中国目前的工业发展正处于全面推进、高质量发展的阶段。自改革开放以来，中国的工业经济取得了惊人的成就，成为全球最大的制造业国家和工业生产大国之一。而在工业生产环境中，工业设备的状态分析和故障检测的重要性不言而喻。这些分析和检测任务旨在实时监测和评估设备、机器或系统的运行状态，以便及时发现潜在的故障、异常或不良情况。随着工业设备的复杂性和精密性日益增高，可能导致其发生故障的因素和发生故障后的维修成本都在日益增多和增高，因此分析和预测变压器的运行状态，发现可能存在的故障具有重要的价值和意义。工业时序数据是反应工业物理场景状态变化的一组数据，其在时序上具有趋势性，且多变量之间会相互影响。现代工业逐渐朝着数据化和自动化的方向发展的同时，每天都会产生大量的工业时序数据，其中包含着许多生产设备的状态信息，对这些海量数据的分析和预测可以有效地帮助我们判断工业设备地运行状态和发现可能存在地故障

China's industrial development is currently at a stage of comprehensive and high-quality development. Since the reform and opening up, China's industrial economy has made amazing achievements and become one of the largest manufacturing countries and industrial production powers in the world. And the importance of condition analysis and fault detection of industrial equipment in an industrial production environment cannot be overstated. These analysis and inspection tasks are designed to monitor and evaluate the operational status of equipment, machines or systems in real time in order to detect potential failures, abnormalities or adverse conditions in a timely manner. With the increasing complexity and sophistication of industrial equipment, the factors that may cause it to fail and the cost of repairing it after a failure are increasing and increasing, so analyzing and predicting the operational status of transformers and detecting possible faults is of great value and significance. Industrial time series data is a set of data reflecting the change of state of industrial physical scenes, which has a trend in time series and multiple variables can affect each other. While modern industry is gradually developing in the direction of data and automation, a large amount of industrial timing data is generated every day, which contains the status information of many production equipments.

目前对工业时序数据的分析中多采用人工的手段，一般是先通过传感器获取相应时序数据，基于工作人员的经验和领域知识，通过制定一系列规则，阈值，模式和指标来进行数据分析和故障检测，以判断数据是否异常或故障是否发生。有些地方的工作人员还会手动选择和提取时序数据的特征， 并使用一些统计方法或建模来分析时序数据。以上人工手段存在不少问题，比如人工分析需要花费大量的时间，人工分析可能会导致结果的不准确性，过于依赖工作人员的经验和技能，缺少可视化展示界面等，因此实现一个能够分析预测工业时序数据并提供可视化展示界面的原型系统有十分重要的价值。同时由于工业真实故障样本稀少，系统还应能够对故障样本进行增强。

At present, the analysis of industrial timing data is mostly done manually, generally by acquiring the corresponding timing data through sensors, and then performing data analysis and fault detection by developing a series of rules, thresholds, patterns and indicators based on the staff's experience and domain knowledge to determine whether the data is abnormal or whether a fault has occurred. In some places, staffs also manually select and extract features of the timing data, and use some statistical methods or modeling to analyze the timing data. The above manual methods have many problems, such as manual analysis takes a lot of time, manual analysis may lead to inaccurate results, too dependent on staff experience and skills, lack of visual display interface, etc. Therefore, it is very important to implement a prototype system that can analyze and predict industrial time series data and provide visual display interface. The system should also be able to enhance the failure samples due to the scarcity of real industrial failure samples.

本文以工业油浸式变压器为例子，首先介绍了工业时序数据分析预测现有的业务场景以及使用该软件后的业务场景，然后介绍了功能需求和非功能需求。最后，根据需求分析的结果，设计了本原型系统的框架。

In this paper, taking an industrial oil-immersed transformer as an example, we first introduce the existing business scenario of industrial timing data analysis and the business scenario after using this software, and then introduce the functional and non-functional requirements. Finally, based on the results of the requirements analysis, the framework of this prototype system is designed.

然后介绍了为原型系统设计实现的几个模型，分别是基于transformer的工业长期时序数据异常检测模型，基于GMM的长期故障时序数据聚类模型，基于VARMAX的工业时序数据预测模型，基于LSTM的工业短期故障时序数据分类模型和基于时域和频域的故障时序数据增强算法。

Then several models designed and implemented for the prototype system are introduced, which are transformer-based anomaly detection model for industrial long-term timing data, GMM-based clustering model for long-term fault timing data, VARMAX-based prediction model for industrial timing data, LSTM-based classification model for industrial short-term fault timing data and fault timing data enhancement algorithm based on time domain and frequency domain.

基于Transformer模型的时序数据异常检测模型主要是通过引入LayerDrop层，对抗训练等方式改进transformer模型，实现了对工业长期时序数据异常检测功能。数据预处理主要包括数据补全，归一化处理和滑动窗口处理，LayerDrop通过随机地丢弃部分层来减小模型的容量，对于具有高维度且变化较为复杂的工业时序数据，可以避免模型过度学习训练数据中的噪声或特定的模式，从而减轻了过拟合的风险。对抗训练则通过对数据重构，分析重构异常等方式实现。同时还有增加层等改进。

The transformer model-based anomaly detection model for industrial long-term timing data is mainly improved by introducing LayerDrop layer and adversarial training to improve the transformer model. LayerDrop reduces the capacity of the model by randomly dropping some layers. For industrial time-series data with high dimensionality and complex changes, the model can avoid overlearning the noise or specific patterns in the training data, thus mitigating the risk of overfitting. Counter-training is then achieved by reconstructing the data and analyzing reconstruction anomalies. There are also improvements such as adding layers.

基于高斯混合模型的故障数据聚类模型主要通过引入小波包变换特征提取实现了对检测出的故障油气数据的聚类分析功能，它能够将时序数据分解成一系列小波包系数，这些系数包含了小波变换中的所有尺度和频率成分，因此能够更全面地反映出数据的特征。

The Gaussian mixture model based fault data clustering model achieves the clustering analysis function of detected fault oil and gas data mainly by introducing wavelet packet transform feature extraction, which can decompose the time-series data into a series of wavelet packet coefficients, which contain all scale and frequency components in wavelet transform and thus can reflect the characteristics of the data more comprehensively.

基于VARMAX的工业长期时序数据预测则通过引入自适应超参数优化等方式实现，模型首先提取油气时序数据的时间序列特征，同时通过Optuna实现模型参数的自适应优化，通过TPE算法找到模型的自回归阶数，滑动平均阶数，趋势项和模型的正则化参数的最佳组合。

VARMAX-based industrial long-term time series data forecasting is achieved by introducing adaptive hyperparameter optimization, etc. The model first extracts the time series characteristics of oil and gas time series data, and at the same time implements adaptive optimization of model parameters by Optuna, which finds the best combination of autoregressive order, sliding average order, trend term and regularization parameters of the model by TPE algorithm.

基于LSTM的短期时序数据实时故障分类则是通过引入one-hot编码等方式实现，将标签转换为one-hot编码，具体来说是将一个离散的分类变量映射为一个稀疏向量，其中只有一个元素为1，其他元素均为0，这个元素的索引位置对应于该变量所属的类别，通过多层LSTM实现对故障数据的分类。同时模型还包括一个RMSprop优化器，使用自适应学习率，可以根据不同权重的历史梯度信息调整学习率。

LSTM-based real-time fault classification of short-term time-series data is achieved by introducing one-hot coding, etc., converting labels into one-hot coding, specifically by mapping a discrete categorical variable into a sparse vector, where only one element is 1 and all other elements are 0. The index position of this element corresponds to the category to which the variable belongs, and by multi-layer LSTM The classification of faulty data is achieved through a multilayer LSTM. The model also includes an RMSprop optimizer that uses an adaptive learning rate, which can be adjusted according to the historical gradient information of different weights.

基于时域和频域的故障数据增强算法则通过时域伸缩，频域转换，噪声注入等方式实现。通过向油气和油温故障数据注入高斯噪声生成新的故障数据。特别的，根据不同油气参量以及油温变化幅度的不同，每个参量都有一个噪声因子来控制噪声的强度，每个参量的噪声因子根据各个油气参量以及油温的物理变化规律设定。同时最后会有一个非正判断保证噪声注入之后的数据不会存在负值。时间缩放则是通过拉伸和压缩时间轴来生成新的故障数据，压缩方式为隔点取样，按照压缩比例隔点采样。拉伸的方式则依靠样条插值实现。频域部分则通过噪声注入和频域转换实现，主要使用了快速傅里叶变换及其反变换的方法。

The time and frequency domain based fault data enhancement algorithm is then implemented by time domain scaling, frequency domain conversion, and noise injection. New fault data are generated by injecting Gaussian noise into the oil-gas and oil-temperature fault data. In particular, each parameter has a noise factor to control the noise intensity according to the different oil and gas parameters and the oil temperature variation magnitude, and the noise factor of each parameter is set according to the physical variation law of each oil and gas parameter and oil temperature. At the same time, there is a non-positive judgment at the end to ensure that the data will not have negative values after noise injection. Time scaling is used to generate new fault data by stretching and compressing the time axis, and the compression is done by sampling at intervals, according to the compression ratio. The stretching is achieved by interpolation of splines. The frequency domain part is achieved by noise injection and frequency transformation, mainly using the fast Fourier transform and its inverse transformation method.

最后，我们构建了一个完整的原型软件系统。该系统由服务器，后端和可视化交互式界面组成。可视化交互界面用React框架进行开发，完成了系统的可视化界面，同时完成了提供了用户的交互功能，其中数据可视化图表部分使用antv G2plot实现，提供了可交互式的可视化图表。后端主要使用Flask框架，提供数据管理服务，模型算法管理，时序数据分析预测和故障数据增强服务，模型实现使用TensorFlow，PyTorch等库，系统选择InfluxDB作为时序数据库部署在服务器上，数据库连接使用InfluxDB-Python库实现，数据库主要存储工业变压器的油气数据，线圈数据等工业时序数据。

Finally, we built a complete prototype software system. The system consists of a server, a backend and a visual interactive interface. The visual interactive interface is developed with React framework, which completes the visualization interface of the system and also completes the interactive functions provided to the users. The data visualization chart part is implemented using antv G2plot, which provides the interactive visualization chart. The back-end mainly uses Flask framework to provide data management services, model algorithm management, timing data analysis and prediction and fault data enhancement services, model implementation using TensorFlow, PyTorch and other libraries, the system chooses InfluxDB as the timing database deployed on the server, database connection using InfluxDB-Python library implementation. The database mainly stores the oil and gas data of industrial transformers, coil data and other industrial timing data.

为了验证原型系统的效果，我们使用了中国长江电力股份有限公司的三峡电厂右岸21B主变压器2022年5月份到2023年3月份监测的数据，同时通过真实数据的特点和各种故障的物理规律模拟了多组故障数据，对各个模型进行了对比实验，通过体征模型参数并计算模型的准确率等指标，实验结果证明系统有较高的准确性和可用性。

In order to verify the effectiveness of the prototype system, we used the data monitored from May 2022 to March 2023 for the right bank 21B main transformer of the Three Gorges Power Plant of China Yangtze River Power Co. We also simulated multiple sets of fault data through the characteristics of real data and the physical laws of various faults, and conducted comparative experiments on each model by corporately characterizing the model parameters and calculating the accuracy of the model and other indicators, and the experimental results proved that the system has high accuracy and usability.

同时模型还有一些方面值得改进，一方面系统对变化近似的不同故障数据的分析并不是很明确，同时实验使用的故障数据大多是模拟的数据，实验存在一定的局限性，还需要进一步的改进。

At the same time there are some aspects of the model that deserve improvement. On the one hand the system is not very clear in the analysis of different fault data that vary in approximation, and at the same time the fault data used in the experiments are mostly simulated data, the experiments have some limitations and need further improvement.