A ChatGPT-like Solution for Power Transformer Condition Monitoring

Joseph Accurso, Raul Mendy, Aira Torres, and Yufei Tang
Department of Electrical Engineering and Computer Science
FPL Center for Intelligent Energy Technologies
Florida Atlantic University
Boca Raton, FL
Email: accu3879@ravens.benedictine.edu, {rmendy2019, airatorres2020, tangy}@fau.edu

***Abstract*—This paper explores the application of large-scale foundation models (LSF-Models) for the improved prognostic management of smart grid power transformers. One of the foundational architectures of many LSF-Models, including ChatGPT, is the Transformer architecture. This architecture helps LSF-Models perform complex classification predictions and provide users with an appropriately predicted response. The advanced natural language processing techniques of the Transformer architecture interpret the nuances of power transformer-related data and efficiently extract valuable insights for fault detection and diagnosis. The multiple self-attention mechanisms of the architecture provide an enhanced form of feature extraction which expedites the creation of comprehensive models for large multiclass datasets. During testing, this fault classification method performed with 97.2% accuracy according to the Matthews Correlation Coefficient when evaluating 45 classes of simulated power transformer internal faults and external transient disturbances. To better represent the diverse multimodal data that compose smart grid systems, this research explores a unified system that integrates multiple machine-learning models into one simple and easy-to-use interface. The backend of this interface again takes advantage of the Transformer architecture to perform conversation-based classification and to provide a response or prediction from the appropriate system-integrated model. The success of the Transformer architecture in diverse applications within the same overall system showcases its potential to analyze the wide range of data typically found throughout a robust smart grid system.**

***Index Terms*—Large-Scale Foundation Models, ChatGPT, Prognostic Health Management, Power Transformer, Multiclass Classification, Fault Detection**

I. Introduction

In 2022, the United States utilized 4.05 trillion kilowatt-hours of electricity to support residential, commercial, and industrial sectors [1]. With new emerging technologies connecting to the grid every day, the variety of energy consumption has put a diverse strain on the electrical grid system. The transmission of appropriate electrical energy from current distribution systems is only possible due to the proper maintenance of electrical power transformers. These devices step up or step down voltage according to particular physical properties that take advantage of electromagnetic induction. This process, however, dissipates energy to the environment in the form of heat, thereby exposing the power transformer to potential thermal malfunctions. The general wear and tear built up throughout the lifespan of a power transformer contributes to additional transient disturbances and mechanical failures. Not only does this compromise the mechanical design of the power transformer but it also affects the stability of its electrical output. Anomalous voltage drops, voltage rises, and phase shifts all impact the overall reliability of the electrical grid and can be potentially damaging to connected devices. Therefore, a power transformer operating with untreated mechanical faults becomes a weak link that, if left untreated, ultimately disrupts the entire chain of power distribution to industrial, commercial, and residential sectors.

The rise in electrical power dependence requires an even greater improvement in reliability. Due to the fact that internal faults and external transient disturbances also impact the electrical signal itself, modern sensors incorporated into smart grid systems can be used to collect signal data and alert maintenance crews to defective power transformers. Although contemporary systems have shown promising results in detecting faults [2], classifying those detections poses a greater computational challenge. The amount of data needed to construct a comprehensive fault classification model for an entire smart grid system is vast and requires machine learning (ML) models with the capability of handling extensive datasets. This is where large-scale foundation models (LSF-Models) such as GPT-3.5 (ChatGPT) [3] and Segment Anything Model (SAM) [4] have seen the greatest success. Applying an LSF model for smart grid system condition monitoring is a comprehensive solution that provides corroborated information by utilizing multiple models and simulations.

The success of many LSF-Models is attributed to the innovative Transformer deep learning architecture [5]. The Transformer architecture has become a popular foundation for many natural language processing (NLP) and computer vision (CV) models in the few years since its infancy in 2017. Its recent success in the form of ChatGPT proves its capability in handling large multiclass classification problems. This research proposes to use the Transformer architecture to classify three-phase signals in smart grids and to provide a similar conversational prediction to that of ChatGPT. Instead

This work was supported in part by the US National Science Foundation (NSF) through Grants CMMI-2145571 and CNS-1950400, and the Florida Power & Light (FPL) Center for Intelligent Energy Technologies (InETech) Summer Internship Program.

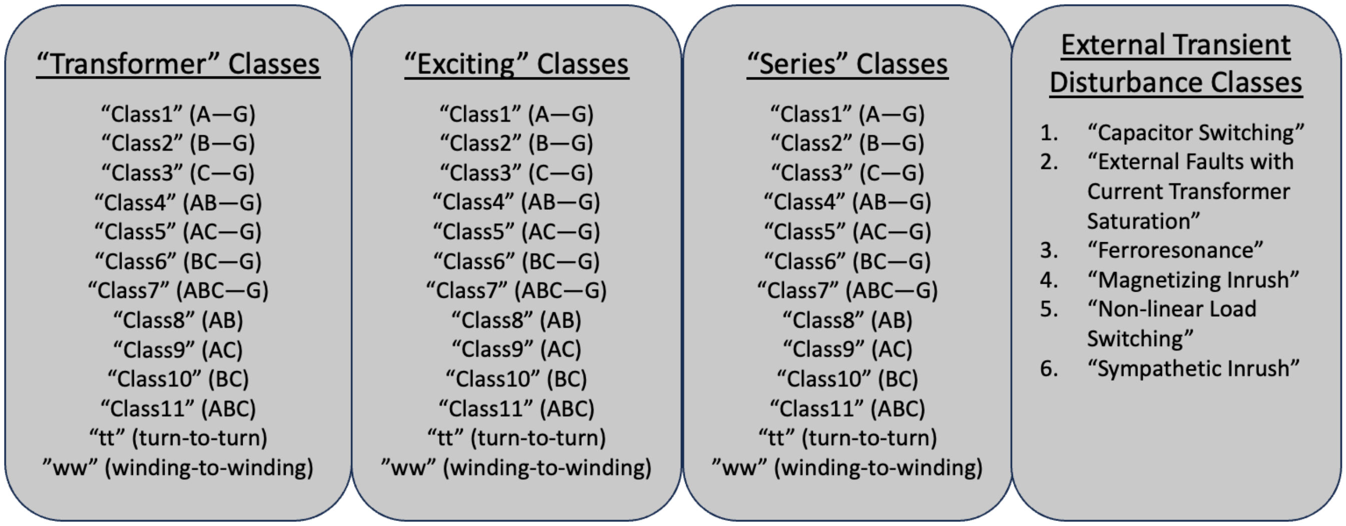


Fig. 1: Dataset used in this research. A total of 45 fault classes require a multiclass classification solution.

of processing global web data, the model was trained on anomalous time-series signals to provide a fault classification prediction. This multiclass fault classification model was integrated with a basic natural language processor which also used a simplified structure of the Transformer architecture to constitute the backend of the conversational user interface. The success of this unique flexibility acted as a proof-of-concept and demonstrated how future models can, in a similar way, deeply integrate diverse types of data throughout various power transformer condition monitoring applications.

The rest of this paper is organized as follows. Section II explains the background related to this research. Section III provides an overview of the proposed solution. Section IV details experimental results and demonstrates the proposed integrated system. Section V concludes the research and outlines future work.

II. Background and Preliminaries

A. LSF-Models for Smart Grids

Over the past few years, the popularity, efficiency, and universality of LSF-Models have all expanded rapidly. Foundation models attempt to encompass multiple facets of data within a particular network while remaining generally applicable to multiple tasks within the same sector. Models such as Bidirectional Encoder Representation from Transformers (BERT) [6], Enhanced Representation through kNowledge IntEgration (ERNIE) [7], and Large Language Model Meta AI (LLaMA) [8] have showcased unique and successful approaches to large-scale NLP applications. Additionally, the GPT series has pushed the boundaries of the Transformer architecture through the development of increasingly complex optimization structures [9] [10].

Despite the historic success of NLP and CV LSF-models, complex smart grid systems require flexible deep learning models that can appropriately interpret several modes of data, including signals, images, videos, and text. Even the textual mode requires the analysis of various data structures such as maintenance records, ongoing maintenance work orders, and culminating project reports. Holistic assistive systems, therefore, require a deep integration of multiple ML models. Current research into such a capable multimodal model is still years behind comparable designs from the NLP and CV fields [11]. The majority of proposed solutions include wavelet-based convolutional neural networks (CNNs) [12] [13] and recurrent neural networks (RNNs) [14] which both contain significant disadvantages within smart grid system applications.

B. Transformer Architectures for Fault Classification

The historical success of the Transformer architecture for large text-based multiclass classification for projects such as ChatGPT prompted the search for a custom architecture capable of accurate signal-based classification. At the beginning of 2023, electrical engineering researchers at the National Institute of Technology Calicut constructed two models based on the Transformer architecture. The research compared a basic "vanilla" Transformer to an innovative Differential Architecture Search (DARTS) algorithm which produced accurate results for detecting 23 fault types and 15 fault locations for power transformers [18]. The baseline model preprocesses the input data by equalizing the samples per class and converting the arrays to TensorFlow tensors. This preprocessed data is transformed into attention embeddings through several levels of multi-head attention mechanisms and feedforward networks. These methods analyze a series of 'queries', 'keys', and 'values' to calculate the relative connections between individual data tokens.

This process mirrors the structure of the foundational Transformer architecture encoder. Although proven to be a more efficient method for deep data analysis than RNNs, the mathematical processes of the attention encoder are not time-dependent. In order to retain time-dependent information, the model also applies positional embeddings to the data. This allows the context of the data to be reconstructed directly following the completion of the main attention encoder process.

C. Dataset Validation using Traditional ML Techniques

The initial steps of this research evaluated a subset of a three-phase time-series signal-based dataset [15] assembled by researchers at Syracuse University. This dataset recorded differential current as a function of time for a 5-bus system simulated using PSCAD/EMTDC software. 100,908 transient cases were simulated by modifying various system parameters. Differential current measurements for each transient example were saved in comma-delimited plain text files. Measurements were taken every 100 microseconds for 0.0726 seconds to produce 726 data points for each of the three phases. That is, the raw data was organized as a 726×4 matrix with the first column as time and columns two, three, and four as the differential current of phases A, B, and C, respectively.

Three types of transformers were observed in this simulation: power transformers, Indirect Symmetrical Phase Angle Regulator (ISPAR) series transformers, and ISPAR exciting transformers. 13 internal faults were observed for each of the three transformer types. The first 11 faults were labeled "Class1" through "Class11" and represented the following transformer winding locations: phase A to ground, B to ground, C to ground, A to B to ground, A to C to ground, B to C to ground, A to B to C to ground, phase A to phase B, A to C, B to C, and A to B to C. The twelfth and thirteenth classes for each transformer type reflected turn-to-turn and winding-to-winding fault locations, respectively. Six additional external transient disturbances were observed as individual fault classes: capacitor switching, external faults with current transformer (CT) saturation, ferroresonance, magnetizing inrush, non-linear load switching, and sympathetic inrush. The 13 internal faults for each of the three transformer types along with the six additional external transient disturbances composed the 45 total fault classes. An overview of the dataset is shown in Fig. 1.

The validity of this dataset was verified through its use as training data for traditional multiclass classification ML techniques. The overall workflow of these algorithms was inspired by previous analysis of various ML classifiers for signal classification using Discrete Wavelet Transform (DWT) decomposition for feature extraction [16]. DWT involves the convolution of a discrete signal and a preselected mother wavelet. The two signals are multiplied at increasing distances from the initial position of the discrete signal. Various families of mother wavelets can be applied in various individual patterns to find the most optimal comparison for a particular application and signal type. The convolution output passes through a series of high- and low-pass filters which produce waveforms of detail and approximation coefficients, respectively. The waveforms of these coefficients represent the high- and low-frequency components of the original signal as a function of time. The abnormal frequencies present in signal anomalies are exaggerated in these decomposed forms and can therefore be more easily detected and categorized during traditional signal analysis processes.

This research compared the features of each decomposed detail coefficient level and the last approximate coefficient level with the corresponding levels of other signals. Twelve statistical features including entropy, 5th percentile, 25th percentile, 75th percentile, 95th percentile, median, mean, standard deviation, variance, and root mean square were calculated for each decomposition level and phase. For example, a level-five decomposition analysis would evaluate twelve statistical features for six sets of DWT decomposition coefficients for three phases of data. 216 features, in this example, would be extracted per data file in the randomly selected training set. For all training instances, the number of decomposition levels was logistically related to model accuracy while maintaining a more linear relationship to the number of features and overall training time. Therefore, due to the diminishing returns of the asymptotic logistic relationship, only five decomposition levels were used for the ML-based dataset validation.

III. Problem Formulation and Proposed Solution

A. Research Objectives

Many energy companies utilize reactive and scheduled maintenance techniques to address damaged infrastructure and prevent abnormal transmission behaviors. Both of these strategies typically deploy maintenance personnel into the field in order to perform on-site system diagnostics. Scheduled maintenance is a broad procedure that requires a large sample of transmission infrastructure to be tested and observed for damage or anomalous behavior. Sending crews into the field is an expensive operation, and the repetitive examination of such a vast system is neither cost- nor time-efficient.

Reactive maintenance contains similar drawbacks in that it also requires expensive and repetitious deployment of on-site maintenance crews. Moreover, this technique relies upon accurate fault detection sensors and an internal system that can quickly communicate the significance of such signals to reactive repair teams. This framework often introduces significant delays in the process of system-wide rehabilitation as the malfunction may persist until on-site assessment and subsequent repair attempts are completed.

This complex situation highlights the need for smart grid systems to adopt more proactive strategies such as those studied in the field of prognostics and health management (PHM). Transitioning towards technologically advanced solutions (such as self-monitoring diagnostic systems for remote power transformer health assessment) would significantly mitigate several of the aforementioned challenges. By leveraging real-time data analytics and predictive algorithms, companies could swiftly identify irregularities or potential malfunctions based on pre-established sensor data within the smart grid systems. This strategy would not only work to proactively address signal irregularities but would also intelligently ap-

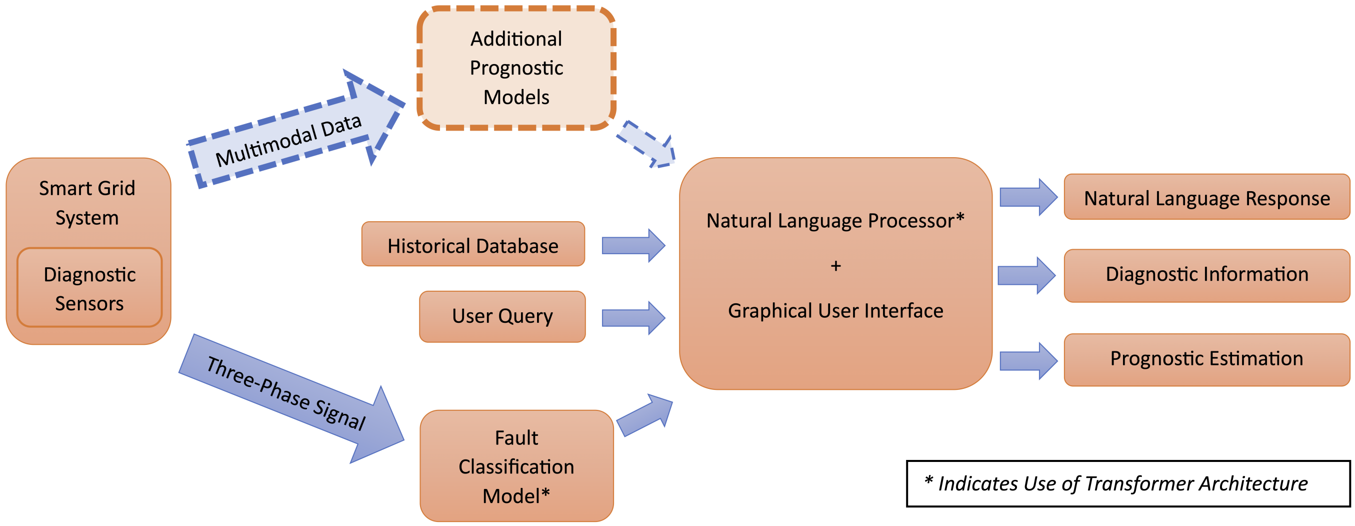


Fig. 2: Two intuitive Transformer-based models are implemented throughout the overall prognostic system and ultimately converge into a single conversational user interface. The scope of work reflects both actualized and proof-of-concept features.

ply holistic network-based insights to individual maintenance response situations.

B. Proposed Solution

This research aimed to establish proactive maintenance strategies by developing a ChatGPT-like interface capable of quickly and intelligently communicating power system insights to a user, shown in Fig. 2. The initial approach leaned on conventional machine-learning techniques, encompassing the use of random forest, gradient boost, and support vector classification algorithms. In order to efficiently observe abnormalities in three-phase signals, the DWT decomposition feature extraction techniques mentioned in Section II were applied. Multi-stage testing, data analysis, and iterative enhancements were used to steadily advance the fault classification performance of the overall system. All three traditional machine-learning methods consistently classified a subset of the faults with an accuracy of 90% on average. These preliminary tests indicated that pairing the full dataset with a more robust multiclass classification model may have promising results in practical applications. This represented a promising start, showcasing the potential for data-driven solutions to effectively address the PHM difficulties currently faced within the energy sector.

The limited efficiency and precision of ML models restricted the initial tests from classifying more than four fault classes at a time. In pursuit of a more powerful model capable of handling tens of classes of power transformer faults and transient disturbances, this research replaced the ML algorithms with a Transformer-based network. The decision to take advantage of the Transformer architecture was mainly inspired by its recent success in handling large multiclass classification problems within popular LSF-Models.

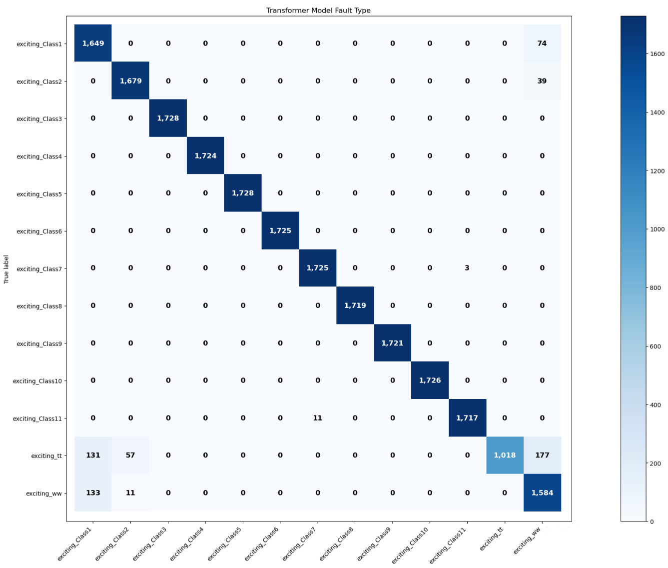
C. Structure of Fault Classification Architecture

The baseline Transformer architecture for three-phase multiclass classification from [18] was originally hard-coded to be used in conjunction with a proprietary private dataset. Without access to this dataset, this research utilized reverse engineering to extract the key components and structures of the algorithm. To prove that this type of model could be used for a range of applications related to condition monitoring of power transformers, the program was also dynamically generalized for a diverse range of data formats.

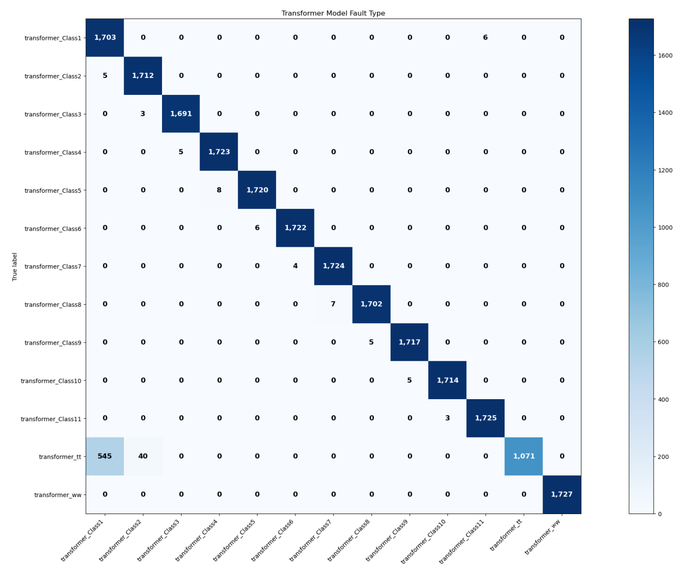
Methods for comma-delimited text file data processing were first incorporated to properly manipulate the data into manageable Numpy arrays. The included time measurements were insignificant for model training; therefore, the data file was reorganized into a 726×3 matrix for the three-phase differential current measurements. Since the number of fault examples varied between a majority of the classes, class normalization techniques were used to equalize the training data. The model struggled to improve when trained solely on the normalized data. To remediate this, the first epoch trained the model on standard data and then completed the remaining epochs with the proper normalized data. When trained with an appropriate number of total epochs, this inconsistency had a negligible impact on overall model accuracy.

The model was initialized by instantiating a Keras tensor with the same dimensions as the restructured data files (726-by-3). After additional normalization, the tensor was reshaped to incorporate a batch size dimension and then flattened with respect to each timestep. A dense layer of dimension 1024 was applied and followed by a dropout of 20%. A dense layer of dimension 64 was applied and followed by another 20% dropout. The rectified linear unit (ReLU) activation function [19] was used for all intermediate dense layers:

$$y = f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{1}$$



(a) ISPAR Exciting Transformer Confusion Matrix



(b) Power Transformer Confusion Matrix

Fig. 3: Two subsections of the overall 45-class confusion matrix display the most prominent outliers. Most notably, the turn-to-turn fault location was commonly confused with “Class1” (phase A to ground) for both types of transformers.

The process of positional embedding was applied using the batch size as the embedding vocabulary size and the dense layer dimension as the dense embedding output dimension. A stack of four Transformer encoders composed the body of the neural network. Each encoder performed layer normalization on the input tensor and sequentially performed multi-headed attention that consisted of four attention heads of size 64 with a dropout probability of 20%. With d_k as the dimension of K, the scaled dot-product self-attention mechanism as a function of the queries Q, keys K, and values V can be expressed as

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2}$$

A residual connection was established by adding the output of the self-attention mechanism to the original normalized tensor. After additional normalization, a feedforward neural network (FNN) was applied. This process consisted of several iterations of the ReLU activation function dense layers and dropouts. A final residual connection was established between the input and output of the FFN.

Following the data preprocessing stage, the model started preparing for the process of fault classification. The data from the stacked encoder was flattened and processed through the following series of dense ReLU layers and dropouts: dense(256), dropout(20%), dense(128), dense(32), and dropout(20%). The inclusion of frequent dropout layers worked to keep the model from overfitting the data. This is an important process that also helps ensure future generalizability. The final dense layer dimension equaled the total number of evaluated classes. This layer also applied a softmax activation function which normalized the data into representative probabilities. This model took advantage of the Adam optimizer [20], an extended version of stochastic gradient descent optimization, with a learning rate α of 0.001. After 150 epochs, training checkpoints were saved as an H5 file, training history as a Pickle file, and the overall trained model as a Keras file.

D. Structure of Natural Language Processor

In addition to using the Transformer architecture for three-phase time-series fault classification, this network was also exercised as a simple NLP. The implementation was supported by an accompanying JSON library that stored individual topics of conversation organized into distinct classes. Each class contained a tag, a list of patterns, and a list of responses. The pattern list contained words or phrases that would be associated with the corresponding class. The library was first processed into organized lists that were then labeled for training. The text data was vectorized by converting the text into integers and removing all punctuation. The sequence of numbers was then split into lists of word-based tokens to retain important relational components. They were then reorganized into an $m \times 20$ matrix for m total pattern entries.

The NLP model first utilized an embedding layer with a vocabulary size of 1000 and a dense embedding dimension of 32. This layer also kept the length of input sequences constant at 20. A one-dimensional global average pooling operation layer was added to map the features of the embedding layer. Two dense ReLU activation function layers of dimension 16 were added sequentially. The final dense layer contained the same number of nodes as the total evaluated classes and applied the softmax activation function to obtain the distribution of probabilities. This model used a sequential grouping to create the complete model object. and they were compiled using the Adam optimizer. The model was run for 550 epochs and saved as an H5 file upon completion.

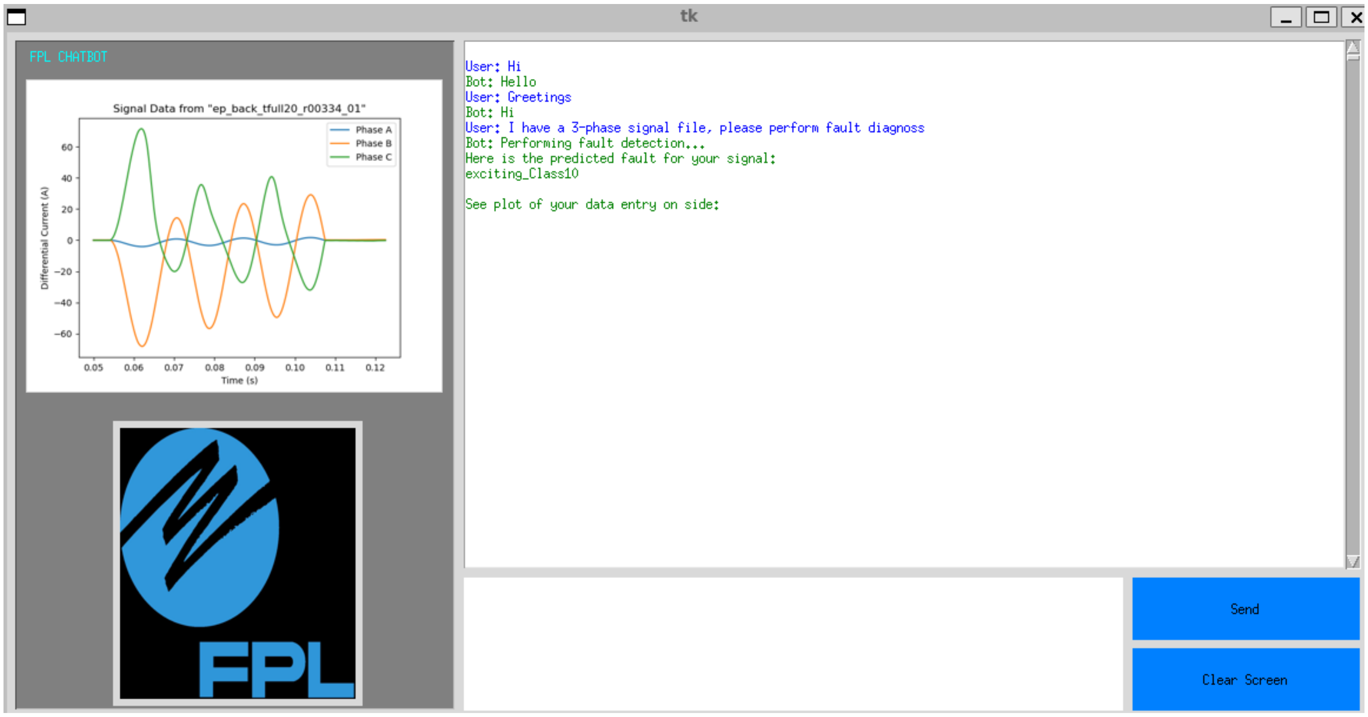


Fig. 4: The GUI performing a fault diagnosis. The predictive model uploads the graphed signals shown on the left and produces the fault class in the conversation window on the right. Users have the option to clear the screen or continue responding in the input box below. As shown above, the conversational nature of the GUI understands context and even with spelling errors, fault diagnostics can still be activated.

IV. Results Analysis and System Demonstration

A. Fault Classification Model Performance

The predictive Transformer model was designed to analyze three-phase signals recorded over time for feature extraction and classification. It aimed to detect and classify faults among 45 different power transformer fault types using a dataset of over 100,000 simulated examples [15]. This model achieved a Matthew's Correlation Coefficient accuracy score of 97.2% in classifying the 45 classes of fault signals. The matrix, partially shown in Fig. 3, also shows that a specific handful of classes were more prone to error than others, including the *exciting turn-to-turn (exciting-tt)*, *exciting winding-to-winding(exciting-ww)*, *transformer-Class1*, *transformer-tt*, *magnetic inrush*, and *series-tt* classes. Among the *exciting* categories, a significant number of testing samples from *exciting-tt* and *exciting-ww* were classified as *exciting-Class1* and *exciting-Class2*, and vice versa. Additional outliers included the erroneous classification of *magnetic inrush* as *sympathetic inrush*. The most prominent outlier from the testing data was the misclassification of *transformer-tt* as *transformer-Class1*. Further analysis revealed that the most inaccurate predictions stemmed from the *exciting-tt* and *transformer-tt* classes, which exhibited a total misclassification rate of 65%. In light of these findings, this study demonstrated the feasibility of classifying large-scale phase signals, while recognizing the presence of certain outlier classes. Viewing the concept of predictability and fault classification through the Transformer architecture demonstrated its potential for contributing to a largely accurate mechanism for reliable multiclass classification of three-phase time-series signals found in smart grid systems.

B. Features of Graphical User Interface

To showcase the diagnostic capabilities of the Transformer architecture, this research developed a Linux-based graphical user interface (GUI), shown in Fig. 4. The primary page of the GUI enables dynamic interaction with the chatbot. Key components include:
- **Input Box:** Users can input text for communication with the chatbot. The "send" button submits text to the chatbot, while the "clear screen" button begins a new conversation.
- **Conversation Viewer:** Displays user queries and chatbot responses in real-time.
- **Graph Viewer:** Presents graphs of classified three-phase time-series signals, aiding comprehension and visualization.
- **History Log (not shown):** Logs conversations and corresponding graphs for future review.

C. Interface Example for Power Transformer Fault Diagnosis

Utilizing the Transformer's predictive power, users engage with the chatbot through the interface for fault diagnosis. By uploading signal files, the model efficiently analyzes and classifies the data. The conversational chatbot responds to user queries and instructions with the following aspects:
User greeting and request of fault diagnosis – The chatbot greets users, initiates interactions and responds appropriately

to user prompts. The model performs text-based classification to predict whether the user wants the system to converse, perform fault diagnosis, or execute any other programmed task. Depending on the user input, the interface appropriately replies back or initializes the requested fault classification prediction model.

User prompted to upload three-phase signal files – Users are prompted to designate an anomalous three-phase signal file from the local drive as input for the fully trained model.

Display results – A graph of the anomalous signal is displayed above the logo, and the interface conversationally communicates the predicted fault type.

Conversational nuances – The chatbot accommodates conversational nuances and context, largely in part due to the model's ability to accurately generalize. This, in turn, promotes user-friendly interactions as the system does not get confused by small spelling mistakes or grammatical errors. Note in Fig.4 how the interface correctly classifies the request for fault diagnosis even though fault diagnosis is misspelled as "fault diagnosis."

V. Conclusions and Discussions

The presented study introduces a predictive Transformer model designed to analyze three-phase signals for power transformer fault classification. The model's commendable accuracy of 97.2% in identifying fault types highlights its efficacy. Leveraging MCC, the results probe further discussion of finer metrics that delve deeper into the model's performance, which aligns with an improvement from the initial machine-learning models. The results are promising, with the majority of the confusion matrix aligning accurately with expectations. The high overall accuracy, although impressive, is tempered by the challenges posed by outlier classes. Notably, the *transformer* and *exciting* classes exhibit inconsistencies in classification that require further investigation into their outliers and erroneous patterns. Although misclassifications exist, they serve as valuable insights into areas where the model's sensitivity can be refined. This study emphasizes that while automated fault classification is obtainable, continuous refinement is crucial to address nuanced variations in fault behavior.

The development of a Linux-based GUI further enhances the accessibility of the model. The interface offers a user-friendly platform for dynamic interaction and fault diagnosis, amplifying the model's applicability to real-world scenarios. In the future, other enhancements to this interface could prove crucial to improving the predictive model. These potential improvements include 1) expanding the framework to cross-reference historical prognostic databases and 2) encompassing online sensor condition monitoring in real-time. In conclusion, this research contributes to the evolving landscape of PHM and fault diagnosis for smart grid power transformers. This project serves as a proof-of-concept solution, showcasing how a single user interface can potentially improve the reliability and performance of power transformers throughout smart grid systems.

Acknowledgments

The authors would like to thank Siyuan Du for his contributions to the GUI and general research assistance.

References

[1] U.S. Energy Information Administration - EIA - independent statistics and analysis. Use of electricity - U.S. Energy Information Administration (EIA). (2023, April). https://www.eia.gov/energyexplained/electricity/use-of-electricity.php

[2] H. Ma, T. K. Saha, C. Ekanayake and D. Martin, "Smart Transformer for Smart Grid—Intelligent Framework and Techniques for Power Transformer Asset Management," in IEEE Transactions on Smart Grid, vol. 6, no. 2, pp. 1026-1034, March 2015, doi: 10.1109/TSG.2014.2384501.

[3] L. Ouyang, J. Wu, X. Jiang, D. Almeida and C. Wainwright, et al., "Training language models to follow instructions with human feedback," in Proc. Advances in Neural Information Processing Systems, 2022, pp. 27730-27744.

[4] A. Kirillov, E. Mintun, N. Ravi, H. Mao and C. Rolland, et al., "Segment Anything," [Online]. Available: https://arxiv.org/abs/2304.02643.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit and L. Jones, et al., "Attention is All You Need," Advances in neural information processing systems, vol. 30, 2017.

[6] J. Devlin, M. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," [Online]. Available: https://arxiv.org/abs/1810.04805.

[7] Y. Sun, S. Wang, Y. Li, S. Feng and X. Chen, et al., "ERNIE: Enhanced Representation through Knowledge Integration," [Online]. Available: https://arxiv.org/abs/1904.09223.

[8] H. Touvron, T. Lavril, G. Izacard, X. Martinet and M. Lachaux, et al., "Llama: Open and efficient foundation language models," [Online]. Available: https://arxiv.org/abs/2302.13971.

[9] A. Radford, K. Narasimhan, T. Salimans and I. Sutskever, "Improving language understanding by generative pre-training," [Online]. Available: https://openai.com/blog/chatgpt.

[10] OpenAI, "GPT-4 Technical Report," [Online]. Available: https://arxiv.org/abs/2303.08774.

[11] Li, Y., Wang, H., and Sun, M. (2023). ChatGPT-Like Large-Scale Foundation Models for Prognostics and Health Management: A Survey and Roadmaps. [Online]. Available: https://arxiv.org/abs/2305.06472

[12] H. Wang, Z. Liu, D. Peng and M.J. Zuo, "Interpretable convolutional neural network with multilayer wavelet for Noise-Robust Machinery fault diagnosis," Mech. Syst. Signal Pr., vol. 195, pp. 110314, 2023.

[13] Y. Wang and H. Wang, "Wavelet attention-powered neural network framework with hierarchical dynamic frequency learning for lithium-ion battery state of health prediction," Journal of Energy Storage, vol. 61, pp.106697, 2023.

[14] H. Liu, J. Zhou, Y. Zheng, W. Jiang and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," ISA T., vol. 77, pp. 167-178, 2018.

[15] Pallav K Bera, Can Isik, Vajendra Kumar, February 13, 2020, "Transients and Faults in Power Transformers and Phase Angle Regulators (DATASET)", IEEE Dataport, doi: https://dx.doi.org/10.21227/1d1w-q940.

[16] "A guide for using The wavelet transform in machine learning," ML Fundamentals, https://ataspinar.com/2018/12/21/a-guide-for-using-the-wavelet-transform-in-machine-learning/ [Accessed Aug. 29, 2023].

[17] F. Wasilewski, "Discrete wavelet transform (DWT)," Discrete Wavelet Transform (DWT) - PyWavelets Documentation, https://pywavelets.readthedocs.io/en/latest/ref/dwt-discrete-wavelet-transform.html [Accessed Aug. 29, 2023].

[18] J. B. Thomas and S. K.V., "Neural architecture search algorithm to optimize deep transformer model for fault detection in Electrical Power Distribution Systems," Engineering Applications of Artificial Intelligence, vol. 120, p. 105890, 2023. doi:10.1016/j.engappai.2023.105890

[19] V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines." Available: https://www.cs.toronto.edu/h̃inton/absps/reluICML.pdf

[20] D. Kingma and J. Lei Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION," Jan. 2017. Available: https://arxiv.org/pdf/1412.6980.pdf