agricultural planning procedures (Svoboda et al., 2002). Importantly, the USDM is "not a statistical model" (National Drought Mitigation Center, 2021) and it does not forecast future drought conditions. In contrast to the aforementioned SPI/SPEI, the USDM categories measures a combination of *meteorological, hydrological and agricultural drought* (Mishra & Singh, 2010), making it a more holistic measure (Svoboda et al., 2002). To the best of our knowledge, deep learning has never been applied to predicting future USDM categories. As the USDM is curated by climate experts, these labels are of high quality, and are therefore well-suited to evaluate the usefulness of the input features in DroughtED.

To the best of our knowledge, all prior attempts to forecast droughts using machine learning, even when applied on a national scale, train individual models for each climate region (Adede et al., 2019; Agana & Homaifar, 2017; Jalili et al., 2013). These approaches include nationwide studies training individual models for multiple distinct climate regions within a country (Jalili et al., 2013), or exclusively cover a single homogeneous region (Adede et al., 2019; Agana & Homaifar, 2017). In domains other than drought forecasting, location-specific and location-agnostic methods have been compared: Pylianidis et al. (2021) predict pasture nitrogen response rate using a location-specific and -agnostic model, and conclude using a model spanning several locations aids model generalization.

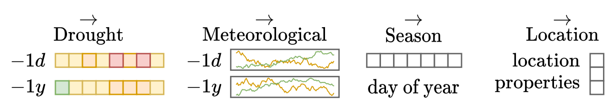


Figure 1. The final dataset consists of previous drought observations and meteorological observations, each with observations leading up to the current time and 1 year prior. To improve generalization across regions, a location identifier includes parameters specific to each location.

3. Dataset

In this work, we introduce a drought dataset which combines 180 daily meteorological observations with geospatial location meta-data for all 3,108 counties across the continental US to allow building a location-agnostic drought forecasting model. All input features are sourced from globally and publicly available datasets, ensuring that our dataset can be extended to new geographical locations such as countries outside of the US. These inputs are paired with USDM drought categories (Svoboda et al., 2002) which are delayed between 1- to 6 week time-frames. We do this by using the following data features outlined in Figure 1. We use USDM categories rather than SPI/SPEI as the USDM captures expert opinion on drought conditions rather than meterological deviations from the norm alone, making it more than a precipitation/temperature forecasting model.

The NASA Prediction Of Worldwide Energy Resources (POWER) project (Sparks, 2018) offers meteorlogical real-time and historical data as a combination of the MERRA-2 (Gelaro et al., 2017) and GEOS (Keller et al., 2021) datasets. The data contains measurements on precipitation, surface pressure, relative humidity dew/frost point, wind speed, and temperature with daily resolution. We include observations preceding the other observations by one year. We do this to make it possible for a model to learn from deviations from the previous year.

Previous drought observations are also included in the dataset. The USDM drought level belongs to one of six categories; no drought (None), abnormally dry (D0), moderate (D1), severe (D2), extreme (D3) and exceptional (D4). We convert these categories to a numeric feature by assigning 0 to no drought and 5 to D4 (exceptional drought). As with the meteorological data, we provide the model with 180 observations leading up to the desired drought prediction and the measurements from the year prior. This variable could be replaced with SPI/SPEI for regions in which manually created drought labels are not available. We additionally include the **season**, which is the day of the year using sine and cosine. As drought is a seasonal phenomenon (Van Loon et al., 2014), this input aids generalization across seasons.

As our dataset encompasses all of the US, which contains different climates zones, we include a *location indicator* which summarises each location as a combinations of parameters derived from the **Harmonized World Soil Database** (Fischer et al., 2008). This data includes the slope, aspect and elevation of terrain at each location, as well as the land use of each location (e.g. rain-fed cultivated land or woodland), and the soil quality in terms of, for example, toxicity or nutrient availability. We hypothesise that this location information, in terms of soil properties, can enable models to generalise across large areas, with training data from one area possibly improving predictions in another area.

The DroughtED dataset contains drought, meteorological, season and location vectors (Figure 1). The drought vector with USDM categories can be replaced with SPI/SPEI and the meteorological vector can be extended to include further temporal data. The location vector can be extended by adding binned latitude/longitude values indicating the approximate geographical area of each location. Given that DroughtED contains data derived from sources available globally, our work can enable drought prediction research to be applied to new countries. Additionally, DroughtED is open source and we have published scripts to support the collection of DroughtED data features which we hope will aid future research[1].

[1] https://kaggle.com/cdminix/us-drought-meteorological-data

4. Models

Both a LSTM and Transformer model were used to create baseline predictions of drought levels given the drought, meteorological, season and location vectors. LSTMs have been successfully used in related time series tasks such as predicting short and mid-term sea surface temperatures (Xiao et al., 2019), precipitation forecasting (Tong & de Witt, 2021) and in previous drought forecasting work (Dikshit et al., 2021). Transformers have been successfully applied to time series forecasting in different domains (Li et al., 2019; Wu et al., 2020). To encode the inputs using the Transformer, we project the input features using a linear layer and also add a special `[REG]` token in this space to obtain an embedding which can be used for regression, similar to the `[CLS]` token used for pre-training BERT (Devlin et al., 2018) (Appendix D). We perform hyper-parameter tuning on both architectures given a compute budget of 30min of training time on a GeForce GTX 1080 Ti graphics card per run, using the AdamW optimizer (Loshchilov & Hutter, 2017) and 1-cycle learning rate schedule (Smith, 2018) and arrive at a set of hyperparameters (see Appendix B).

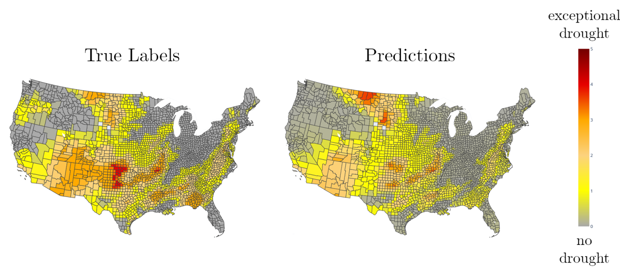


Figure 2. True labels and predictions for December 2017 using the best Transformer model and a 4-week future forecasting period.

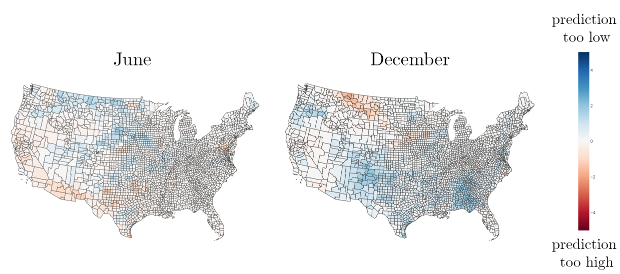


Figure 3. The difference between true labels and predictions for the best Transformer model for June & December 2017.

The drought, meteorological and season vectors are combined and passed as sequential inputs to the LSTM/Transformer. An encoding of said inputs is then passed to a single-layer feed-forward neural network (FFNN), which also takes in the Location vector and outputs 6 weeks of drought predictions. Predictions and errors on validation data are shown in Figure 2 and 3, respectively.

	WEEK 1		*WEEK 2*		*WEEK 3*		*WEEK 4*		*WEEK 5*		*WEEK 6*	
	MAE	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1
LSTM	0.178	**81.1**	0.237	**72.3**	**0.265**	**63.9**	**0.328**	**56.8**	**0.395**	50.5	**0.433**	**47.5**
Transformer	**0.159**	67.0	**0.215**	62.9	0.267	60.3	0.335	56.4	0.398	**50.6**	0.435	46.7
model average ensemble	0.135	83.6	0.198	72.7	0.254	65.1	0.321	58.2	0.388	51.6	0.427	48.0

Table 1. Results on the test set for the Transformer and LSTM with additional features. Combining LSTM and Transformer averaging predictions consistently yields better MAE and F_1 scores.

5. Experiments and Results

Given the aforementioned setup, we seek to explore a) how much the different types of input features contribute to model performance, b) if location-agnostic models can be trained and perform similarly to regional state-of-the-art drought prediction models and c) how location-specific models compare to the location-agnostic one. We measure model performance using macro F_1 scores over drought categories, and average the scores for each weeks' prediction.

For a) we first train and evaluate both LSTM and Transformer on drought and meteorological observations alone, in line with previous work (Dikshit et al., 2021). We then assess the impact of individual features by retraining the model with said features (Appendix A). For all features, we find no significant improvements when using the LSTM model but find relative improvements of 2.5% for seasonal information, 7.6% for location information, 9.5% for binned latitude/longitude values and 16.7% for including drought and meteorological observations from the year prior on the Transformer model. In combination, the features improve the model by 16.6%. We conclude that the features included in the dataset can improve attention-based model performance, with future work needed to explain the differences between LSTM and Transformer model. The single best model performance is achieved when combining LSTM and Transformer with the listed input features in a model average ensemble (see Table 1).

To put our model's results in context we compare to a similar drought forecasting effort in a sub-region of Australia conducted by Dikshit et al. (2021). This work is one of the most recent and largest area forecasting approaches which reports the multi-class weighted ROC-AUC score of SPEI values, which can be compared to USDM predictions, which are also categorical. Our best single model achieves a multi-class weighted ROC-AUC score of 78.1, while theirs achieves 83.0. We see this as comparing favourably, especially since the USDM is a more holistic measure of drought and possibly harder to predict than SPEI. Additionally, we compute the root mean squared error (RMSE) of our model by converting USDM drought categories to numerical values and scaling values to the same range as the SPI values forecast by Tan & Perkowski (2018). For 1-month lead times, the resulting RMSE of 0.19 is in line with RMSE values of 0.29 and 0.13 obtained by Tan & Perkowski (2018) for SPI12 and SPI24, respectively.

For c) we randomly select three US states (which are Iowa, Montana and Oklahoma) and train the LSTM model on i) the training data in the state alone ii) all training data. We find that for each state, the model trained on all training data outperforms the state-specific model, by 4.6% (relative) on average (Appendix E). This shows that given our setup, location-agnostic models outperform location-specific ones.

6. Conclusion

We present DroughtED, a novel dataset for data-driven drought forecasting. We also establish a baseline on DroughtED and showcase the first research to apply a single deep learning model to forecast national drought conditions. We demonstrate that models trained on nationwide data benefit from learning climate patterns that are applicable across heterogeneous regions. With the release of DroughtED, our primary aim is to enable future research on large-scale drought prediction and lower the barrier to entry to apply drought forecasting models to new geographical regions.

Future work could explore the impact of the individual components of the meteorological and location data in our dataset on forecasting performance. Additionally, we have laid the groundwork for extending DroughtED beyond the US by utilising globally available data, it could be investigated which globally available drought indicators could be used in combination with our dataset. With rising environmental challenges resulting from climate change, our hope is that the dataset and deep learning models presented in this report can be integrated into existing drought management systems to help mitigate the adverse impacts of drought. We have initiated discussions with a global environmental organisation to discuss potential applications for improved drought monitoring. Our hope is to support data-driven investigations for drought forecasting and, ultimately, to help contribute to the advance of drought prediction research.

References

Adede, C., Oboko, R., Wagacha, P., and Atzberger, C. A mixed model approach to drought prediction using artificial neural networks: Case of an operational drought monitoring environment. *arXiv preprint arXiv:1901.04927*, 2019.

Agana, N. A. and Homaifar, A. A deep learning based approach for long-term drought prediction. In *Southeast-Con*, 2017.

Agnew, C. T. Using the SPI to identify drought. *Drought Network News*, 2000.

Bolen, B., Sodwatana, M., and Hampson, H. Drought prediction in the western US using a long short-term memory (LSTM) model. 2021.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, 2018.

Dikshit, A., Pradhan, B., and Alamri, A. M. Long lead time drought forecasting using lagged climate variables and a stacked long short-term memory model. *Science of The Total Environment*, 2021.

Easterling, D. R., Arnold, J., Knutson, T., Kunkel, K., LeGrande, A., Leung, L. R., Vose, R., Waliser, D., and Wehner, M. Precipitation change in the united states. 2017.

Elsner, M. M., Cuo, L., Voisin, N., Deems, J. S., Hamlet, A. F., Vano, J. A., Mickelson, K. E., Lee, S.-Y., and Lettenmaier, D. P. Implications of 21st century climate change for the hydrology of washington state. *Climatic Change*, 2010.

Fischer, G., Nachtergaele, F., Prieler, S., Van Velthuizen, H., Verelst, L., and Wiberg, D. Global agro-ecological zones assessment for agriculture (GAEZ). 2008.

Fung, K. F., Huang, Y. F., and Koo, C. H. Coupling fuzzy–SVR and boosting–SVR models with wavelet decomposition for meteorological drought prediction. *Environmental Earth Sciences*, 2019.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 2017.

Guttman, N. B. Accepting the standardized precipitation index: A calculation algorithm. *JAWRA*, 1999.

Hayhoe, K., Wuebbles, D. J., Easterling, D. R., Fahey, D. W., Doherty, S., Kossin, J. P., Sweet, W. V., Vose, R. S., and Wehner, M. F. Our changing climate. impacts, risks, and adaptation in the united states: The fourth national climate assessment, volume II. *U.S. Global Change Research Program*, 2018.

Hochreiter, S. and Schmidhuber, J. Long short-term memory. *Neural Computation*, 1997.

Jalili, M., Gharibshah, J., Ghavami, S. M., Beheshtifar, M., and Farshi, R. Nationwide prediction of drought conditions in iran based on remote sensing data. *IEEE Transactions on Computers*, 2013.

Keller, C. A., Knowland, K. E., Duncan, B. N., Liu, J., Anderson, D. C., Das, S., Lucchesi, R. A., Lundgren, E. W., Nicely, J. M., Nielsen, E., Ott, L. E., Saunders, E., Strode, S. A., Wales, P. A., Jacob, D. J., and Pawson, S. Description of the NASA GEOS composition forecast modeling system GEOS-CF v1.0. *Journal of Advances in Modeling Earth Systems*, 2021.

Khan, N., Sachindra, D., Shahid, S., Ahmed, K., Shiru, M. S., and Nawaz, N. Prediction of droughts over pakistan using machine learning algorithms. *Advances in Water Resources*, 2020.

Kubicz, J. The application of standardized precipitation index (SPI) to monitor drought in surface and groundwaters. In *E3S Web of Conferences*, 2018.

Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., and Yan, X. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *NeurIPS*, 2019.

Loshchilov, I. and Hutter, F. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Mishra, A. K. and Singh, V. P. A review of drought concepts. *Journal of Hydrology*, 2010.

Mishra, A. K., Desai, V. R., and Singh, V. P. Drought forecasting using a hybrid stochastic and neural network model. *Journal of Hydrologic Engineering*, 2007.

National Drought Mitigation Center. What is the U.S. drought monitor?, 2021. URL https://droughtmonitor.unl.edu/About/WhatistheUSDM.aspx.

Poornima, S. and Pushpalatha, M. Drought prediction based on SPI and SPEI with varying timescales using lstm recurrent neural network. *Soft Computing*, 2019.

Pylianidis, C., Snow, V., Holzworth, D., Bryant, J., and Athanasiadis, I. N. Location-specific vs location-agnostic machine learning metamodels for predicting pasture nitrogen response rate. In *Pattern Recognition. ICPR International Workshops and Challenges*, 2021.

Rhee, J., Im, J., and Park, S. Drought forecasting based on machine learning of remote sensing and long-range forecast data. *ISPRS*, 2016.

Serrano, V., Beguería, S., and López, M. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index–spei. *Journal of Climate*, 2010.

Smith, A. B. 2010-2019: A landmark decade of U.S. billion-dollar weather and climate disasters, Jan 2020.

Smith, L. N. A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. *arXiv preprint arXiv:1803.09820*, 2018.

Sparks, A. H. nasapower: A NASA POWER global meteorology, surface solar energy and climatology data client for R. *The Journal of Open Source Software*, 2018.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., et al. The drought monitor. *Bulletin of the American Meteorological Society*, 2002.

Tan, R. Wavelet-coupled machine learning methods for drought forecast utilizing hybrid meteorological and remotely-sensed data. In *DMIN*, 2018.

Tong, C. and de Witt, C. S. RainBench: Enabling data-driven precipitation forecasting on a global scale. 2021.

Van Loon, A. F., Tijdeman, E., Wanders, N., Van Lanen, H. A. J., Teuling, A. J., and Uijlenhoet, R. How climate seasonality modifies drought duration and deficit. *Journal of Geophysical Research: Atmospheres*, 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. In *NIPS*, 2017.

Wu, N., Green, B., Ben, X., and O'Banion, S. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv preprint arXiv:2001.08317*, 2020.

Xiao, C., Chen, N., Hu, C., Wang, K., Gong, J., and Chen, Z. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sensing of Environment*, 2019.

A. Hyperparameter Experiments - Results

	WEEK 1		*WEEK 2*		*WEEK 3*		*WEEK 4*		*WEEK 5*		*WEEK 6*	
	MAE	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1	*MAE*	F_1
LSTM	0.137	81.4	0.211	72.0	0.272	64.7	0.329	58.5	0.381	52.8	0.426	48.4
+ seasonal encoding	0.140	81.2	0.212	72.1	0.273	64.2	0.329	58.2	0.381	53.5	0.426	48.6
+ location identifier	0.137	81.1	0.212	71.1	0.274	65.0	0.331	57.2	0.382	52.3	0.426	48.2
+ lat/lon (binned)	0.141	81.5	0.213	71.5	0.272	**65.2**	0.329	57.5	0.380	53.1	0.425	48.7
+ previous year	0.147	77.9	0.209	66.2	0.261	56.8	**0.312**	52.4	**0.357**	54.4	**0.399**	48.8
combination	0.138	**81.8**	0.211	**72.2**	0.272	63.6	0.328	57.6	0.379	52.9	0.423	48.0
Transformer	0.142	74.4	0.215	61.6	0.272	56.1	0.332	47.5	0.380	42.2	0.424	39.4
+ seasonal encoding	0.157	75.0	0.232	64.4	0.289	57.0	0.346	46.7	0.396	47.2	0.438	39.2
+ location identifier	0.142	77.6	0.218	66.3	0.282	60.2	0.342	51.9	0.384	46.4	0.429	43.5
+ lat/lon (binned)	0.139	78.3	0.218	68.8	0.284	61.0	0.346	53.3	0.398	48.1	0.444	42.5
+ previous year	0.148	76.1	0.215	67.8	0.267	64.5	0.320	**60.7**	0.370	**56.4**	0.403	**49.6**
combination	**0.136**	79.7	**0.205**	69.6	**0.259**	63.7	0.314	57.9	**0.357**	54.5	**0.399**	49.4

Table 2. Mean Absolute Error (MAE) and macro F_1 score on the evaluation set for the LSTM and Transformer. Best results are in **bold** and the *underlined* data represents the best result for that model.

B. Hyperparameter Experiments - Model Setup

Hyperparam	**LSTM**	**Transformer**
Number of Layers	2	4
Hidden Size	512	512
Batch Size	128	128
Dropout Probability	0.1	0.1
Weight Decay	0.01	0.01
Learning Rate	7e-5	7e-5
Number of Epochs	7	7
FFNN inner hidden size	N/A	4096
Attention Heads	N/A	2
Initial Projection Size	N/A	256

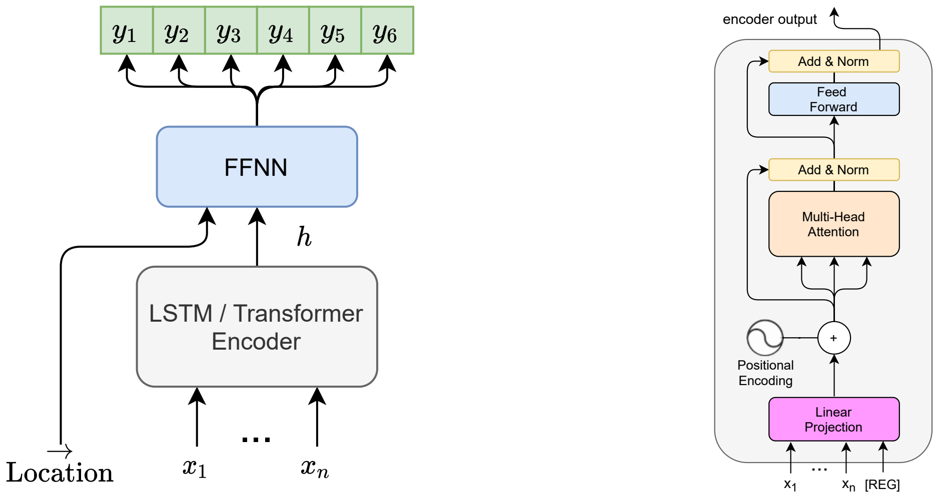
Table 3. Hyperparameters used for LSTM and Transformer models

C. Comparing SPI to USDM Drought Categories

The U.S. Drought Monitor (USDM) website published a table corresponding to the mapping between SPI values and USDM drought categories, which are shown in the table below (National Drought Mitigation Center, 2021).

USDM Category	**Description**	**SPI**
D0	Abnormally Dry	-0.5 to -0.7
D1	Moderate Drought	-0.8 to -1.2
D2	Severe Drought	-1.3 to -1.5
D3	Extreme Drought	-1.6 to -1.9
D4	Exceptional Drought	-2.0 or less

D. Model Architectures and Overview



(a) (b)

Figure 4. Figure (a) illustrates how we pass the encoded input sequence $x_1, ...x_n$ (combined drought, meteorological and season vectors) and the location vector to a feed-forward neural network (FFNN). Figure (b) illustrates the architecture of the Transformer model used in this paper with the `[REG]` token.

E. Experiment (c) Results - Comparing Model Performance on Local vs National Training Data

Training Data	Evaluation Data	*WEEK 1*	
		MAE	F_1
Iowa	Iowa	0.102	88.4
Montana	Montana	0.344	53.1
Oklahoma	Oklahoma	0.212	70.9
	Iowa	0.093	90.1
All	Montana	0.323	55.8
	Oklahoma	0.183	75.8

Table 4. Performance of the LSTM on specific states after training on all available data compared to just the data from those states.