Forecast of Wildfire Potential Across California USA Using a Transformer

1st Russell Limber
Bredesen Center for Interdisciplinary Research
The University of Tennessee
Knoxville, TN, USA
limberrl@ornl.gov

2nd William W. Hargrove
Southern Research Station
USDA Forest Service
Asheville, NC, USA
william.w.hargrove@usda.gov

3th Forrest M. Hoffman
Computational Earth Sciences
Oak Ridge National Laboratory
Oak Ridge, TN, USA
hoffmanfm@ornl.gov

4nd Jitendra Kumar
Environmental Science
Oak Ridge National Laboratory
Oak Ridge, TN, USA
kumarj@ornl.gov

2024 IEEE International Conference on Big Data (BigData) | 979-8-3503-6248-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/BigData62323.2024.10825778

***Abstract*—Wildfires are a major issue facing the United States, a matter further exacerbated by an ever-changing climate. In California alone, wildfires are responsible for billions of dollars in damages and take lives each year. Accurately predicting fire danger conditions allows preparation awareness before wildfires start. Transformers are a class of deep learning models designed to identify patterns in sequential datasets. In recent years, transformers have gained popularity through their impressive performance in natural language processing and other applications of signal recognition. This analysis demonstrates the ability of a transformer with a residual connection to forecast fire danger potential over the state of California. Wildland fire potential index (WFPI) maps collected from the US Geological Survey database from January 1st 2020 to December 31st 2023 were used to tune, train and evaluate the transformer. Meteorological inputs (provided by Daymet daily weather and climatological summaries), the normalized difference vegetation index (NDVI) (calculated from the Moderate Resolution Imaging Spectroradiometer (MODIS)), and outputs from the Scott and Burgman fire behavior fuel models (to characterize maps of fuel types), were used as inputs. Our results show that a transformer can effectively emulate the US Forest Service modeled WFPI maps of California USA for four week long forecasts over the month of July, 2023, with correlations ranging from 0.85 – 0.98.**

***Index Terms*—transformer, residual connection, wildfires, time series, remote sensing.**

I. Introduction

The state of California is having a wildfire crisis. The increased frequency and severity of wildfires across California in recent years can be attributed to a complex system of ecological and climatological drivers. Summer air temperatures have been found to be steadily increasing as far back as 1950, leading to more frequent droughts and drier conditions. This increase in temperature paired with changing precipitation patterns and earlier snow melt, has led to a prolonged fire season [1]. A comprehensive cost analysis, taking into account property damage, lost earnings, closures, etc. found that the incurred losses to the state of California between 2017 and 2021, as a result of wildfires, was $117.4 billion [2]. This included the most deadly wildfire in California's history, the "Camp Fire," which took the lives of 85 people in November of 2018 [3]. Being better able to forecast fire ignition and spread potential gives residents more time to prepare and evacuate. Improving early warning systems can mitigate the impacts of wildfires by granting first responders more time to take action, which can protect property and save lives.

The goal of this analysis is to apply a state of the art deep learning model, a time series based residual transformer, to forecast fire danger potential. The wildland fire potential index (WFPI) dataset used to train the transformer in this analysis was provided by the US Geological Survey in partnership with the US Forest Service (USGS/USFS). The transformer model is built as an emulator to predict WFPI. This will provide a computationally efficient and scalable method to forecast fire danger potential across large regions, using fewer inputs as well as inputs that are easier to collect on a large regional scale. While recurrent neural networks make updates sequentially on a given cell state, transformers do not have this bottleneck, making them easier to parallelize [4]. This is a helpful feature in geospatial analysis, as geospatial analysis tends to examine many grid cells over space and time leading to very large datasets. Due to the sequential pattern recognition capability of this framework, we hypothesize that it will be capable of making robust forecasts up to one week out - comparable to the current one week forecast window provided by the current USGS/USFS WFPI product.

This research was supported in part by the United States Department of Agriculture, US Forest Service, Southern Research Station. Additional support was provided by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area (SFA), which is sponsored by the Regional and Global Model Analysis (RGMA) activity in the Earth and Environmental Systems Sciences Division (EESSD) of the Biological and Environmental Research (BER) office in the US Department of Energy Office of Science.

979-8-3503-6248-0/24/$31.00 ©2024 IEEE

II. Related Work

The fusion of remote sensing and machine learning to predict the occurrence, and or risk, of wildfires is an ever growing field of research. Machine learning methods have been used to expand upon physics-based fire prediction models. Cheng et al. devised a parameter optimization framework for more computationally efficient tuning and improved performance of deterministic fire prediction models. Their analysis also used wildfires in California for evaluation [5]. Utilizing machine learning for fire risk modeling (similar to this analysis) is another common application. XGBoost was shown to be superior to random forest and logistic regression when evaluating the normalized difference vegetation index (NDVI) as a proxy for fire risk [6]. Fire detection (classification) is also a common application of machine learning. Sayad et al. utilized support vector machines and a deep neural network (DNN) to classify fire over a large geospatial dataset housed on the Databricks platform. Their detection model was highly effective, with a recall of 98% [7]. A DNN using a weight-selection strategy was also used over Alaska USA to detect wildfires with a recall of 96% [8]. In many use cases, machine learning and statistical techniques that have been available for decades, such as logistic regression, DNNs and support vector machines, are the typical tools of choice in wildfire detection [9] with more modern machine learning methods often outperforming traditional methods [10].

III. Data

A. Wildland Fire Potential Index

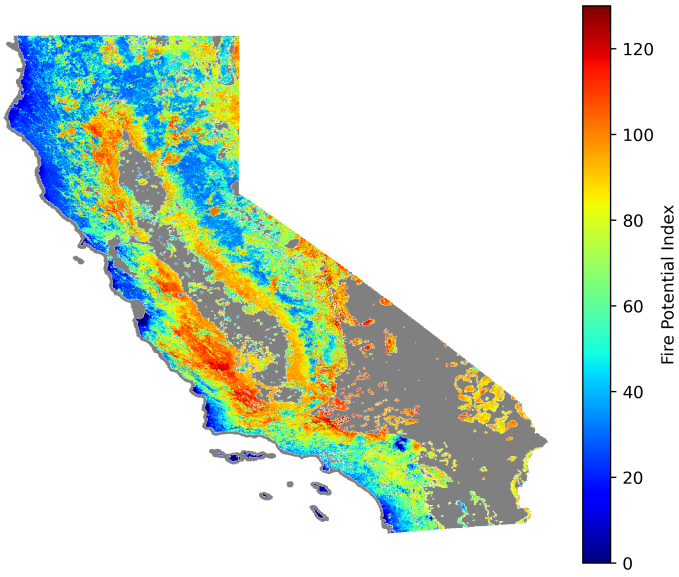

Fig. 1. Map of California USA showing an example wildland fire potential index (WFPI) map for June 15th 2022. The colorbar represents the WFPI values; grey shading are regions where either the US Forest Service does not identify a wildfire threat, or had missing data and were unable to be determined.

In this study we developed a transformer model for an autoregressive task. The goal was to forecast wildland fire potential indices (WFPI). The WFPI is a unitless measure with values in the range of 0 to 150. They are designed to correlate to vegetation flammability, whereby larger values signify a greater chance of fire ignition and spread. WFPI values are derived from a statistical and deterministic algorithm that uses the ratio of live to dead fuel, wind speed ($\frac{\mathrm{m}}{\mathrm{s}}$), dry bulb temperature (°C), precipitation (mm) and normalized difference vegetation index (NDVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) [11]. The USGS/USFS provides WFPI products at spatial resolution of around 1.25 km $\times$ 1.25 km. We designed our transformer to emulate and predict WFPI maps spanning California, with WFPI data serving as the dependent variable to tune, train and evaluate the transformer.

The USGS/USFS WFPI database provides forecasts for daily fire danger potential, up to seven days into the future. The USGS/USFS WFPI data is treated as the ground truth. We use the one day forecasts for training the transformer model as they are considered to be the closest to true WFPI. An example WFPI map for June 15th 2022 is shown in Fig. 1. All WFPI data were downloaded from the USGS/USFS database: https://firedanger.cr.usgs.gov/apps/staticmaps.

B. Daymet Daily Surface Weather and Climatological Summaries

Daymet provides a gridded dataset, at a spatial resolution of 1 km $\times$ 1 km, of daily surface weather and is derived by interpolating and extrapolating meteorological data from in situ instruments and meteorological stations [12]. Daymet has been shown to demonstrate high precision in modeling meteorological variables across North America [13]. Daily minimum air temperature (°C), maximum air temperature (°C), shortwave radiation ($\frac{\mathrm{W}}{\mathrm{m}^2}$), snow water equivalent ($\frac{\mathrm{kg}}{\mathrm{m}^2}$) and precipitation (mm) were obtained from Daymet to provide inputs to the transformer model. The Daymet dataset was downloaded from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC): https://daac.ornl.gov/cgi-bin/dataset_lister.pl?p=32.

C. Moderate Resolution Imaging Spectroradiometer

The MODIS (Moderate Resolution Imaging Spectroradiometer) is a vital instrument on board the satellites Terra and Aqua. Composite images using a 16 day temporal window at a 250 meter resolution were collected. While 36 spectral bands are provided by MODIS, we opted to use the higher-order integrative NDVI. NDVI is a unitless metric calculated using two of the bands collected by MODIS, i.e. near infrared (NIR: 0.845 – 0.885 μm) and visible red (R: 0.630 – 0.680 μm), radiation, Equation 1:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \tag{1}$$

NDVI is a venerable indicator of how forests are impacted by precipitation or lack thereof [14], [15]. Furthermore, NDVI from MODIS is a common metric applied to many wildfire prediction and impact models [6]–[8] and maintains

continuity with the calculation of WFPI. All MODIS images were downloaded from the Land Processes Distributed Active Archive Center [16]: https://appeears.earthdatacloud.nasa.gov/task/area.

D. Scott and Burgan Fire Behavior Fuel Dataset

40 Scott and Burgan Fire Behavior Fuel Model (FBFM40) developed by US Forest Service LANDFIRE program provides distinct distributions of fuel loading found among surface fuel components (live and dead), size classes, and fuel types. The Scott and Burgan fire behavior fuel data are developed by a set of standard fire behavior fuel models in conjunction with Rothermel's surface fire spread model [17]. The fuel models are described by the most common fire carrying fuel type (grass, brush, timber, or slash), loading and surface area-to-volume ratio by size class and component, fuel bed depth, and moisture of extinction [18]. These products are developed using expert rulesets to understand how different types of disturbance would change pre-disturbance fuel models to post disturbance fuels, based on the severity and time since disturbance. The LANDFIRE FBFM40 2020 covering the state of California for the year 2020 was used in this analysis and serves as the only static variable used in the model. The product originally available at 30 m spatial resolution, was resampled to 1.25 km for use in our analysis. Data provides information about the ground fuel conditions for our forecasting model. While the fuel models are provided as descrete classification maps, we used the following continuous metrics, characterizing them as inputs in our modeling: fine fuel load ($\frac{\text{tons}}{\text{acre}}$), characteristic surface-area-to-volume ($\frac{1}{\text{m}}$), packing ratio, extinction moisture content (%), flame length (m), rate of spread ($\frac{\text{m}}{\text{s}}$). The Scott and Burgan fire behavior fuels dataset were downloaded from the LANDFIRE program of the U.S. Department of Agriculture, Forest Service and U.S. Department of the Interior: https://landfire.gov/fuel/fbfm40

E. Data Synthesis and Processing

Datasets described in the above section were synthesized and processed as illustrated in the upcoming section. A summary of the variables used in the analysis is available on Table I.

1) Data import: The gridded geospatial raster datasets were downloaded via respective databases and processed for use in our analysis. All raster datasets were resampled to 1.25 km $\times$ 1.25 km spatial resolution, matching the WFPI data products. Rasters with continuous data were resampled using a bilinear interpolation model while the discrete/classification fuels data was downscaled using k-nearest neighbor. Rasters were re-projected to a geographical latitudinal and longitudinal coordinate reference system and combined through space and time to form a time series of matrices:

$$\mathbf{X}_i,\ \mathbf{X}_{i+1}, \mathbf{X}_{i+2}, \mathbf{X}_{i+3} \dots\ \mathbf{X}_{i+k} \tag{2}$$

Each matrix in Equation 2 refers to a four dimensional matrix with the dimensions: latitude, longitude, lookback window, inputs (or features); for a given iteration or day, *i*, to the end of the time series (*i* + *k*). The matrices were staggered by the forecast window which in this analysis was one day. Therefore, matrix $\mathbf{X}_{i+1}$ was the same as matrix $\mathbf{X}_i$ except its lookback window starts and ends one day later. The lookback window length used in this analysis was 30 days. This was the depth of past information that will be available to the transformer to make a forecast on the subsequent WFPI values. For example, since the start of the time series is January 1 2020, the first matrix, $\mathbf{X}_i$, represents all of the inputs from January 1 2020 to January 30 2020 over California, which will then be used to predict WFPI values for January 31st 2020, represented as $\mathbf{y}_{i+1}$. Note that predicted forecast matrix $\mathbf{y}_i$ is a subset of matrix $\mathbf{X}_i$ whereby $\mathbf{y}_i \subseteq \mathbf{X}_i$ on the inputs dimension. This means that the past WFPI values over the lookback window are able to be leveraged by the transformer to forecast the subsequent WFPI values at $\mathbf{y}_{i+1}$. After checking for outliers, matrices $\mathbf{X}_i \dots\ \mathbf{X}_{i+k}$ were normalized using the global minimum and maximum values over the time series for each respective input variable.

2) Missing Values: Of the 1,460 days from January 1, 2020 to December 31, 2023, 46 were missing WFPI values. These days were simply omitted from the analysis. To account for these gaps, temporally embedded features (*TEFs*) were added to the inputs dimension. These features were derived as shown by Equation 3 and Equation 4:

$$h(d_i) = (d_i - 1)\left(\frac{360}{364}\right)\left(\frac{\pi}{180}\right) \tag{3}$$

$$[\text{TEF}_x,\ \ \text{TEF}_y] = \begin{cases} \cos(h(d_i)) \\ \sin(h(d_i)) \end{cases} \tag{4}$$

Where d_i is the Gregorian day of the year. Leap years were removed following a 365 day calendar in line with

TABLE I
TABLE OF VARIABLES USED IN THE ANALYSIS WITH THE SOURCE OF THE DATA, FEATURES USED FROM THE DATA SOURCE AND THE ORIGINAL SPATIAL RESOLUTION, SHOWN.

Data Source	**Feature(s)**	**Original Spatial Resolution**
US Geological Survey and the US Forest Service	wildland fire potential index (WFPI)	1.25 km $\times$ 1.25 km
Daymet Daily Surface Weather and Climatological Summaries	daily minimum air temperature (°C), maximum air temperature (°C), shortwave radiation ($\frac{\text{W}}{\text{m}^2}$), snow water equivalent ($\frac{\text{kg}}{\text{m}^2}$) and precipitation (mm)	1.0 km $\times$ 1.0 km
Moderate Resolution Imaging Spectroradiometer	normalized difference vegetation index (NDVI)	250 m $\times$ 250 m
US Forest Service LANDFIRE Program; 40 Scott and Burgan Fire Behavior Fuel Model	fine fuel load ($\frac{\text{tons}}{\text{acre}}$), characteristic surface-area-to-volume ($\frac{1}{\text{m}}$), packing ratio, extinction moisture content (%), flame length (m), rate of spread ($\frac{\text{m}}{\text{s}}$)	30 m $\times$ 30 m

the meteorological inputs from Daymet. These *TEFs* provide temporal context to the transformer by emulating the days of the year as points around a unit circle and then taking the components of that circle using cos and sin.

In every MODIS image there are regions with missing data, especially during the cloudiest periods of the year (generally late spring and early summer) present throughout California. This means that spatially, NDVI values can have regions that are missing values. The same holds true for WFPI values. However, since the average input matrix in Equation 2 contains about 150,000 complete 30 day time series, we chose to not patch missing data areas, since the per day sample sizes were sufficient for model training.

3) Data pipeline: All of the input matrices in Equation 2 were flattened over the latitudinal and longitudinal dimensions. A generator function was created to read the data for $\mathbf{X}_i$ and retrieve $\mathbf{y}_i^{(l)}$ from matrix $\mathbf{X}_{i+1}$ where $\mathbf{y}_i^{(l)}$ are the WFPI values of the final day for the next day matrix $\mathbf{X}_{i+1}$ (l superscript, representing the final slice along the lookback window dimension). The data was partitioned in a training, validation, testing split of 70%, 15%, 15%, respectively. Although a forecast greater than seven days will likely never be necessary with such a model, we chose to apply these cutoffs as it places the start of our testing data right at the beginning of the fire season in California (July 1, 2023). This makes for an ideal evaluation to assess how well the model does during the fire season of the final year.

IV. Residual Transformer

The transformer in this analysis (Fig. 3) was implemented using Tensorflow and Keras with hyperparameter tuning functionality built on top of the Keras tuner API [19]. All computations were carried out on the National Energy Research Scientific Computing Center (NERSC), Perlmutter supercomputer. Perlmutter, based on the HPE Cray Shasta platform, is a heterogeneous system comprised of 3,072 CPU-only (AMD EPYC 7763 CPUs) and 1,792 GPU-accelerated nodes (NVIDIA A100 GPUs).

Input matrix $\mathbf{X}_i$ was flattened over the latitudinal and longitudinal dimensions, operated on by the transformer, then reshaped back for visualization and evaluation. A residual connection was implemented, since past dependent variable values $\mathbf{y}_i^{(l)}$ are autocorrelated and therefore can effectively help forecast subsequent values $\mathbf{y}_{i+1}^{(l)}$. Practically speaking, this means that the WFPI values over the grid prior to the day being forecasted were mapped to the output of the transformer (function F) and the sum taken, Equation 5:

$$\hat{\mathbf{y}}_{i+1}^{(l)} = \mathbf{y}_i^{(l)} + F(\mathbf{X}_i) \tag{5}$$

This residual connection is represented by the schematic in Fig. 2. Should the transformer be used to forecast more than one day at a time, the predicted WFPI values will be used in place of the actuals. This concept is illustrated in Equations 6:

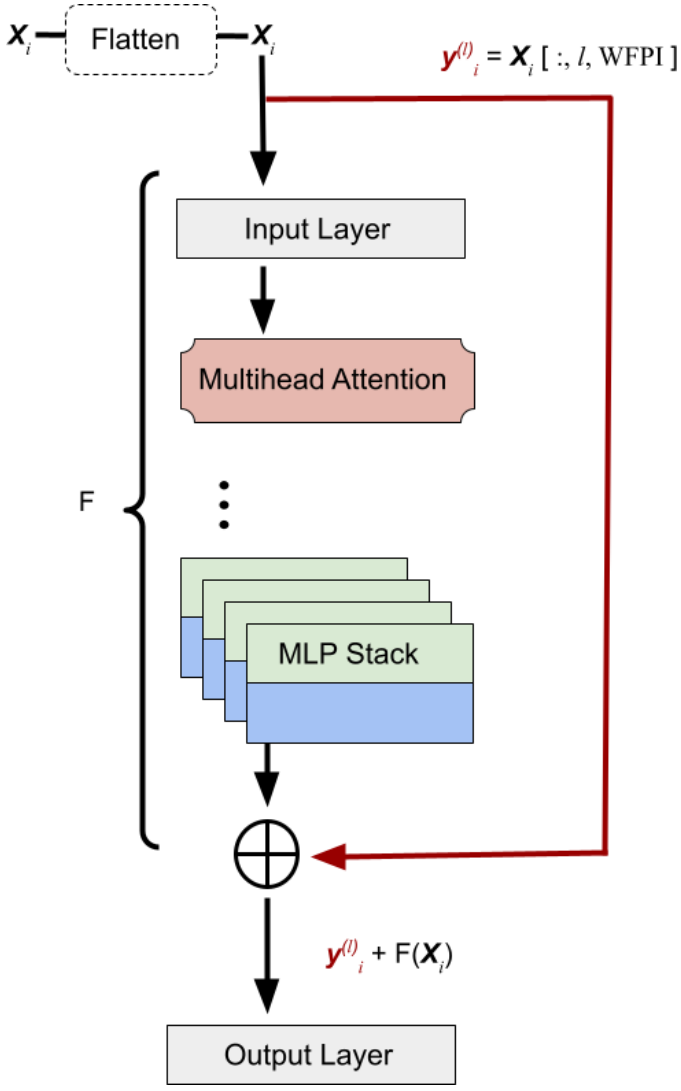

Fig. 2. Schematic showing the residual connection of the transformer. Function F: represents the hidden layers and subprocesses of the transformer. $\mathbf{X}_i$: the input matrix which is flattened on the latitudinal and longitudinal dimensions prior to entry into the model. $\mathbf{y}_i^{(l)}$: the dependent variable, wildland fire potential index (WFPI), for the last day of the time series in the lookback window of $\mathbf{X}_i$ represented by l.

$$\begin{aligned}
\hat{\mathbf{y}}_{i+2}^{(l)} &= \hat{\mathbf{y}}_{i+1}^{(l)} + F(\hat{\mathbf{X}}_{i+1}); \\
\hat{\mathbf{y}}_{i+3}^{(l)} &= \hat{\mathbf{y}}_{i+2}^{(l)} + F(\hat{\mathbf{X}}_{i+2}); \\
&\dots \\
\hat{\mathbf{y}}_{i+k}^{(l)} &= \hat{\mathbf{y}}_{i+k-1}^{(l)} + F(\hat{\mathbf{X}}_{i+k-1});
\end{aligned} \tag{6}$$

In a real-world application, the other inputs will likewise be replaced by forecasted information, such as one week weather forecasts in place of Daymet, seasonally imputed NDVI values, etc. This analysis uses the continuous time series of input variables for training through to evaluation.

During all epochs for tuning and training the transformer is evaluated using the means squared error (MSE) of the validation data. MSE is used in place of loss because the

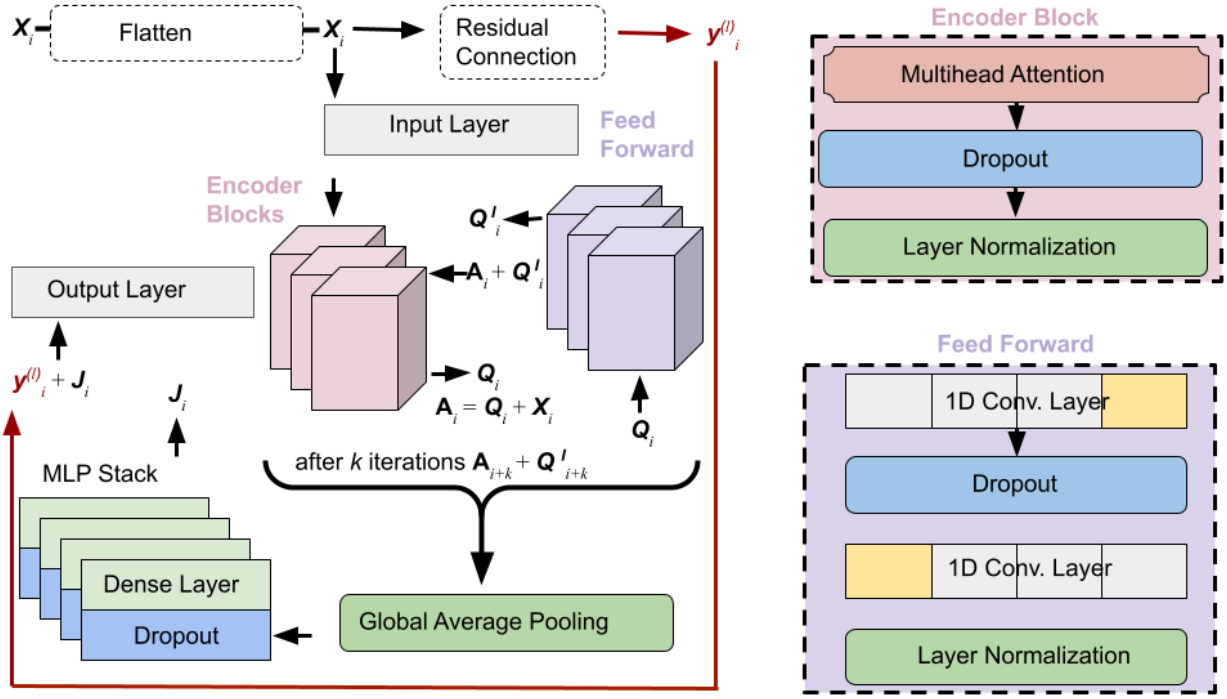

Fig. 3. Schematic high level overview of the entire transformer algorithm. The red line shows the steps of the residual connection as the dependent variable for the current iteration is temporarily removed to be used later to estimate the subsequent iteration. The process of data moving through the encoder blocks and being operated on by the feed forward partition is illustrated. Here k is some number of iterations equivalent to the number of encoder blocks which in this case was three. The MLP (multi-layer perceptron) stack, symbolizes the data being processed by the final dense layers with dropout. Note: the actual model only used one such layer in the MLP stack as was selected through tuning, the four pictured here are just for illustration.

transformer is learning the difference between the prior day WFPI values and the shift in those WFPI values, function F (Equation 5). Weighing more heavily against erroneous large fluctuations in WFPI between $\mathbf{y}_i^{(l)}$ and $\mathbf{y}_{i+1}^{(l)}$ is a helpful strategy. It penalizes the transformer to a greater degree for deviating too much and too quickly from past WFPI values.

On average, each input matrix $\mathbf{X}_i$ over the lookback window dimension (30 days) has about 150,000 complete time series over all 16 input features. We use a sub-batch of 50,000 randomly selected cells at each iteration so as to save on compute time. This means that on average one third of the grid cells are chosen at random and read into the transformer with every iteration.

A. Bayesian Hyperparameter Selection

As the model architecture and hyperparameter space are relatively large in this analysis, a grid search optimization scheme would have been too computationally intensive. While there exist several parameter search algorthims that tend to outperform a random selection such as hyperband [20], we opted to use a Bayesian optimization scheme, as it has strong convergence potential over relatively few trials [21]. Table II shows the proposed hyperparameter search space, incremental adjustment size by trial (step), and what the ultimate selection was. The final selections made in Table II were determined by finding the hyperparameter arrangements that minimized validation MSE. We ran 64 hyperparameter trials for 25 epochs each with a patience of seven. A patience of seven means that if the model did not show improvement for seven straight epochs, it would terminate the tuning at that trial and select the epoch with the lowest validation MSE and record it for comparison.

TABLE II

TABLE SHOWING THE BAYESIAN HYPERPARAMETER OPTIMIZATION SCHEME APPLIED TO THE TRANSFORMER. EACH HYPERPARAMETER WAS ASSESSED BY MINIMIZING THE MEAN SQUARED ERROR (MSE) OF THE VALIDATION DATASET. HYPERPARAMETER SPACE, STEP SIZE BY TRIAL AND THE FINAL HYPERPARAMETER SELECTED FOR THE TRANSFORMER, ARE SHOWN.

Hyperparameter	Range of Values	Step	Selection
learning rate	1.0×10^{-5} – 1.0×10^{-1}	10*	1.0×10^{-4}
batch size	96 – 448	32	256
encoder blocks	2 – 4	1	3
feed forward dimensions	32 – 64	8	64
encoder dropout	0 – 0.5	0.05	0.2
transformer heads	{2, 4, 8}	NA	4
attention head size	16 – 128	8|16**	64
MLP layers	1 – 3	1	1
MLP neurons	16 – 256	16	64
MLP activation	{relu, leaky-relu, tanh}	NA	relu
MLP dropout	0 – 0.5	0.05	0.2

*Selected along a logarithmic curve: $1.0 \times 10^{-5}, 1.0 \times 10^{-4} \ldots 0.1$
**Either 8 or 16 is multiplied by the head size being tuned

V. Results

A. Accuracy and Spatial Patterns of WFPI Forecast

The primary goal of our study was to develop a transformer model capable of emulating the USGS/USFS WFPI maps for the state of California. Landscape scale patterns of fire potentials are critically important for the practical and operational use of WFPI datasets for resource planning and allocations. Thus, beyond pixel-wise accuracy of prediction, accurate spatial patterns and coherence is of utmost importance. We evaluated the WFPI forecasts based on transformer model for spatial correlation with the benchmark test datasets. Fig. 4, shows an example one day prediction for July 22, 2023. The spatial patterns of the transformer-based forecast (Fig. 4A were very similar to those from USGS/USFS products Fig. 4B). Overall, the transformer model one day forecast, achieved a spatial correlation of 0.98 for the test dataset. Examination of the distribution of WFPI values across the study region (Fig. 4C), shows the model to have a slight bias towards underpredicting the most extreme WFPI values, however it exhibit similar distribution of the WFPI values overall.

B. Seven Days Forecast of WFPI

The USGS/USFS provides operational forecast of WFPI for up to seven days, updated daily. We applied our trained transformer model in a similar fashion to seven days forecast window comparable to the USGS/USFS products. We started on the first day of July 2023 and ran the model for seven days forecast. We then repeated the process for the second, third and fourth weeks of July.

Fig. 5 shows the heatmap plot of spatial correlation between predicted and actual WFPI across California for the 7 days forecasts conducted for the four weeks of July 2024. The transformer model performed very well during all four weeks for forecasting one and two days ahead, however, the correlations were slightly weaker for three to seven day ahead forecasts. This gradual loss of accuracy can be explained by the use of modeled WFPI values for the residual connections (Equations 6) and the potential compounding of errors when applied to longer lead forecsts. Nonetheless, the correlations between maps remains high, with the lowest correlation on the third forecasted day of the fourth week (July 24, 2023) the lowest at 0.85, and all other correlations for the 4 weeks tested between 0.88 – 0.98.

C. Computational Performance of Transformer Model

Hyperparameter tuning was conducted using 12 nodes on Perlmutter, each node with 4 GPUs and required 42 hours of computation. Each node operated on a single trial and upon completion, performed an allreduce, along with updates to the hyperparameter weights to determine next steps for searching the hyperparameter space. Hyperparamter search instructions were passed using a chief to worker node paradigm, whereby one of the nodes was set to be the chief. The transformer was set to train for up to 200 epochs with a patience of 20 (with the option to add more if patience was not yet reached).

The transformer reached its minimum validation MSE at 31 epochs for the hyperparameter configuration shown in Table II. Since the model had actually maxed out the number of epochs for the optimal hyperparameter trial, it effectively trained for 25 epochs prior. This makes the total number of epochs for training effectively 56 since training began from the final model parameter configuration of the last epoch of tuning. The training step utilized 6 Perlmutter nodes with 4 GPUs per node and took a total of 24 hours to complete.

The trained transformer was able to map WFPI forecasts for all of California for four weeks of July 2023 (results from which are presented here) in six minutes and thirty-two seconds, using two CPUs each with 64 cores.

VI. Discussion

A. Accuracy of Transformer Model Forecast

Overall the results of the residual transformer are very promising. In particular, the results of the one and two day forecasted WFPI values are very good with correlations reaching 0.97 and 0.98. The subsequent forecasts show good correlation with room for improvement. While it is normal for autocorrelated algorithms to diverge the further into the future a forecast is made, we believe this model architecture can be improved. Nonetheless, this analysis shows that a residual transformer can be a powerful tool in emulating WFPI mapping, which can aid first responders and communities in preparing ahead for wildfires. This can ultimately help improve the computational efficiency of WFPI forecasts and the time required to update and issue forecasts. These actions support fire related operational decisions, saving property and lives.

B. Assumptions and Data Limitations

Two key assumptions were made in the input datasets used in the study, which impact the model accuracy and limits the application of the model in an operational scenario. The first pertains to the meteorological features available from Daymet. Daymet variables are produced by interpolating and extrapolating over space and time, using many different meteolorgical stations and data collection instruments [12]. However, Daymet is not available as a forecast product. The forecasts ranging from one to seven days in the USGS/USFS WFPI maps are produced through a reanalysis data product with forecasts [11]. This means that the meteorological variables used from Daymet may not correspond well with the forecasted weather data used in the WFPI maps which can create contrasts between the predictions and the actual WFPI maps, especially the further into the future forecasts are made. One alternative might be to supplement the Daymet data in the analysis with data from the European Centre for Medium-Range Weather Forecasts Reanalysis 5th Generation (ERA5) dataset [22]. The ERA5 product would be more likely to resemble forecasted weather data and can support operational deployment of a transformer. Daymet does tend to have a lower degree of uncertainty across North America [13] relative to ERA5 [23], which results from how the two products are derived. Therefore, using the Daymet inputs for model tuning

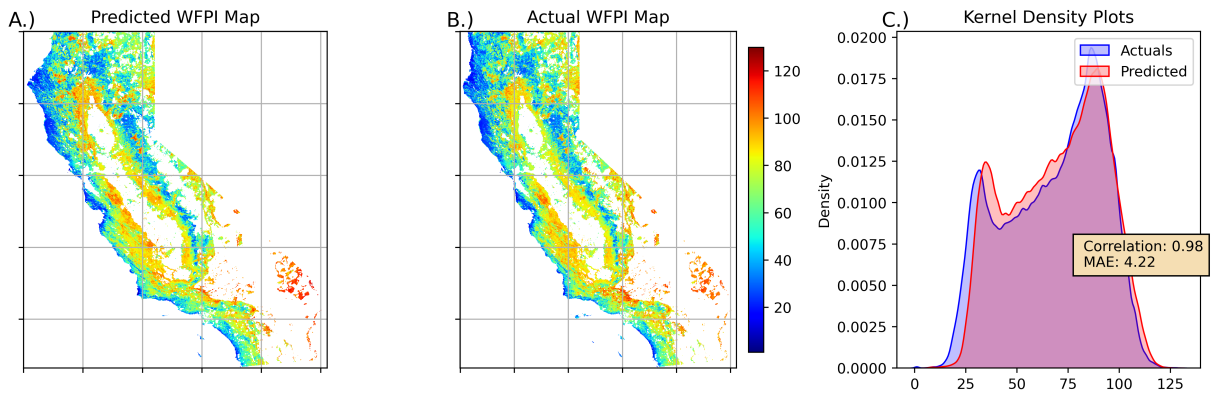

Fig. 4. Comparison of actual and predicted WFPI maps. A: map of the wildland fire potential index (WFPI) values as modeled by the transformer for a next day prediction. Values are for a one day prediction, July 22, 2023. B: a map of the actual WFPI values for July 22, 2023. C: kernel density plots comparing the distributions of the predicted WFPI values to the actual WFPI values. Plot also shows the Pearson correlation coefficient between the non-missing values of the two maps and the mean absolute error (MAE).

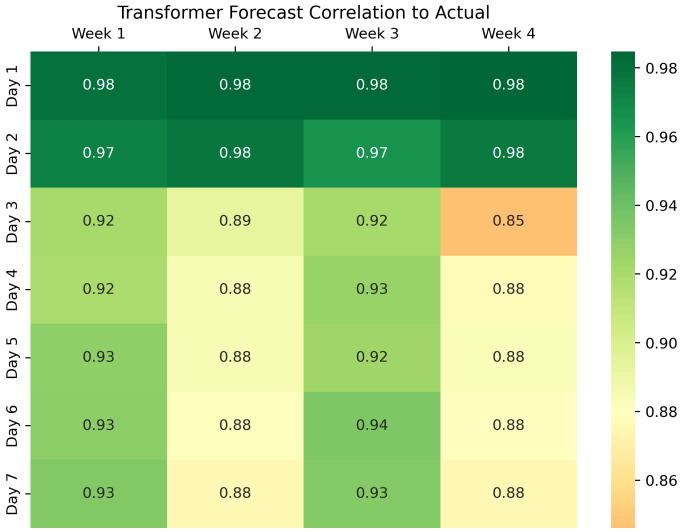

Fig. 5. Correlation heat map between the predicted wildland fire potential index (WFPI) and the actual WFPI maps, using the Pearson correlation coefficient. Heat map shows the results of the 7 day forecasts for four weeks during the middle of the 2023 fire season in California, used for evaluation. Week 1: 07/01 – 07/08; Week 2: 07/08 – 07/15; Week 3: 07/15 – 07/22; Week 4: 07/22 – 07/29

and training to take advantage of the low uncertainty and then leveraging the ERA5 product for its forecasting capability would be a logical next step for our future work to transition from research to operations.

The second assumption made in the inputs was the fuel type maps that are static and do not capture the dynamically changing fuel conditions on the landscape. This static fuel map was derived directly from the expected values of six fuel variables commonly used in fuel classification by the USGS/USFS LANDFIRE. Since land cover and fuel classification typically do not change as frequently as inputs like phenology (NDVI) or precipitation, treating these as static variables was a reasonable assumption. However, for how long it remains suitable is an active topic of research in fire ecology. It would likely be beneficial to incorporate additional fuel type maps that capture the dynamics of fuel landscapes at seasonal to annual basis. The manner in which these variables are embedded within the forecast model could also be improved, by incorporating additional clustering steps in an effort to better discriminate between land fuel types.

C. Improvements to Residual Transformer

While the Bayesian hyperparameter tuner included batch size as a hyperparameter, this value was modified during training as the training time became prohibitively long. Tuning revealed that a reduced batch size (96) seemed to lead to better model performance, but as this was taking a relatively long time, a batch size of 256 was used instead for the training. By reducing the batch size we would almost certainly see an improvement in model performance. In addition, we postulate that the multihead attention of the transformer, paired with the seasonal relationships in fire weather data, would still allow models such as this one to make effective forecasts as far out as seven days. While the model showed good correlation to the actual maps before the third day of the weeks tested, from the third day onward, predictions were noticeably less correlated. One possible strategy would be to include a long-term memory layer prior to the global average pooling step, in order to track long term memory in the system. The use of gated recurrent units might also be effective in accomplishing this goal.

VII. Conclusion

Wildfires are a major issue facing the state of California. As temperatures continue to rise and rainfall patterns in much of the state become more erratic, the wildfire season is projected to grow longer each year. Sequential deep learning models such as transformers have the potential to emulate, and may in some cases even replace, many of the current more-manual early warning detection models used for wildfires and other similar disasters. We collected daily modeled wildland fire

potential index (WFPI) maps from the LANDFIRE program developed by US Geological Survey in partnership with the US Forest Service (USGS/USFS), from January 1st 2020 to December 31st 2023 at a one day forecast, to serve as ground truth for model tuning, training and evaluation. We then leveraged Daymet meteorological variables and normalized difference vegetation index (NDVI) values from the Moderate Resolution Imaging Spectroradiometer (MODIS); and used the Scott and Burgman fire behavior fuel model outputs to estimate land fuel types. We paired these input variables with temporally embedded features that we extracted using the cos and sin of days of the calendar year converted to radians of the unit circle. A transformer with a residual connection was tuned to predict WFPI values using a Bayesian hyperparameter search algorithm. Post-training, the model showed very good results, and was able to emulate the actual WFPI maps for the month of July, 2023, with Pearson correlation coefficient values ranging from 0.85 – 0.98. Our study demonstrates the potential for transformer models to modernize the forecasts of WFPI at large scales, by leveraging a wide range of datasets and computational resources, to help address the wildfire crisis.

ACKNOWLEDGMENT

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231 for project m2467. This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] "Public comment draft california's historical fire activity before modern fire suppression," California Air Resources Board, 1001 I Street, Sacramento, CA 95814, Technical Report, November 2021.

[2] J. Paci, M. Newman, and T. Gage, "The economic, fiscal, and environmental costs of wildfires in california," Gordon and Betty Moore Foundation, 1661 Page Mill Road, Palo Alto, CA 94304, Technical Report, August 2023.

[3] J. Byrnes, I. Sandoval-Cervantes, P. Acosta, J. Andronowski, J. Beatrice, W. Belcher, C. Bird, J. Bird, Z. Crossland, R. Depp *et al.*, *The Marginalized in Death: A Forensic Anthropology of Intersectional Identity in the Modern Era*. Lexington Books, 2022. [Online]. Available: https://books.google.com/books?id=PIOFEAAAQBAJ

[4] S. Karita, N. Chen, T. Hayashi, T. Hori, H. Inaguma, Z. Jiang, M. Someki, N. E. Y. Soplin, R. Yamamoto, X. Wang, S. Watanabe, T. Yoshimura, and W. Zhang, "A comparative study on transformer vs rnn in speech applications," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 449–456.

[5] S. Cheng, Y. Jin, S. P. Harrison, C. Quilodrán-Casas, I. C. Prentice, Y.-K. Guo, and R. Arcucci, "Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling," *Remote Sensing*, vol. 14, no. 13, 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/13/3228

[6] Y. Michael, D. Helman, O. Glickman, D. Gabay, S. Brenner, and I. M. Lensky, "Forecasting fire risk with machine learning and dynamic information derived from satellite vegetation index time-series," *Science of The Total Environment*, vol. 764, p. 142844, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720363749

[7] S. Younes Oulad, M. Hajar, and A. Hassan, "Predictive modeling of wildfires: A new dataset and machine learning approach," *Fire Safety Journal*, vol. 104, pp. 130–146, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0379711218303941

[8] Z. Langford, J. Kumar, and F. Hoffman, "Wildfire mapping in interior alaska using deep neural networks on imbalanced datasets," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 770–778.

[9] G. Bayat and K. Yıldız, "Comparison of the machine learning methods to predict wildfire areas," *Turkish Journal of Science and Technology*, vol. 17, no. 2, p. 241–250, 2022.

[10] K. Gholamnia, T. Gudiyangada Nachappa, O. Ghorbanzadeh, and T. Blaschke, "Comparisons of diverse machine learning approaches for wildfire susceptibility mapping," *Symmetry*, vol. 12, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2073-8994/12/4/604

[11] J. C. Eidenshink and S. M. Howard, "United states geological survey fire science: Fire danger monitoring and forecasting," U.S. Geological Survey, Reston, VA, Tech. Rep. 2012-3121, 2012, geographic Analysis and Monitoring Program. Contributing office: Earth Resources Observation and Science (EROS) Center. [Online]. Available: https://pubs.usgs.gov/fs/2012/3121/

[12] M. Thornton, R. Shrestha, Y. Wei, P. Thornton, S.-C. Kao, and B. Wilson, "Daymet: Annual climate summaries on a 1-km grid for north america, version 4 r1," 2022. [Online]. Available: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2130

[13] P. E. Thornton, R. Shrestha, M. Thornton, S.-C. Kao, Y. Wei, and B. E. Wilson, "Gridded daily weather data for north america with comprehensive uncertainty quantification," *Scientific Data*, vol. 8, no. 1, p. 190, 7 2021. [Online]. Available: https://doi.org/10.1038/s41597-021-00973-0

[14] F. Lloret, A. Lobo, H. Estevan, P. Maisongrande, J. Vayreda, and J. Terradas, "Woody plant richness and ndvi response to drought events in catalonian (northeastern spain) forests," *Ecology*, vol. 88, no. 9, pp. 2270–2279, 2007. [Online]. Available: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/06-1195.1

[15] H. Wang, Z. Li, L. Cao, R. Feng, and Y. Pan, "Response of ndvi of natural vegetation to climate changes and drought in china," *Land*, vol. 10, no. 9, 2021. [Online]. Available: https://www.mdpi.com/2073-445X/10/9/966

[16] D. Golon, "The land processes distributed active archive center (lp daac)," U.S. Geological Survey, Reston, VA, Tech. Rep. 2016-3070, 2016, contributing office: Earth Resources Observation and Science (EROS) Center. [Online]. Available: http://dx.doi.org/10.3133/fs20163070

[17] R. Rothermel, *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*, ser. USDA Forest Service research paper INT. Intermountain Forest & Range Experiment Station, Forest Service, U.S. Department of Agriculture, 1972. [Online]. Available: https://books.google.com/books?id=27n_RugVVK0C

[18] J. H. Scott and R. E. Burgan, "Standard fire behavior fuel models: A comprehensive set for use with rothermel's surface fire spread model," United States Department of Agriculture; Forest Service, Rocky Mountain Research Station 240 West Prospect Road Fort Collins, CO 80526, General Technical RMRS-GTR-153, June 2005, last updated: 2022.

[19] T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi *et al.*, "Kerastuner," https://github.com/keras-team/keras-tuner, 2019.

[20] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," 2018. [Online]. Available: https://arxiv.org/abs/1603.06560

[21] I. Dewancker, M. McCourt, and S. Clark, "Bayesian optimization for machine learning : A practical guidebook," 2016.

[22] Copernicus Climate Change Service, "Climate Data Store, (2018): Seasonal forecast daily and subdaily data on single levels. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)." DOI: 10.24381/cds.181d637e (Accessed on 11-07-2023), 2018.

[23] A. Akintomide Afolayan, J. Chunyong, W. Jiali, and K. Veerabhadra Rao, "Evaluation of precipitation across the contiguous

united states, alaska, and puerto rico in multi-decadal convection-permitting simulations," *Scientific Reports*, vol. 14, no. 1, p. 1238, 2024. [Online]. Available: https://doi.org/10.1038/s41598-024-51714-3