EXPLORING MODELS TO IMPROVE WILDFIRE HOTSPOT PREDICTION

Abstract—**Wildfires are resposible for environmental, economic, and human losses, highlighting the need for accurate and timely prediction models. In order to predict the wildfire hotspots, this study show's conventional Long Short-Term Memory (LSTM) models with improved architectures, like LSTM with Attention layers and Time Series Transformers. Using NASA's MODIS satellite information from 2000 to 2023, important prediction features such as fire radiative power and brightness temperatures were studied. The Results shows that the Time Series Transformers substantially outdated traditional LSTM models, which achieved least RMSE among the testing between the period (2021–2023).In particular, the Transformer error has decreased. RMSE by 11.63% in 2021, 20.29% in 2022, and 4.82% in 2023 when compared to the foundation LSTM model. The results from these illustrate the signficance of improved architectures to record complecated temporal trends and to enhance stability of wildfire hotspot prediction, facilitates the earlier assistance and best disaster management.**

Index Terms—**Wildfire prediction, LSTM, Transformers, Attention models, MODIS data**

I. Introduction

Wildfires are the most horrible natural disasters, which lead to significant environmental liability, greenhouse gas emissions are the significant reason for loss of biodiversity and are responsible for climate change. In the affected areas by wildfires are reasons for the economic and human tolls which are served equally severely, threatening livelihood, properties, and infrastructure in affected areas. The increase in global temperatures and raised human activities from the past decades have led to an increase in the intensity of wildfires, this makes the prediction a more essential tool for useful disaster management.

The significant challenge due to the complexity caused by temporal patterns in environmental data will remain accurate prediction of wildfire hotspots. The conventional models which are like Long Short-Term Memory (LSTM) networks, which will show effective on degree, will be hard to integrate long-term spatial variations and dependencies in resulting data. This will affect precision and dependability in crucial actual situations, as there will be an increase in data complexity.

- The use of advanced deep learning models will help in improving the prediction of wildlife hotspots.
- To determine how well the performance of Time Series Transformers and LSTM with Attention Layers is in comparison to traditional LSTM models.
- To find if advanced architecture can improve the precision and dependability of predictions by testing null and alternate hypotheses.

This study has NASA's MODIS satellite data (2000–2023) to show the effectiveness of distinct time-series prediction models. The better complex temporal patterns can be captured more effectively by integrating attention mechanisms and transformer architectures. The effective disaster management strategies provide actionable insights by integrating advanced models which improve prediction performance research.

II. Literature Review

A. Base Paper Summary:

The study by Kadir [1] investigates the application of the Long Short-Term Memory (LSTM) deep learning algorithm for forecasting and tracking wildfire hotspots in Indonesia. This study discusses the growing frequency of forest fires brought on by warming temperatures, man-made activities, and the distinct environmental conditions of tropical areas like Indonesia. The research highlights the following key features:

There is severe and frequent wildfire occurrence in Indonesia, particularly in areas like Sumatra and Kalimantan, where peatland contributes to the spread of fires. There is significant ecological damage, air pollution, and health risks, such as respiratory issues due to wildfires. Additionally, these fires lead to carbon emissions, exacerbating global warming. Taking preventive measures requires an efficient forecasting method to assist policymakers.

The dataset includes over 700,000 records of wildfire hotspots derived from NASA's MODIS system and covers the years 2010 to 2022. Preprocessing of data was performed by applying different techniques: filtering, aggregation, and normalization, to enhance forecasting precision. Data were then grouped into classes of multiple parameters like acquisition date, brightness, and geographic coordinates in latitude and longitude at varying confidence levels.

The LSTM algorithm was chosen because it could model sequential data and capture long-term dependencies effectively.

The study divided the dataset into training, 80%, and testing, 20%, subsets and applied error evaluation metrics such as RMSE, MAE, and R^2 to check the accuracy of the forecasting. The research was focused on the prediction of the number of fire hotspots for future years and mapping their spatial distribution.

The LSTM model was reliable, with an average error of 7% in the forecast of the number of wildfire hotspots. Analysis revealed that hotspots due to wildfires are concentrated mainly in Sumatra and Kalimantan. Monthly patterns showed an increase in the wildfire incidents toward the dry season (September–December). Developed a predictive model for wildfire hotspot mapping and forecasting in tropical regions. Demonstrated the application of LSTM for sequential data analysis on hotspot forecasting with high accuracy. Highlighted the need to integrate the wildfire forecast into environmental monitoring systems.

The application of MODIS data in moderate resolution (250m) can only allow the detection of small-scale fires. External factors, such as unmeasured meteorological variables, might influence the result of prediction. The authors propose improving the LSTM model by adding more features and using higher-resolution data. Recommendations for expanding the study to include other regions with similar environmental conditions.

B. Related Work

Time series forecasting is crucial in fields like finance, transportation, and environmental sciences. Traditional models, such as ARIMA, often struggle with non-stationary data, while deep learning methods like LSTM have shown significant promise in capturing long-term dependencies.

Wen and Li [2] suggested an enhancement by including an attention mechanism together with LSTMs to improve prediction performance. The attention layer dynamically weights the importance of previous inputs and resolves several drawbacks of encoder-decoder models, such as those concerning input sequences of big length. After experimentation using air quality, electricity consumption, and stock datasets, it was found that this model performed better than conventional methods based on LSTMs, especially for larger time steps.

This work identifies that the integration of LSTMs with attention mechanisms in enhancing time series prediction helps establish a benchmark for state-of-the-art models dealing with complex temporal patterns [2].

Though originally developed for NLP, transformers have lately shown great promise for time-series analysis. Ahmed [3] performed an extensive tutorial on the very basics of Transformer architecture: self-attention, positional encoding, and multi-head mechanisms, with the rationale for using them in the development of time-series forecasting and classification tasks. They also reviewed some limitations of RNNs, such as the problem of vanishing gradients and the inefficiency of sequential processing.

The tutorial investigated various enhancements to the original Transformer architecture, such as ProbSparse attention, temporal encodings, and gating mechanisms, which optimize the performance of long-sequence time-series tasks. Ahmed et al. also discussed best practices on how to train Transformers, such as adaptive model initialization and optimization techniques for large datasets. Their review consolidated the transformative role of Transformers in tackling temporal dependencies and achieving state-of-the-art results in a wide variety of time-series applications.

Using factors from district to village-level and matching it with multiple time scale analyses, Edwards [4] assessed anthropogenic drivers of forest fires in Indonesia. He noticed that there were definite incidences of El Niño events which are attached with enhanced drought, greatly determined the fire activity in more than 50 percent year-to-year variation of hotspots. The authors also presented evidence on the role of political decentralization, as the creation of new districts was associated with up to 60% more fires, and they found a strong correlation between fire activity and regional economic growth, underscoring the fact that rural development is typically related to higher land-clearing fires. At the village scale, the critical determinants of fire prevalence were poverty, remoteness, and dependence on traditional fire-based agriculture. The study underlines a need for targeted interventions at district and village scales to reduce fire risks, especially during El Niño events.

Cheng [5] developed a parameter-flexible wildland fire forecasting algorithm by using machine learning and ROM techniques to overcome computational difficulties related to traditionally used models for fire predictions. In an efficient manner, the forward modeling approach was derived using the ML-based surrogates for the forecast of burnt areas, including CAE and PCA. It also proposed, for the first time, the inverse modeling framework for estimation of parameters in an innovative way using latent assimilation techniques. It provided a computationally robust approach for large-scale wildfire events, with satellite observations for validation. Results highlighted the need to reduce computational costs and adaptability in dynamic wildland fire modeling to provide near real-time fire propagation forecasting for emergency response.

The work by Bhowmik [6] introduces a novelty in the form of a spatiotemporal machine learning framework for wildfire prediction, focused on California's recent intense wildfire seasons. The authors developed a holistic database of wildfires, the CALMZ dataset, that integrated historic fire datasets, environmental and meteorological sensor values, and geological factors at an astonishing 37 million data points. Using this dataset, they came up with the ULSTM network, a deep hybrid model of U-Net and LSTM, especially in learning spatial and temporal features. ULSTM achieved over 97% accuracy in wildfire predictions, outperforming traditional CNNs (76%). Leading indicators, like temperature and wind speed, proved to be more predictive than trailing indicators, such as particulate matter levels. It is able to predict 85.7% of the fires surpassing 300,000 acres well in advance of up to two weeks.It also points out that ULSTM has a great potential contribution to large-scale wildfire management, resource allocation, and early warning systems, though there are limitations with regards to the smaller, human-caused fires. Future directions include the expansion of the model scope globally and the development of personalized early warnings for at-risk populations. This system holds promise for mitigating the social, environmental, and economic impacts of wildfires.

Sayad et al. [7], on the other hand, have developed a model in order to predict a wildfire with the data set that was provided using satellite-based remote sensing. They had used MODIS products with parameters like NDVI, LST, Thermal Anomalies amongst others on areas in Canada. Later, they applied ANNs and SVMs on Databricks' platform after the respective preprocessing and extrapolation of those datasets in order to get the desired accuracies of 98.32% and 97.48% correspondingly. The comparative analysis demonstrated the superiority of their methodology when compared to current models in terms of predictability. Future work will focus on refining the model by the inclusion of meteorological variables such as wind speed and air temperature.

Most of the current studies on wildfire prediction have been focused on the use of the SVM and LSTM models, rarely exploring the attention mechanisms or transformer models that are more capable of capturing spatiotemporal dependencies. Most of the related works evaluate the models based on a rather smaller dataset, missing recent-year insights-for example, from 2021 to 2023-that are important in considering the model's resilience in changing fire patterns.

We integrate advanced architectures like attention mechanisms and transformers to improve spatiotemporal predictions. Further, we evaluate our models on extended datasets (2000-2023), with more focus on recent years to make sure robustness and adaptability are considered in the climate-driven wildfire trends.

III. Methodology

A. Dataset Overview

This research utilizes a dataset derived from NASA's MODIS satellite system, which offers extensive fire detection data for Indonesia spanning from 2000 to 2023. The dataset provides critical spatial and temporal information essential for wildfire prediction, including latitude and longitude coordinates, brightness temperature, fire radiative power (FRP), confidence levels, and a day/night indicator. A comprehensive total of 1,492,022 records were compiled by merging annual CSV files into a unified dataset employing Python. Key features relevant to wildfire prediction were retained, while duplicates and irrelevant data were removed to ensure the dataset's quality and reliability. This dataset constitutes the basis for the analysis, supplying robust data for training and assessing predictive models.

B. Data Preprocessing

The data preprocessing phase encompassed several critical steps to ensure the dataset was adequately prepared for modeling. Initially, the dataset underwent a comprehensive analysis to identify any missing values or duplicate entries, both of which were confirmed to be absent. This enabled a seamless progression to feature engineering, where temporal attributes such as year, month, and day were extracted from the acquisition date. Subsequently, the daily fire hotspot counts were consolidated to create a time-series dataset. To effectively scale the features for modeling, a MinMaxScaler was employed to normalize the data within a range of [0, 1], which is vital for optimizing the training of neural networks. Outliers were examined to evaluate their impact on model performance, and any notable discrepancies were appropriately addressed during the preprocessing phase. Ultimately, the dataset was segmented into a training set (2010–2020) and testing sets (2021–2023) to facilitate performance assessment across various time frames.

C. Model Descriptions

1) LSTM Model:: The Long Short-Term Memory (LSTM) model was selected as the primary methodology due to its exceptional capability to understand temporal dependencies. This architecture features two stacked LSTM layers, each comprising 24 units, supplemented by dropout layers to reduce the likelihood of overfitting. Furthermore, dense layers are incorporated for final output predictions.

- Forget Gate The forget gate determines which information from the previous cell state (C_{t-1}) should be discarded:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

  where f_t is the forget gate output, x_t is the input at time t, h_{t-1} is the hidden state from the previous time step, W_f and b_f are the weight matrix and bias for the forget gate, and σ is the sigmoid activation function.
- Input Gate The input gate decides which new information will be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

  Here, i_t is the input gate output, $\tilde{C}_t$ is the candidate cell state, and W_i, W_C and b_i, b_C are the corresponding weights and biases. The function $\tanh$ is the hyperbolic tangent activation function.
- Cell State Update The cell state is updated by combining the forget gate and input gate:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (4)$$

  where C_t is the updated cell state and C_{t-1} is the previous cell state.
- Output Gate The output gate determines the information to output as the hidden state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (6)$$

  Here, o_t is the output gate output, h_t is the current hidden state, and W_o and b_o are the weight matrix and bias for the output gate.

Summary The LSTM model combines the above equations to regulate information flow through the network:

- The **forget gate** removes irrelevant information from the previous state.
- The **input gate** decides what new information to add to the cell state.
- The **cell state** stores long-term information.
- The **output gate** determines the relevant information to output.

Key hyperparameters, such as a dropout rate of 0.26, a learning rate of 0.0042, and a time step of 10, were optimized using Optuna. The training phase extended over 44 epochs with a batch size of 16. The baseline model achieved RMSE metrics of 55.64 for 2021, 43.63 for 2022, and 141.57 for 2023. These results established a foundation for future improvements.

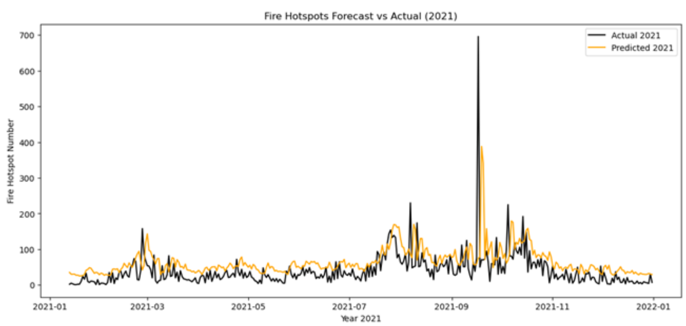

Fig. 1. Base LSTM Model: Highlights the model's performance in capturing fire hotspot trends in 2021. Some deviations between predicted and actual values may indicate underperformance in certain regions or temporal patterns.

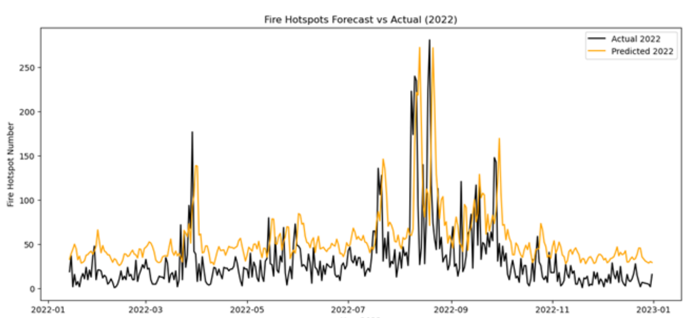

Fig. 2. Base LSTM Model: Shows an improvement in prediction accuracy in 2022 compared to 2021, with a closer alignment between predicted and actual values.

2) LSTM with Attention Layers:: Building on the established foundation, the LSTM model was enhanced with an attention mechanism to improve its capacity for identifying temporal relationships and emphasizing significant time intervals. This attention mechanism dynamically modifies the importance of preceding time intervals during the prediction process, allowing the model to concentrate on relevant data points. The refined architecture comprises two LSTM layers, each containing 88 units, with an integrated attention block situated between them.

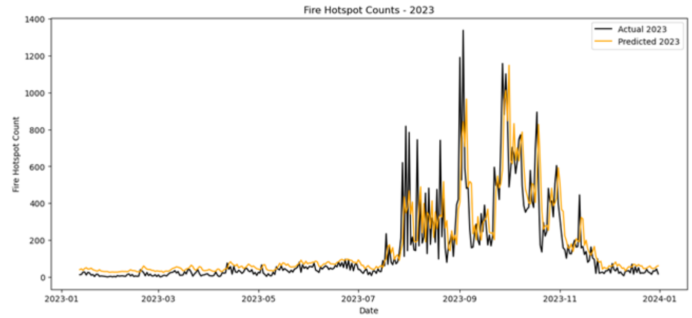

Fig. 3. Base LSTM Model: Indicates a potential drop in accuracy, likely due to the model's inability to generalize well to more complex patterns or significant variations in the dataset in 2023

Attention Mechanism: The attention mechanism computes the context vector c_t based on the importance of hidden states h_t. The following equations define the steps:

1) Score Calculation:

$$e_t = \tanh(W_q h_t + W_k h_{t'}), \tag{7}$$

where W_q and W_k are trainable weight matrices, and e_t represents the alignment score between the query and key.

2) Attention Weights:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^{T} \exp(e_t)}, \tag{8}$$

where α_t is the normalized attention weight for each time step.

3) Context Vector:

$$c_t = \sum_{t=1}^{T} \alpha_t h_t, \tag{9}$$

where c_t is the weighted sum of the hidden states, capturing relevant information from the sequence.

4) Final Output:

$$y_t = W_y[c_t; h_t] + b_y, \tag{10}$$

where W_y and b_y are trainable weights and biases, and $[c_t; h_t]$ denotes the concatenation of the context vector and hidden state.

To reduce the likelihood of overfitting, dropout layers with a rate of 0.23 were included. The learning rate was precisely adjusted to 0.000045, while the model was trained for 33 epochs using a batch size of 16 and a time step of 8. The incorporation of the attention mechanism substantially improved performance, yielding RMSE values of 54.26 for 2021, 42.78 for 2022, and 147.89 for 2023, highlighting its efficacy in capturing complex temporal dependencies.

3) Time Series Transformer Model:: This study explores advanced modeling methodologies through the development of a Time Series Transformer, which employs an encoder-decoder architecture integrated with additive attention mechanisms. The model leverages input embeddings to convert raw data into a high-dimensional space, subsequently utilizing transformer encoders to effectively capture long-term dependencies.

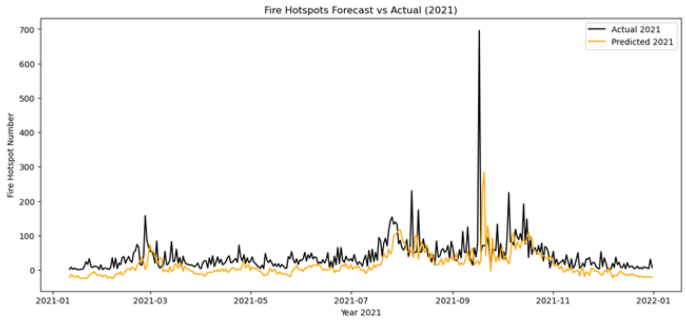

Fig. 4. LSTM With Attention: (2021): Demonstrates a more accurate prediction compared to the Base LSTM, with fewer significant outliers.

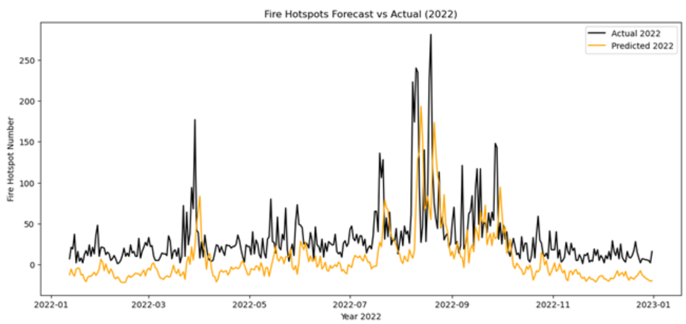

Fig. 5. LSTM with Attention Model: (2022): Depicts a strong correlation between predicted and actual values, reflecting the model's ability to adapt to patterns in the data.

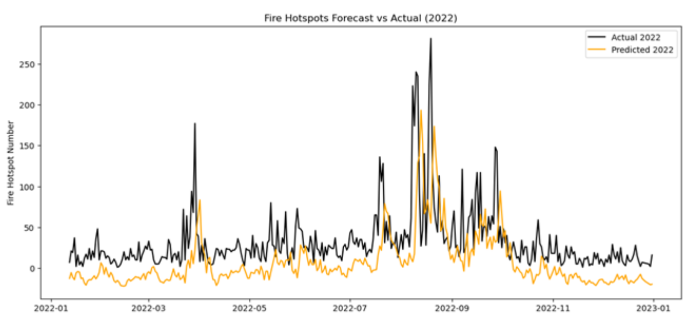

Fig. 6. LSTM with Attention Model(2023): Shows better generalization compared to the Base LSTM, though some discrepancies persist in areas of high variability.

IV. Transformer Equations and Implementation Details

The transformer relies on the self-attention mechanism to model dependencies in the sequence. The following equations define its core components:

A. Input Embedding

$$x_i^{\text{embed}} = W_e x_i, \tag{11}$$

where W_e is the embedding matrix used to map the input x_i into a high-dimensional space.

B. Scaled Dot-Product Attention

$$A = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{12}$$

where $Q = XW_Q$, $K = XW_K$, $V = XW_V$, and d_k is the dimensionality of the keys. Here, A represents the attention output.

C. Multi-Head Attention

$$\text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W_O, \tag{13}$$

where each attention head learns distinct relationships, and W_O is the output projection matrix.

D. Feed-Forward Network

$$FFN(Z) = \text{ReLU}(ZW_1 + b_1)W_2 + b_2, \tag{14}$$

where W_1, W_2 and b_1, b_2 are trainable parameters, and Z is the input to the feed-forward layer.

E. Encoder Output

$$Z = \text{MHA}(Q, K, V) + X, \tag{15}$$

where X is the input sequence.

F. Final Prediction

$$y_t = W_o Z + b_o, \tag{16}$$

where W_o and b_o are trainable weights and biases used for regression or classification tasks.

The incorporation of additive attention enhances the model's focus on critical features. The architecture consists of three attention heads, a single encoder layer, and three decoder layers. To mitigate the risk of overfitting, dropout regularization was set at 0.23. The training process was conducted with a learning rate of 0.00013, a batch size of 32, and a time step of 2. The Time Series Transformer outperformed LSTM-based models, achieving RMSE values of 49.18 for 2021, 34.78 for 2022, and 144.25 for 2023, highlighting its exceptional capacity to model complex temporal dynamics.

The experimental setup was defined by comprehensive preparation and rigorous testing to ensure the dependability of results. We employed GPU-accelerated hardware for the training and evaluation of the models, thereby shortening computation times. Our model development utilized Python-based frameworks, including TensorFlow and PyTorch, supported by libraries such as NumPy, Pandas, and Matplotlib for data processing and visualization tasks. The Root Mean Squared Error (RMSE) was selected as the primary evaluation metric to gauge the models' predictive accuracy. Additionally, hyperparameter tuning was carried out using Optuna to refine crucial parameters, such as the number of LSTM units, dropout rates, learning rates, and attention configurations. To validate

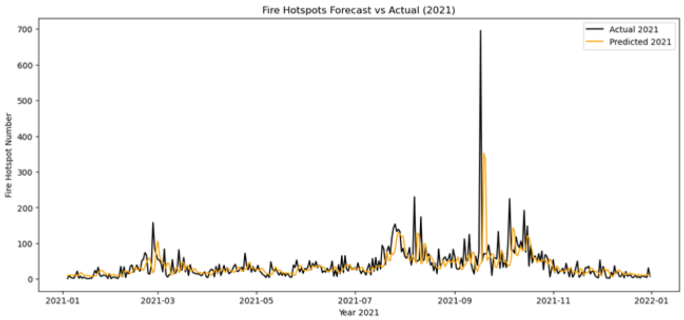

Fig. 7. Time Series Transformer: (2021): Indicates a high degree of accuracy, with predictions closely matching actual values. The Transformer model demonstrates its capability to handle time-series data efficiently.

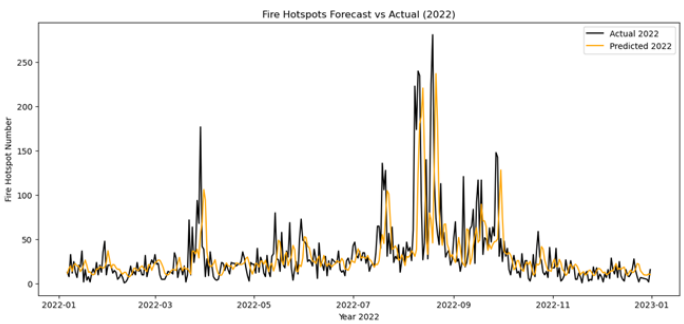

Fig. 8. Time Series Transformer:(2022): Further improves predictive accuracy, showing minimal deviations and closely following the trend of actual fire hotspots.

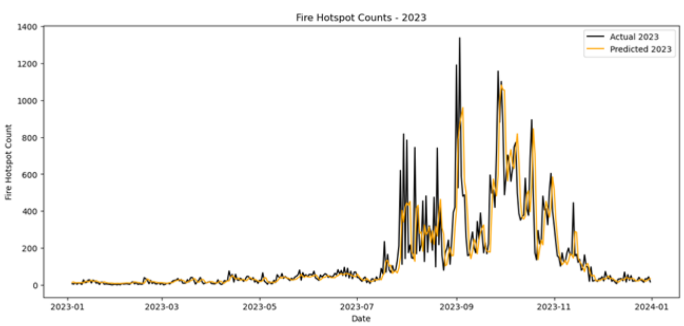

Fig. 9. Time Series Transformer (2023): Maintains strong performance even in a more challenging year, with predictions staying consistent and capturing significant variations in fire activity.

the models and mitigate the risk of overfitting, 20% of the training data was allocated for validation during the training phase. In the testing phase, data from 2021 to 2023 was employed to assess the models' generalization capabilities.

V. Results

The findings of this research have shown that the Time Series Transformer dominated the LSTM based models throughout every test period across the RMSE metric. Owing to state of the art attention and transformer architectures, it was possible for the model to learn complex temporal knowledge more efficiently which made the model suitable for predicting

Model	RMSE(2021)	RMSE(2022)	RMSE(2023)
Base LSTM	55.64	43.63	151.57
LSTM with Attention	54.26	42.77	147.89
Time Series Transformer	49.18	34.78	144.25

Fig. 10. Root Mean Square Error (RMSE) comparison across different models (Base LSTM, LSTM with Attention, and Time Series Transformer) over three years (2021, 2022, and 2023). Lower RMSE values indicate better predictive performance.

wild fire hotspots. The suggested technique is able to provide a balanced view towards all the phases' data preparation and model testing as well as model performance comparison, indicating how more efficient strategies for polar fire should be designed and used.

The testing models' performances were measured and recorded based on a Root Mean Squared Error (RMSE) performance metric for test datasets ranging from year 2021 to 2023. The base LSTM network model was able to give reasonable results as it recorded an overall RMSE of 55.64 for the year 2021, 43.63 for 2022 and 141.57 for the year 2023. Furthermore, placing attention layers to model this temporal architecture making use of the LSTM technique further reduced its dependence on the time variable. The 2021 figure recorded an RMSE for the year of 54.26, 42.78 for the year 2022 and RMSE for year 2023 of 147.89. Regardless of this, the Time Series Transformer model consistently performed better than both LSTM models in every test period. The values were 49.18, 34.78 and 144.25 respectively for years 2021, 2022 and 2023.

This shows superior ability displayed by the transformer-based architectures for capturing the complex and temporal relationships along with the dependencies in task of wildfire prediction.

The comparative performance of the models is summarized in Figure 1, which highlights the improvements achieved by the advanced architectures.

In order to obtain more detailed understanding into how the model was performing, several visualizations were created for analyzing the accuracy and loss, trends in prediction, and actual fire hotspots. Here are a few graphs and figures created and presented in the document:

1) Accuracy and Loss Trends: : The authors mention that training and validation accuracy/loss curves are populated with the values for each of the models which demonstrates the convergence and overfit/underfit patterns seen after training. The LSTM model on its own converged gradually over time while attention based enhancement of LSTM had some edge in terms of validation loss. Time Series Transformer model on the other hand had a speedy convergence with very low overfitting as such demonstrating its effectiveness.

2) Temporal Pattern Visualization: : Temporal accuracy of each model is assessed by visually comparing the number of fire hotspots predicted and the number in the actual places for the years 2021, 2022 and 2023. Such graphs show that Time Series Transformer when compared temporally with other fire events accurately predicts them at fire peak periods rather than from LSTM based models which can only do so at other periods.

3) Model Performance Trends: : Values of RMSE are given with the years tested at the top for each of the models to show the relative gains made through attention mechanism and transformer architectures.

A. Insights

The results of these studies are highlighting the superior performance of the Time Series Transformer model when compared to the LSTM-based models. This can be explained by the specific features of transformer architectures, which handle time series data the best in terms of complex temporal factors and relationships. In contrast to LSTMs that proceed with the data one step at a time, making them susceptible to the vanishing gradient problem for long sequences, transformers utilize attention layers that permit the model to attend to any time index in the input sequence at one shot. Such a capability to capture short-term and long-term dependencies at the same time creates an ideal environment for utilizing transformers as there is a significant presence of short and long variations in most large datasets, for instance, wildfire hotspot locations. It is worth noting that, despite the popularity of the Time Series Transformer, several problems were faced during the realisation of the project. The first problem is definitely the dataset's scope and richness, which was challenging due to its temporal dependencies, geographical variance, and detection hot spots. There was also the aspect of data preparation to eliminate inconsistencies, standardise values, and filling in the blanks. Moreover, the process of training the transformer model consisted of tuning parameters such as the number of heads, layers and embedding depth which were all resource and time intensive. The LSTM combined with attention layers, though provided a better performance when compared to baseline LSTM, still struggled to reach.

B. Practical Implications

There are significant implications of this study's findings on wildfire management and environmental monitoring. The accurate predictions of the wildfire hotspots will enable the early intervention of the authorities, by allowing the timely allocation of resources likefirefighting teams, equipment etc. This can help fight and stop the spreading of wildfires, by reducing the environmental damage, and also safeguarding the communities in all vulnerable regions.

The Time Series Transformer has the ability to capture the complex patterns in the wildfire activity. This also supports the better decision-making in long-term planning and policy formulation. For example, at any particular instance, if we can identify high-risk periods or regions in time, we can facilitate the implementation of all the preventative measures, like controlled burns or enhanced surveillance during peak fire seasons.

If we integrate the advanced predictive models into existing wildfire monitoring systems, like those that are powered by the satellite data, it can strengthen the disaster management frameworks while improving the overall resilience to climate-related challenges. By using the cutting-edge machine learning techniques, the study helps in understanding that more effective and proactive wildfire management strategies, ultimately contributing to environmental sustainability and public safety might give better results.

VI. Conclusion and Future Work

This has been discussed in the research on the effectiveness of advanced models in machine learning, especially about LSTM, LSTM models with attention layers, and Time Series Transformers, in hotspot predictions for wildfires in Indonesia using the dataset from NASA MODIS. It was found that Time Series Transformer appears to be the best choice when looking at accuracy and robustness regarding the other models in study since it was able to catch difficult temporal dependencies and hidden relationships in time series data. By using attention mechanisms and transformer-based architectures, improvements have been made on predictions of wildfire trends, truthfully preparing the predictions for reliable use. Findings thus stress the need for integrating into wildfire monitoring systems such sophisticated machine-learning techniques for timely response and allocation of appropriate resources.

A. Limitations

While the study achieved promising results, there are several limitations we noticed that need to be addressed. First is the computational cost of training the transformer models. It was significantly higher when compared to the LSTM-based models. This might limit the deployment of the model in environments which are resource-constrained. Additionally, the quality and the resolution of the MODIS dataset posed challenges, like handling the missing data, variability in the confidence levels, and geographic biases. All this constraints might have impacted the models' generalization through out the diverse regions or also extreme wildfire scenarios. Lastly, the models are trained and tested while historical data, that might not be fully capturing the dynamic, evolving nature of the wildfire behavior being under the changing climatic conditions.

B. Future Directions

In future one can focus on overcoming these observed limitations. They can do explore some real-time prediction capabilities while using the streaming data. This will enable some updates iusthat are dynamic and forecasting that are near-instantaneous. By expanding this geographical scope of the study for including the other regions that are wildfire-prone we can validate models' generalizability. We can also provide insights on patterns of global wildfire. Along with this, incorporating datasets of higher-resolution or by integrating the data from two to many satellite sources might enhance the predictions quality. One of the other promising directions might include the process of optimizing the transformer's architectures to reduce the computational costs, exploring the hybrid models which combine the strengths of transformer models and the other deep learning techniques. Advancements like this will make the role of AI-driven wildfire prediction systems in more solid mitigating the drastic impacts of wildfires worldwide.

REFERENCES

[1] E. A. Kadir, T. Suhartono, E. Supriyono, and H. P. Wijaya, ”Wildfire Hotspots Forecasting and Mapping for Environmental Monitoring Based on the Long Short-Term Memory Networks Deep Learning Algorithm,” Environments, vol. 10, no. 124, pp. 1–26, Jul. 2023. [Online]. Available: https://doi.org/10.3390/environments10070124.

[2] X. Wen and W. Li, ”Time series prediction based on LSTM-attention-LSTM model,” IEEE Access, vol. 11, pp. 48322–48331, 2023, doi: 10.1109/ACCESS.2023.3276628.

[3] S. Ahmed, I. E. Nielsen, A. Tripathi, S. Siddiqui, R. P. Ramachandran, and G. Rasool, ”Transformers in time-series analysis: A tutorial,” Circuits, Systems, and Signal Processing, vol. 42, no. 12, pp. 7433–7466, 2023.

[4] R. B. Edwards, R. L. Naylor, M. M. Higgins, and W. P. Falcon, ”Causes of Indonesia’s forest fires,” World Development, vol. 127, p. 104717, 2020.

[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, p. 3228, 2022.

[6] R. T. Bhowmik, Y. S. Jung, J. A. Aguilera, M. Prunicki, and K. Nadeau, ”A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework,” Journal of Environmental Management, vol. 341, p. 117908, 2023.

[7] Y. O. Sayad, H. Mousannif, and H. Al Moatassime, ”Predictive modeling of wildfires: A new dataset and machine learning approach,” Fire Safety Journal, vol. 104, pp. 130–146, 2019.

[8] P. K. Gellert, ”A brief history and analysis of Indonesia’s forest fire crisis,” Indonesia, no. 65, pp. 63–85, 1998.

[9] J. Lever and R. Arcucci, ”Sentimental wildfire: A social-physics machine learning model for wildfire nowcasting,” Journal of Computational Social Science, vol. 5, no. 2, pp. 1427–1465, 2022.

[10] S. Girtsou, A. Apostolakis, G. Giannopoulos, and C. Kontoes, ”A machine learning methodology for next day wildfire prediction,” in 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Jul. 2021, pp. 8487–8490.