Multi-Source Deep Learning for TSO-Level Electricity Demand Forecasting

MSc Research Project
Msc Data Analytics

Jismol Thomas -
Student ID: 23414952

School of Computing
National College of Ireland

Supervisor: Harshani Nagahamulla

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Jismol Thomas -
Student ID:	23414952
Programme:	MSc Data Analytics
Year:	2025
Module:	MSc Research Project
Supervisor:	Harshani Nagahamulla
Submission Due Date:	28/01/2026
Project Title:	Multi-Source Deep Learning for TSO-Level Electricity Demand Forecasting
Word Count:	5748
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Jismol
Date:	27th January 2026

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Multi-Source Deep Learning for TSO-Level Electricity Demand Forecasting

Jismol Thomas -
23414952

Abstract

Proper short-term electricity demand forecasting is critical for the maintenance of secure and cost-effective grid operation at the Transmission System Operator (TSO) level. This paper compares the performance of three deep learning models, namely, Long Short-Term Memory (LSTM), the Temporal Fusion Transformer (TFT), and a hybrid TFT-LSTM residual model, on a multi-source dataset, which combines SMARD consumption and generation data, Open-Meteo weather variables, Eurostat socioeconomic indicators, and calendar features. The models are evaluated based on the point-forecast, probabilistic, and operational measures. The findings indicate that the TFT that has been optimized with Optuna scores the best in total point-forecast accuracy, whereas the hybrid TFT-LSTM Residual model has better peak-load accuracy by fixing systematic underestimation of the TFT results. Even though the transformer-based models tend to under-calibrate uncertainty intervals, the tuned TFT and hybrid architectures are highly effective in terms of early warning. These results indicate the power of transformer-based and hybrid deep learning techniques to forecast the peaks and multi-source TSO-level using various sources.

1 Introduction

Contemporary power systems require proper forecasting of short-term electricity demand. To book reserves, load balance generation portfolios, avoid congestion, and achieve general grid reliability, Transmission System Operators mainly depend on accurate day-ahead forecasts. Due to the fast transformation of the German electricity system, the requirement to find robust data-driven forecasting strategies has become more important. There are four TSOs in Germany, namely 50 Hertz, Amprion, Tennet, and TransnetBW. Even though, Conventional statistical tools like ARIMA and exponential smoothing are historically useful, cannot effectively represent complicated nonlinear and temporal interactions that are found in contemporary power systems. Development of novel deep learning models is necessary in these situations. The recent developments in deep learning, specifically, Long Short-Term Memory (LSTM) networks and Transformer-based architectures, demonstrated good results in load prediction and other energy-related services Giacomazzi et al. (2023). Mainly, LSTMs are useful with the help of gated memory units in modelling sequential patterns. Architectures like the Temporal Fusion Transformer (TFT) use attention to learn complex interactions in time, feature importance, and multi-horizon interactions. There have also been hybrid architectures involving recurrent and transformer components, which seek to use the complementary modelling

strengths with residual Learning. Against this backdrop, this study addresses the following research question:
How accurately LSTM, Temporal Fusion Transformer (TFT), and a hybrid TFT-LSTM residual architecture predict TSO-level electricity demand in Germany when the models are trained on multi-source temporal, weather, and socioeconomic variables?
In order to respond to this question, we have five main aims of the study:

1. Build a multi-source dataset using SMARD electricity data, Open-Meteo weather variables, Eurostat socioeconomic variables, and calendar features.
2. Create and fine-tune baseline LSTM, TFT with a look-back window of 168 hours and a 24-hour forecast horizon. Also, develop and fine-tune a Hybrid TFT-LSTM Residual model.
3. Compare the models in terms of accuracy of forecasting, measuring uncertainty, detecting peak events, and early warning performance.
4. Critically evaluate results to find implications, limitations, and opportunities.
5. Develop an interactive Power BI dashboard to visualise TSO-level demand, model forecasts for supporting interpretability and operational insights.

This research has three significant contributions to the literature. It first offers a new hybrid TFT-LSTM Residual architecture. Secondly, it offers a comparison of LSTM, TFT, and hybrid TFT-LSTM Residual architectures to date, with the most extensive comparison of German TSO-level load forecasting on real-world multi-source data. Third, it develops an interactive Power BI dashboard to visualize the TSO–level forecast for all 4 TSOs of Germany in a single place.
The rest of the paper outlines the following:
Section 2 assess the existing literature on the topic of electricity demand forecasting and deep learning architecture, which forms the theoretical framework of modelling decisions made in the research. Section 3 is Methodology. This section provides a detailed description of the entire research process, data collection, preprocessing, feature engineering, and model selection principles, as well as the evaluation framework used in this research. Section 4 Design and Specification. This section outlines the system architecture, modelling framework, as well as the major design decisions that are detailed. Section 5 is the implementation section. This section describes how the proposed models and data pipeline are implemented, which involves tools and libraries, and the production of final datasets and model results. Section 6 is the Evaluation section. This section fulfills the assessment of all the models in terms of point, probabilistic, and operational measures. Lastly, Section 7 has concluded the findings, limitations, and future work directions of this research.

2 Related Work

The predictability of short-term electricity loads is critical to the stable operation of the power system. This section critically analyzes the current literature through the review of the traditional, deep learning, transformer-based, and hybrid forecasting models.

2.1 TSO-level Load Forecasting and the German Context

An increasing amount of literature dwells on load forecasting on open European sources of data. Behm et al. (2020) demonstrate that ANN models based on relatively simple architectures that are trained on ENTSO-E load and weather data can fairly recreate load profiles on a national scale. With their existing work, the possibility of reproducible forecasting is proven, but not in connection with short-term operational forecasting and deep learning techniques. The traditional statistics modelling is still used in the operational or planning environment. Generalised additive models (GAMs) are the forecasting models used by Marino et al. (2016) to predict the load of the German systems with a focus on interpretability. Nonlinear interactions between weather, socioeconomic, and operational effects are difficult to model in GAMs, restricting their application to volatile short time horizons. Maciejowska et al. (2021) indicate that predictions of the load and renewable energy could improve day-ahead price forecasting performance, which explains the significance of load forecasting in electricity markets. On the same note, Narajewski (2022) and Ziel and Weron (2018) demonstrate that the probability-based information enhances the imbalance price modelling. Although these studies highlight the importance of accurate load forecasts in operations, they do not formulate specialised TSO-level forecasting architectures, but instead use load as an input to market models. Lago et al. (2018) also prove the superiority of deep learning methods over traditional algorithms when it comes to electricity price forecasting, one of the problem domains that has common temporal aspects with load prediction. In general, there is a scarcity of existing deep learning algorithms that are specifically intended to be applied to multi-TSO load data to predict the short-term load in the market. This inspires the idea of strict DL-based TSO-oriented research.

2.2 Deep Learning for System-level Load

The ability to represent nonlinear temporal dependencies has made deep neural networks the focus of STLF. Such a transition is supported by early reviews, which offer crucial theoretical backgrounds. Raza and Khosravi (2015) presents a systematic comparison of the artificial intelligence methods in the field of forecasting smart grids, stating that when adequate training data is present, deep architectures tend to perform better than shallow networks. This observation has been confirmed in a multiplicity of later studies. Employed LSTM networks are shown to be superior to feedforward neural networks in situations with national-scale load series, Gao et al. (2019). These models, however, do not include attention mechanisms or components based on transformers, which is now the state of the art. Wu et al. (2021) builds upon the LSTM strategies by employing deep belief networks along with the copula models to formulate the complex load distributions, but their computational needs would restrain their real-time use. A number of publications suggest improvement methods to LSTM systems. The study by Chen et al. (2018) combines graph convolutional networks (GCNs) with LSTMs to learn spatial and temporal associations of regional load data simultaneously. Nabavi et al. (2021) use discrete wavelet transforms before LSTM modelling and demonstrate better results where they extract multi-scale temporal features. Likewise, Gao et al. (2019) suggests an EMD-GRU based on the hybrid that breaks down load signals and then applies gated recurrent units to it, which results in enhancements in varying load patterns. Although such hybrid LSTM schemes enhance accuracy, they are limited to recurrent models and hardly involve probabilistic forecasting, which is a critical need of risk-sensitive TSO decision-making. In addition to LSTM methods, ensemble methods have also performed well. Massaoudi et al. (2021) show that the stacked ensemble models with gradient boosting (LGBM-XGB) and neural networks can not only enhance the accuracy of short-term load forecasting but also incur a higher computational cost. These ensemble methods are indicative of the possibility of using multiple learning paradigms, but they fail to draw attention mechanisms that allow Transformers to learn long-range dependencies. One of the common weaknesses in LSTM-based research is the emphasis on single regions or single entity load forecasting. Guo et al. (2024) present a hybridised forecasting approach that captures long-term dynamics in day-ahead electricity price series. Their model combines advanced deep learning and statistical techniques to achieve improved predictive accuracy. The single-series LSTM models are not capable of explaining interconnections across borders because interdependent balancing regions, such as German TSOs, are spatially correlated.

2.3 Forecasting based on Transformer and TFT

Transformer architectures have reconstructed sequence modelling, providing parallel and long-range temporal reasoning. They have proven to be significantly better than the current networks with their rapid adoption in STLF. Giacomazzi et al. (2023) compare LSTM and Temporal Fusion Transformer (TFT) models used in a variety of grid layers and demonstrate that LSTM is weaker than TFT with added exogenous factors in the form of weather and calendar variables. In their work, however, hyperparameter optimisation is not investigated, nor is TSO-level forecasting under scrutiny. Wan et al. (2019) use TFT to forecast the demand at the city scale and show steady improvements over plain LSTM architectures. On the same note, Lu and Chen (2024) suggest a variant of the data-slicing transformer in high-renewables systems. Although these methods emphasise the flexibility of the transformers, the majority of uses are at the distribution or municipal level. Different deep architectures have also appeared. Wan et al. (2019) present a temporal convolutional network-based multivariate forecasting model that proves that convolution-based systems are able to effectively learn temporal regularities using dilated causal convolutions. These results support the capabilities of more complex neural architectures in comparison with recurrent designs in the application to TSO-level applications. Transformers also have their problems despite their strengths. According to Zhou et al. (2021), without supplementation by residual learning or decomposition, transformer-based models can have poor peak representation and interval miscalibration. This inspires the hybrid designs implementing transformer interpretability with the good performance of LSTM in the short-term autocorrelation procedure. The hybrid TFT-LSTM residual model is considered in this research.

2.4 Probabilistic and Hybrid Approaches

Probability forecasting is becoming another important factor as renewable penetration is bringing additional uncertainty. The relevance of the probabilistic approaches has been determined by means of a competitive benchmark. The article by Hong et al. (2016) is the result of the Global Energy Forecasting Competition 2014 and revealed that a probability approach can considerably improve the decision-making process in the area of power systems. This work provided the standards of evaluation, such as pinball loss and reliability measures, which have become popular in the field. On the basis of these results, Khan et al. (2021) demonstrate that quantile regression deep learning models can be well calibrated on loss functions of the proper type. Moesmann et al. (2025) come up with a hybrid model of BiGRU-GAM-Gaussian process that is high in interval accuracy, but with the drawback of high computational complexity, with limited system-level testing. The hybrid neural architecture has become popular in STLF robustness. Qin et al. (2024) offer a channel-attention hybrid framework that enhances the distribution network load forecasting. Nevertheless, none of these approaches uses transformer encoders or makes particular reference to TSO-level forecasting. More importantly, a published study has not investigated the TFT-LSTM residual hybrid in system load forecasting at a large scale, a gap that is filled in this thesis.

2.5 Hyperparameter Optimization and Reproducibility

Despite the fact that the hyperparameter optimization (HPO) method has been identified to play a significant role in the performance of a DL model, it is not yet the practice applied to energy forecasting consistency. Despite its inefficiency, traditional grid search is still in use. Bayesian optimization, Snoek et al. (2012), as well as new models like Optuna, Akiba et al. (2019), are more efficient in search and better reproducible. Ziel and Weron (2018) discusses high-dimensional modeling proxies of electricity prediction and underlines the relevance of a systematic feature selection process with model validation. Their study proves that special concern for the model specification and cross-validation strategies is critical to the sound forecasting behavior, in particular, when multiple architectures are compared. However, few studies use HPO to solve transformer-based models, and none directly compares HPO to residual hybrid TFT LSTM-based predictors of TSO level predictions. This paper will fill this void by optimising all the model families using systematic optimization (Optuna).

2.6 Summary and Research Gap

Open data supports the effective TSO-level modelling. LSTMs are strong baselines but have an inability to perform long-range modelling, and transformers are skilled multi horizon, multi-source modelling. Probabilistic and hybrid models are efficient yet not easily applicable to TSO level systems. Advanced structures have limited studies on hyperparameter optimisation. None of the studies gives a consistent comparison between LSTM, TFT, and a hybrid TFT-LSTM residual in short-term multi-source TSO level load forecasting based on probabilistic and business-relevant measures. This gap is the driving force behind this research.

2.7 Summary Table of Influential Work

Table 1 provides a summary of influential works.

Reference	Approach	Strengths	Weaknesses
Behm et al. (2020)	ANN with ENTSO-E + weather	Reproducible effective national profiling	No modern DL or uncertainty modelling
Ziel and Weron (2018)	GAM for German load	Interpretability, robust mid-term forecasts	Limited nonlinear modelling
Raza and Khosravi (2015)	Review of AI techniques	Comprehensive comparison	Pre-transformer era
Hong et al. (2016)	Probabilistic benchmarking	Established evaluation metrics	Limited competition scope
Lago et al. (2018)	DL for price forecasting	Strong empirical validation	Focuses on price, not load
Gao et al. (2019)	LSTM short-term load forecasting	Strong temporal learning	No uncertainty modelling or attention mechanisms
Massaoudi et al. (2021)	Stacked LGBM–XGB–MLP	Robust ensemble results	High computational cost
Ouyang et al. (2019)	DBN–Copula	Effective distribution modelling	High computational cost
Chen et al. (2018)	Deep residual network	Strong feature learning, good accuracy	No explicit probabilistic modelling
Nabavi et al. (2021)	DWT–LSTM	Multi-scale extraction	Mainly buildings, not TSO-level
Giacomazzi et al. (2023)	TFT	Interpretable, strong exogenous handling	Limited hyperparameter optimisation
Lu and Chen (2024)	Data-slicing Transformer	Good multi-feature integration	Limited probabilistic analysis
Khan et al. (2021)	Quantile EEMD–ELM	Well-calibrated intervals	Limited covariate variety
Narajewski (2022)	Probabilistic price/imbalance	Strong uncertainty modelling	Market-specific, high complexity
Qin et al. (2024)	Channel-attention hybrid	Robust feature extraction	No transformer encoder

Table 1: Summary of key related works

3 Methodology

This part outlines the entire methodological approach embraced to predict hourly electricity demand at the level of the German Transmission System Operator (TSO).

3.1 Data Collection

This research used various publicly available and verified data sources, to measure operational, environmental, and structural predictors of electricity demand. Below are the data sources used in the research.

1. SMARD: The SMARD platform offers the hourly consumption of electricity, power generation mix, and day-ahead market prices of the DE/LU bidding area.
2. Open-Meteo: Open-Meteo API provides hourly weather data, per TSO headquarters, comprising temperature, humidity, wind speed, shortwave radiation, rainfall, and surface pressure.
3. Python holidays: Python holidays is a Python library that produces German national indicators of public holidays.
4. Eurostat (NUTS3): Eurostat releases gross domestic product (GDP) and population density data at NUTS3 level, which are released yearly.

Dataset	**Description**	**Frequency**	**Variables Used**
SMARD Consumption, Generation & Day ahead market price	Electricity demand, renewable generation mix, and day-ahead market prices for the DE/LU bidding zone.	Hourly	Consumption (MW), Solar Generation, Wind Onshore, Wind Offshore, Total Generation, Day-Ahead Price
Open-Meteo Weather Data	Meteorological conditions for each TSO headquarters.	Hourly	Temperature, Humidity, Wind Speed, Shortwave Radiation, Rainfall, Air Pressure
Python Holidays	German national public holiday indicators.	Daily	Holiday Flag (0/1)
Eurostat Socioeconomic Data	Socioeconomic measures at the NUTS3 level.	Annual	GDP, Population Density
Zenodo TSO Mapping Dataset	Geospatial mapping between NUTS3 regions and TSO territories.	Static	NUTS3–TSO Mapping shape files

Table 2: Summary of Datasets, Frequency, and Variables Used

SMARD and Open-Meteo datasets include all the time period between 1 January 2020 and 30 September 2025, whereas the Eurostat datasets cover the time period between 2020 and 2022.

3.2 Data Preprocessing

This helps to obtain structural, temporal, and spatial coherence among data sources.

Temporal Alignment

The SMARD times that were initially on CET/CEST were localised and translated to UTC. An hourly time index was produced continuously, and the gaps in the time index were filled in by linear interpolation or forward filling. Variables were resampled to an hourly frequency in order to normalise time granularity.

Cleaning and Normalisation

Numerical data were standardised, and this entailed the elimination of currency symbols, thousand separators, and non-uniform units. Energy values were brought into harmony in MWh and prices in the form of €/MWh. Datasets were standardised to have the same column names so that they can be easily merged.

3.3 Socioeconomic–Spatial Integration

The indicators presented by Eurostat are at the NUTS3 level, but the electricity data are at the TSO level. In order to reconcile these spatial hierarchies:

Name Normalisation

The names of the districts were first converted to lowercase, symbols were removed, and administrative prefixes like Kreis and Landkreis were removed.

Fuzzy Matching (RapidFuzz)

In order to identify the TSO and NUTS3 mapping, refer to the Zenodo dataset. A fuzzy matching strategy, which used RapidFuzz WRatio scorer, was used to match each NUTS3 region to TSO administrative regions. Upon normalizing the name of districts, the fuzzy similarity scores were high enough to make sure that TSO assignments were reliable.

GDP Trend Forecasting

A simple linear regression model was applied to historical GDP trends at the TSO level to fill the gaps in the socioeconomic values to be used between consecutive observations of Eurostat on a yearly basis. This model was further applied in interpolating and predicting GDP values using the intermediate years required to match the hourly data. The obtained continuous series of GDP made the socioeconomic information complete and consistent over time for model training (Hyndman and Athanasopoulos; 2018).

$$\mathrm{GDP}(t) = \beta_0 + \beta_1 t \tag{1}$$

3.4 Dataset Integration

After preprocessing each data source, all datasets were merged into a single unified hourly panel at the TSO level. SMARD consumption, generation, and market price data were joined with Open-Meteo weather variables on the basis of timestamp (UTC) and TSO region. Public holiday indicators were merged using date keys. While socioeconomic variables (GDP and population density), mapped to TSOs through the fuzzy NUTS3–TSO matching procedure, were attached to every hourly observation through forward filling of annual values. This integration process ensured that each timestamp–TSO combination contained a complete set of operational, meteorological, socioeconomic, and calendar features required for model training.

3.5 Feature Engineering

Feature engineering improves model learning behaviour by capturing temporal structure, seasonal variation, and nonlinear relationships.

Lag and Rolling Features

For consumption, generation, weather and price variables, lagged values at 1h, 24h and 168h were constructed. Rolling statistics of 24h and 168h windows were used to find short-term and weekly trends and volatility.

Calendar and Temporal Features

Calendrical characteristics were added to record periodic and seasonal trends in the demand of electricity. These included hour-of-day indicators, day-of-week indicators, month indicators, weekend/weekday indicators, day-of-year indicators, year indicators and public holiday indicators.

Table 3: Summary of Engineered Features Used in the Forecasting Models

Feature Category	**Variables**	**Description / Purpose**
Lag Features	Lag_1h, Lag_24h, Lag_168h (load, weather, generation, price)	Capture short-term, daily, and weekly temporal dependencies and recurring patterns in demand, weather, generation, and price.
Rolling Window Features	Rolling_mean_24h, Rolling_mean_168h, Rolling_std_24h, Rolling_std_168h	Capture smoothed intra-day and weekly trends as well as volatility in load, weather, generation, and price variables.
Calendar Features	Hour, Day-of-Week, Month, Weekend flag, Holiday flag, Day-of-Year, Year	Represent human activity cycles, weekly seasonality, annual progression, and public holiday effects on electricity demand.
Weather Features	Temperature, Humidity, Wind Speed, Solar Radiation, Rainfall, Air Pressure	Core exogenous drivers of electricity demand and renewable generation variations.
Socioeconomic Features	GDP, Population Density	Reflect long-term demand behaviour differences across TSO regions, with annual values interpolated to hourly resolution.
Operational Market Features	Day-ahead Market Price	Represents market responses to the supply–demand balance and operational system conditions.
Interaction Features	Price–Temperature Interaction, Price–Wind Interaction, Renewables Ratio	Capture nonlinear relationships and combined effects between market, weather, and generation variables.

Feature Category	Variables	Description / Purpose
TSO Identity Encoding	TSO Label	Categorical identifier for the transmission operator region (50Hertz, Amprion, TenneT, TransnetBW).

3.6 Exploratory Data Analysis

All input datasets were examined through exploratory data analysis (EDA) to assess data integrity, temporal behaviour, and the relevance of engineered features. Figure 2 shows distinct seasonal trends in the long-term hourly load series, with higher demand in winter and lower consumption in summer. Figure 3 of daily and weekly load profile indicated similar patterns in the structure of the demand at TSOs. This is bimodal in a shape of daily pattern, with highest loads in the morning and evening, and weekdays loads are higher as compared to weekends. Figure 1 indicated the nonlinear relationship that was expected between temperature and load. The low and high temperatures would have an increase in consumption with low demand in mild temperatures. The differences in TSOs are also apparent. On the basis of the socioeconomic and industrial structure, there is a systematic increase in load levels in regions with a larger population (e.g., Amprion and TenneT) and also a decrease in smaller TSOs like TransnetBW. These evaluations affirm that it can be said that the meteorological, temporal, and regional socioeconomic variables are significant covariates in the expression of TSO-level electricity demand.

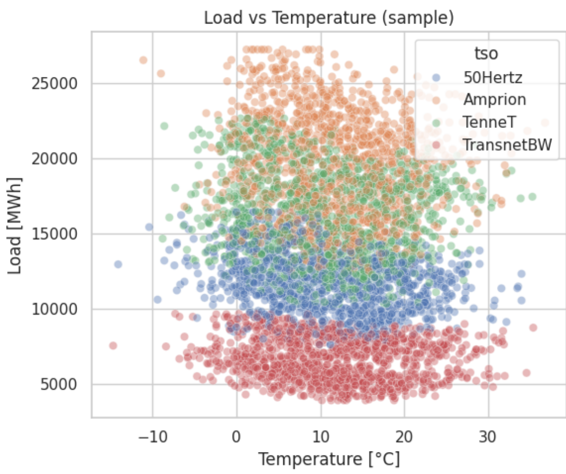

Figure 1: Load Vs Temperature

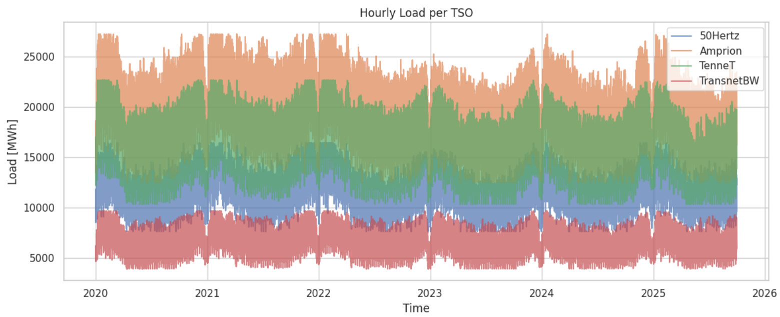

Figure 2: Hourly Per TSO

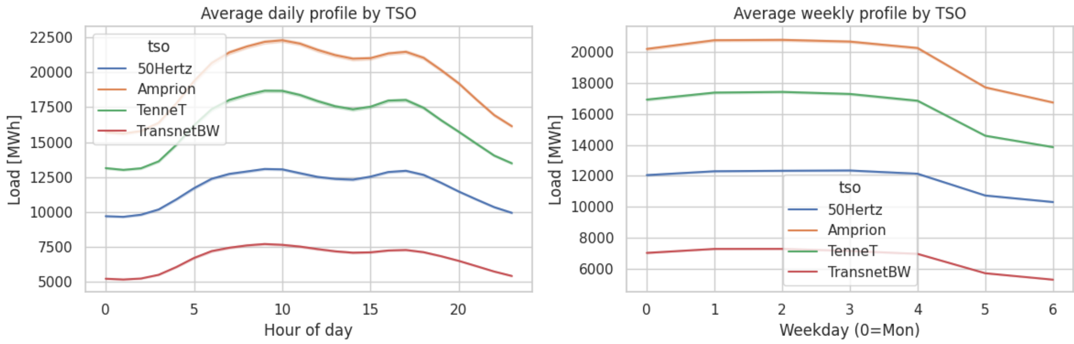

Figure 3: Average weekly profile by TSO & Average Daily profile by TSO

3.7 Model Development

Three forecasting architectures were developed:

LSTM Baseline

PyTorch was used to implement a three-layered 128-hidden-unit multivariate LSTM. It was trained using the Adam optimiser and MSE loss. A realistic validation was taken care of by a chronological split of 60/20/20.

Temporal Fusion Transformer (TFT)

The PyTorch-forecasting library was used to implement a TFT model. The architecture is based on variable selection networks, gated residual networks, and multi-head attention to represent multivariate time series. Probabilistic forecasting was made possible by quantile loss at P10, P50, and P90.

Hybrid TFT - LSTM Residual Model

A hybrid model defined in two stages was developed: (1) LSTM predicts the residual errors, (2) TFT predicts the base demand. The final predictions were calculated as:

$$\hat{y}_{final} = \hat{y}_{TFT} + \hat{r}_{LSTM}$$

3.8 Hyperparameter Optimisation

Optuna was used to tune the hyperparameters of all models. There were search spaces of learning rate, dropout, hidden units, attention heads (TFT), and sequence lengths. The best settings were chosen because of the validation loss.The search spaces of the hyperparameters were selected based on prior literature and preliminary experimental testing. This ensured a balance between predictive accuracy, generalisation capability, and computational feasibility.This design choice also ensures forecasting reliability and operational stability, which are critical requirements for real-time TSO decision-making.

3.9 Evaluation Methodology

The metrics of evaluation were point metrics (MAE, RMSE, MAPE, MASE, R2), probabilistic metrics (P50/P90 loss, coverage, MPIW), and operational metrics (peak-load error and early warning precision/recall). This multi-dimensional assessment is in line with the existing best practices in TSO-level forecasting studies.

3.10 Summary

The methodology incorporates multi-source information, spatial socioeconomic modelling, fuzzy inference systems, advanced feature engineering, and state-of-the-art deep learning architectures. This is a reliable method of predicting TSO-level electricity demands in Germany with the help of this systematic process.

4 Design Specification

Figure 4 gives the system architecture that was created in multi-source TSO-level demand forecasting of electricity. Heterogeneous data are converted into model-ready tensors and utilized to train deep learning forecasting models.

The architecture has five major sources of data: (1) SMARD consumption, generation, and day-ahead market prices (2) Hourly weather data of each TSO headquarters by Open-Meteo, (3) Eurostat socioeconomic indicators such as population density and GDP indicators at the NUTS3 level, (4) Python-holidays to provide information about German public holidays. (5) The Zenodo dataset for mapping the NUTS3 regions to TSO.

Data streams are processed within a unified ingestion and harmonisation layer, where time alignment, format normalisation, and data consistency checks are applied. A Spatial-Socioeconomic mapping, as one of the ingredients of this ingestion layer, is a combination of the Eurostat NUTS3 indicators and TSO regions with the Zenodo dataset, fuzzy matching, and GDP interpolation. The socioeconomic features by the hour were created in accordance with the scope of operations of particular TSOs. The Feature Engineering Layer forms predictive covariates like lag features. Structural, temporal, and socioeconomic descriptors are derived from rolling statistics, calendar indicators, and meteorological variables. A combination of these datasets provides a single TSO-level modelling dataset per hour, the input of all forecasting models. This information is fed through the Model Training and Forecasting layer, which is run with. three architectures; (1) an LSTM baseline model (2) a Temporal Fusion Transformer (TFT) baseline model (3) a Hybrid TFT-LSTM Residual model in which an LSTM learns to introduce residual corrections on the outputs of TFT. The Base Models block will involve the initial training of the three models using the help of default hyperparameters. Optuna Hyperparameter Tuning layer is then the next layer. This layer is an automated optimiser of the major model hyperparameters, such as the learning rate, hidden dimensions, dropout rate, and, in the TFT model, attention head count and gating network dimensions. The optimised versions of individual architectures are generated by reintroducing the tuned configurations into the model training workflow. The base and Optuna-tuned model outputs are fed into the Evaluation Layer.The measures of this layer are point measures (MAE, RMSE, R2, MAPE, MASE), probabilistic measures (pinball loss, prediction, interval coverage, MPIW) and operational measures (peak-load error and early-warning detection). Lastly, the power BI dashboard shows the performance of the Optuna-tuned models. This

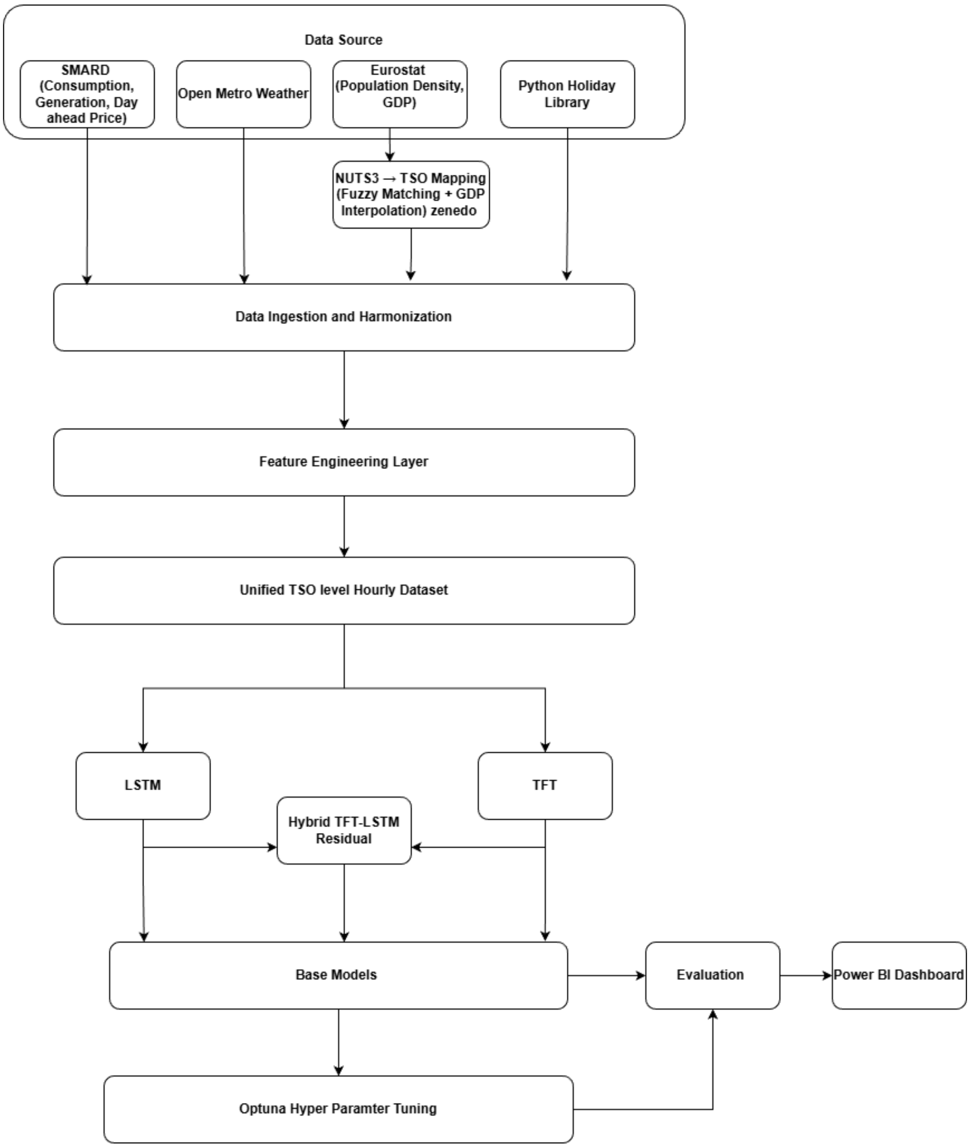

Figure 4: Architecture Diagram

facilitates interactivity on TSO level forecasts, uncertainty distributions, time patterns and performance of the comparative models.

5 Implementation

The forecasting system was implemented within two complementary Python environments. The entire ingestion of the data, the preprocessing of the data, spatial mapping, feature engineering, and the formation of the Unified TSO-Level Modelling dataset are done in a Jupyter Notebook environment. This arrangement allowed effective exploration, consecutive development, and interactive validation of every transformation step of data.

The training of the model, hyperparameter optimization, probabilistic forecasting, and evaluation were done in Google Colab. Colab offered a GPU-based execution environment that was appropriate to be used in training deep learning models, conducting Optuna optimization studies, and producing large-scale forecasting results. Colab notebooks were equipped with all the necessary libraries that were installed directly to enable reproducible execution.

Raw multi-source datasets are converted into model-ready tensors by the modelling pipeline, and the deep learning architectures are trained. Models were all run in PyTorch, and the Temporal Fusion Transformer (TFT) architecture was run with the library of pytorch-forecasting. The TFT base and the residual LSTM present in the Hybrid TFT-LSTM Residual model are custom PyTorch modules. Standard Python data science libraries were used for preprocessing, feature engineering, and spatial-socioeconomic mapping.

The processed data were then transformed into supervised learning samples. A sliding-window generator generated 168 hour look-back sequences and 24 hour ahead targets. This was also in line with the 60/20/20 chronological train/validation/test split in the methodology. All the inputs of the models were normalised and coded into multivariate tensors that are compatible with LSTM, TFT, and Hybrid TFT-LSTM Residual architectures.

The Adam optimiser was used with early stopping to train a model to prevent overfitting. The LSTM, TFT, and Hybrid TFT-LSTM Residual models were originally trained with a standard set of hyperparameters. Optuna was subsequently incorporated into the training process to optimise the most crucial hyperparameters, such as the learning rate, the hidden size, dropout, and the batch size in the case of TFT, attention heads, and gating network dimensions. Optuna was used to automate hyperparameter tuning, and the best-performing configurations were selected based on validation loss. Following this, constructed a Power BI dashboard using the CSV that is generated in Google Colab.

The implementation had the following main results:

- Created a TSO-level dataset at hourly resolution fully cleaned and harmonised (created in Jupyter Notebook).
- Trained baseline and tuned Optuna LSTM, TFT, and Hybrid TFT-LSTM Residual models.
- Forecast (at P10, P50, and P90) of a TFT-based model.
- Measure evaluation and comparison across models.
- Visualisation files that can be used in the Power BI dashboard.
- Power dashboard developed.

Operational value that is delivered by Power BI includes the ability of TSOs to interactively visualise forecast accuracy, peak-load risks, and temporal demand patterns. This dashboard enables operators to compare model outputs in real-time and track errors over time and find any deviations in system load swiftly to aid in making decisions more speedily and more informed.

All the experiments were implemented in a similar workflow in Jupyter Notebook and Google Colab. The preprocessing and training environments were separated to provide an organised implementation and distinct monitoring of every stage of the pipeline.

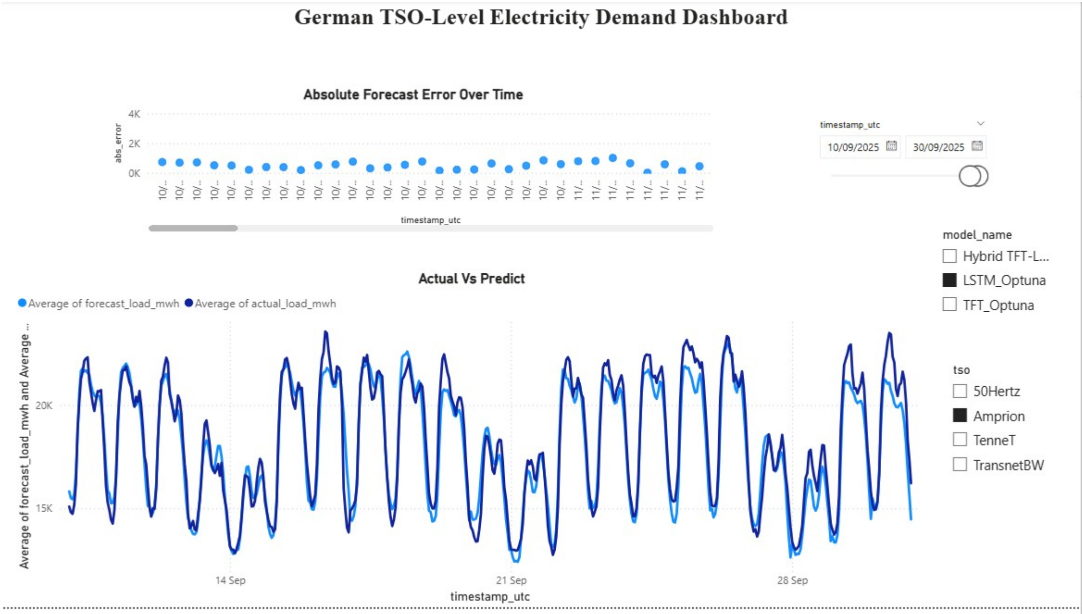

Figure 5: PowerBI Dashboard

6 Evaluation

This section provides a detailed analysis of the experimental findings and principal results of the research.

6.1 Experiment / Case Study 1: Baseline Models

This case study evaluates baseline deep learning models for forecasting German TSO-level load. Two architectures were compared: a Long Short-Term Memory (LSTM) network and a Temporal Fusion Transformer (TFT). These baseline models serve as reference points for the subsequent tuning experiments.

Table 4: Point Forecast Performance of Baseline Models (Test Set)

Model	**MAE (MW)**	**RMSE (MW)**	**MAPE (%)**	**MASE**	R^2
LSTM (Baseline)	1652	2229	12.44	1.538	0.86
TFT (Baseline)	185	246	1.35	0.172	0.9977

Results The baseline TFT model is substantially better than the LSTM baseline, with much lower MAE and MAPE and a higher R^2. These findings support prior work showing that transformer-based architectures are more effective than recurrent networks in capturing long-range temporal dependencies. The baseline TFT also represents the general shape of the load curve well, but it still underestimates some evening peaks, which is a limitation of the model.

6.2 Experiment / Case Study 2: Hyperparameter-Tuned Models

This case study examines the effect of hyperparameter optimization on the LSTM and TFT models using Optuna. The dataset was split chronologically into 60% training, 20% validation, and 20% test sets, and all models were trained with a look-back window of 168 hours and a forecast horizon of 24 hours.

Table 5: Point Forecast Performance After Optuna Tuning

Model	**MAE (MW)**	**RMSE (MW)**	**MAPE (%)**	**MASE**	R^2
LSTM (Optuna)	586	902	4.57	0.586	0.970
TFT (Optuna)	**176**	**235**	**1.24**	**0.164**	**0.9979**

Results The findings indicate that the hyperparameter tuning process significantly improves the performance of LSTMs, but TFT Optuna is more effective in general and reaches the state-of-the-art accuracy that is comparable to published TSO-level prediction research. These results uphold the supposition that attention processes are especially efficient in terms of capturing the associations between weather, calendar, and socioeconomic characteristics. Compared to the untuned TFT, the baseline hybrid (not tuned) is more effective in predicting the peaks. Nevertheless, when tuning is utilized, the hybrid comes to the same global performance to TFT Optuna, which means that the difference in error is insignificant.

6.3 Experiment / Case Study 3: Hybrid Residual Models

This analysis explores the hybrid architecture in which an LSTM is trained on the residual errors of the TFT model. Both a baseline hybrid and an Optuna-tuned hybrid were evaluated.

Table 6: Performance of Hybrid TFT-LSTM Residual Models

Model	**MAE (MW)**	**RMSE (MW)**	**MAPE (%)**	**MASE**	R^2
Hybrid (Baseline)	192	251	1.42	0.321	0.9976
Hybrid (Optuna)	**176**	**235**	**1.24**	**0.164**	**0.9979**

Results In this case study, uncertainty quantification, peak forecasting, early-warning capability, and efficiency are assessed.

6.4 Experiment / Case Study 4: Probabilistic & Business Metrics

This case study evaluates uncertainty quantification, peak forecasting, early-warning capability, and efficiency.

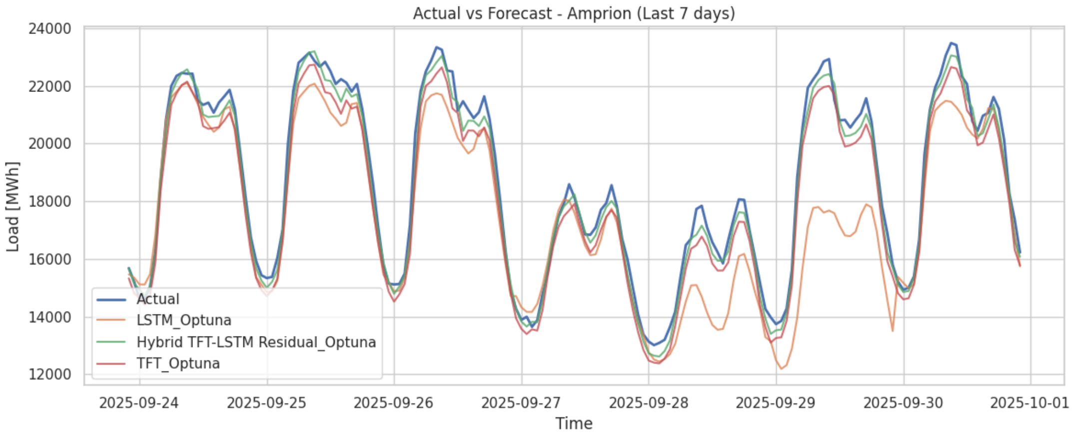

Figure 6: Actual Vs Forecast for Amprion

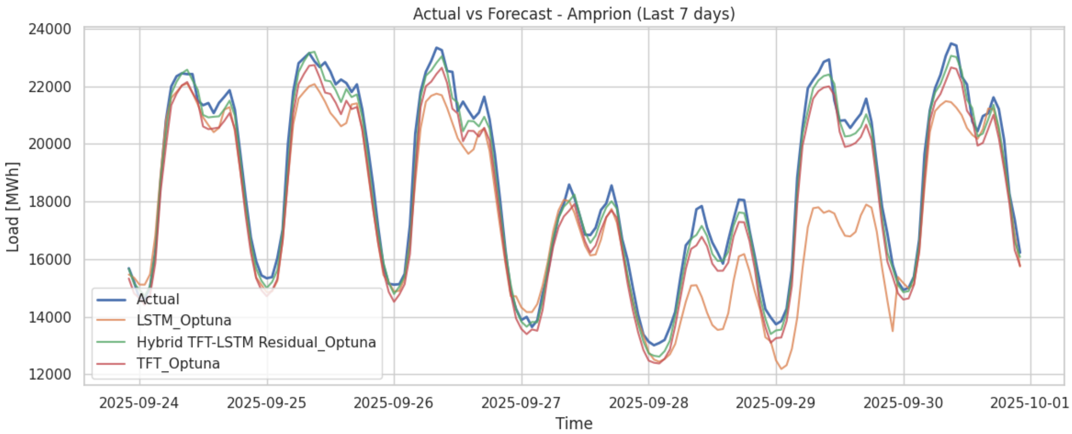

Figure 7: Actual Vs Forecast for 50Hertz

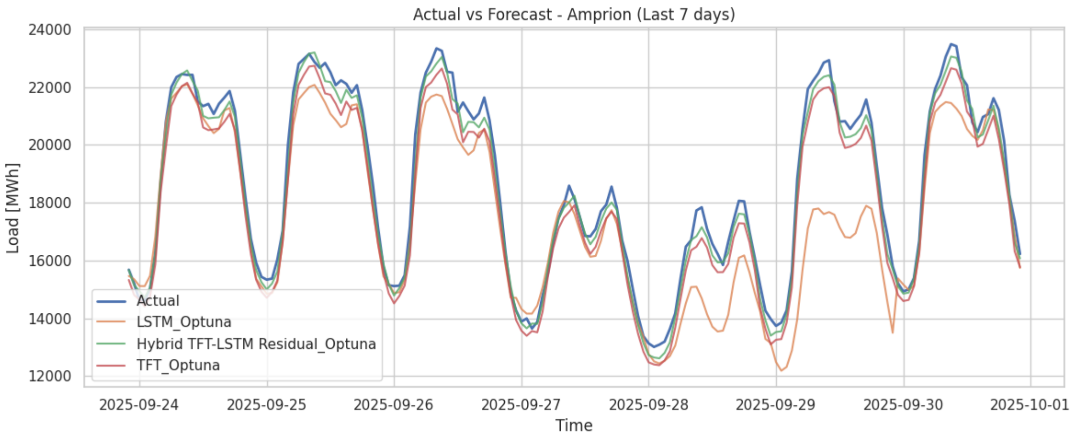

Figure 8: Actual Vs Forecast for TenneT

Results TFT-based models produce sharp and accurate probabilistic forecasts. The hybrid baseline improves peak detection, while LSTM Optuna is the only model with nominal coverage (0.80). TFT Optuna and Hybrid Optuna achieve almost perfect early-warning performance, which is critical for grid stability.

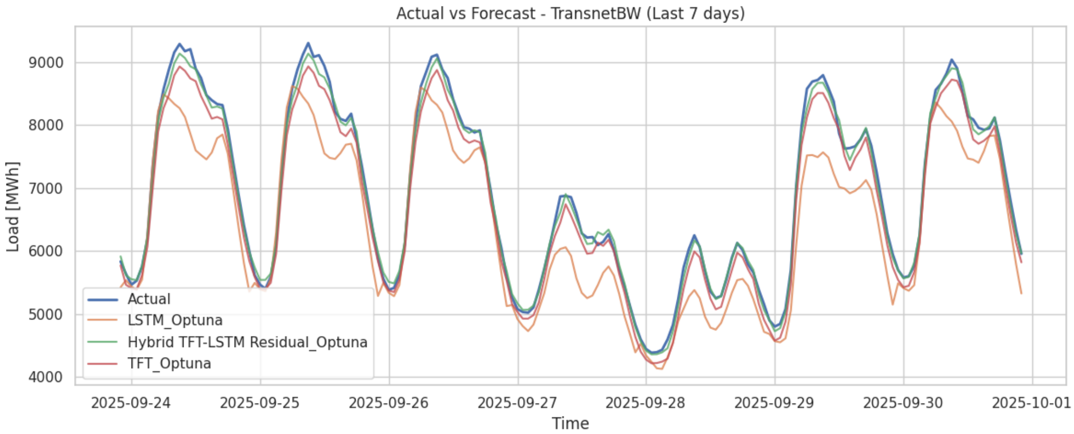

Figure 9: Actual Vs Forecast for TransnetBW

Table 7: Uncertainty and Business Metrics

Model	**P50 Loss**	**Coverage**	**MPIW**	**Peak MAE (MW)**	**F1 Score**
LSTM (Baseline)	1652	0.800	3316	2838	0.501
LSTM (Optuna)	586	**0.800**	1777	969	0.768
TFT (Baseline)	146	0.656	391	315	1.000
TFT (Optuna)	**142**	0.719	419	**265**	**0.999**
Hybrid (Baseline)	96	0.625	392	273	0.941
Hybrid (Optuna)	**142**	0.719	419	**265**	**0.999**

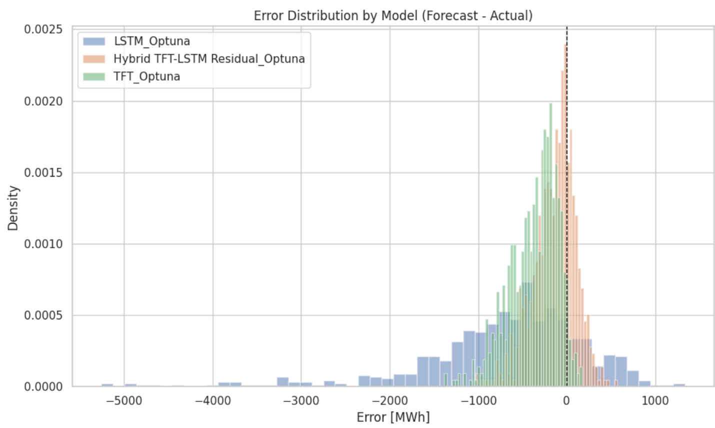

Figure 10: Residual error histograms

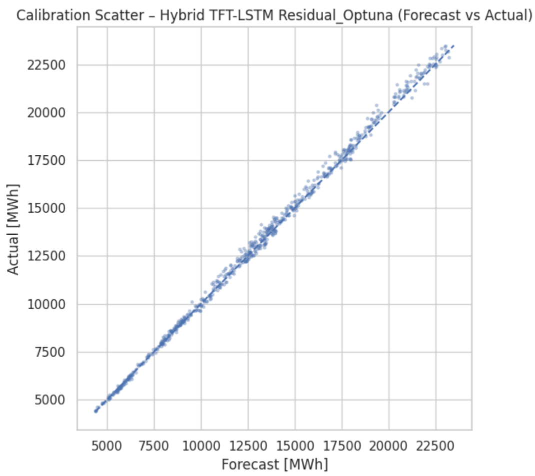

Figure 11: Calibration plot for Hybrid

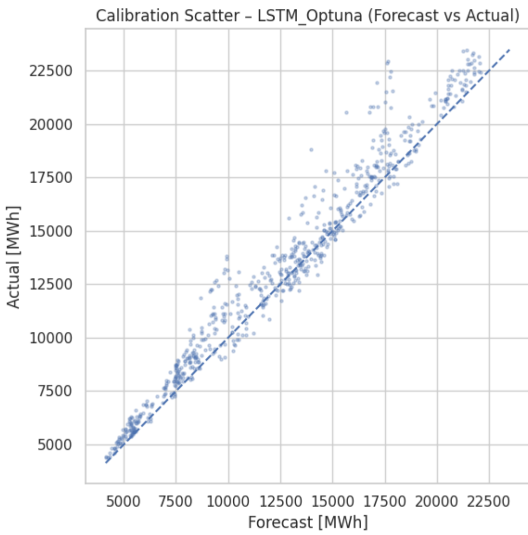

Figure 12: Calibration plot for LSTM

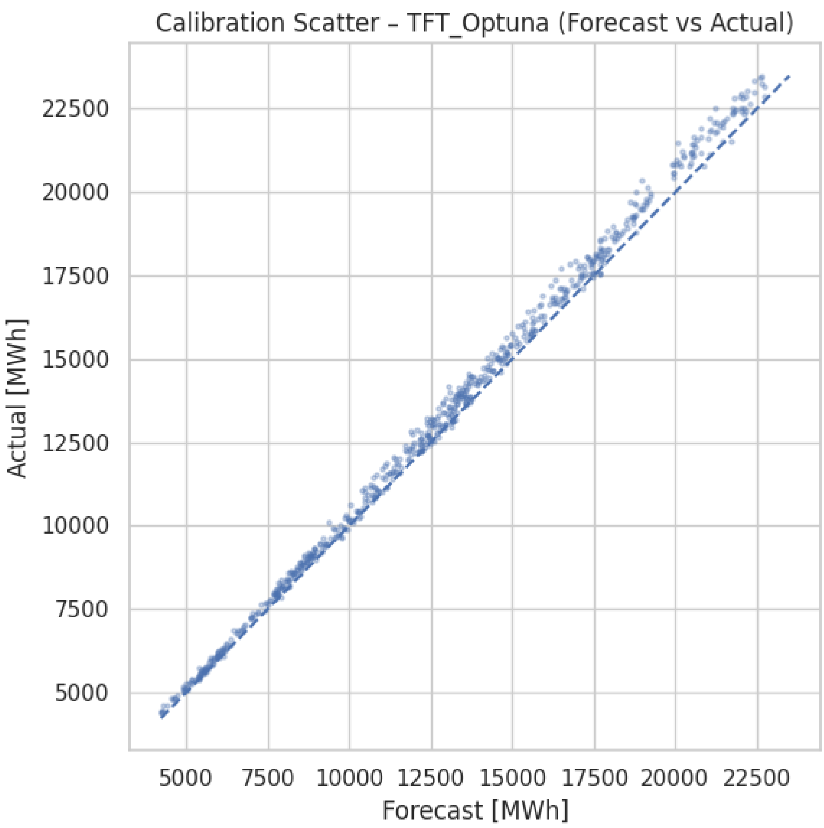

Figure 13: Calibration curves (reliability diagrams) for probabilistic forecasts of TFT-based models.

Table 8: Per-TSO Forecast Performance of Optuna-Tuned Models

TSO	Model	MAE (MW)	RMSE (MW)	R^2	MAPE (%)	Peak MAE (MW)	Peak RMSE (MW)
3*50Hertz	LSTM Optuna	681.26	927.44	0.7416	5.82	995.69	1156.55
	Hybrid TFT-LSTM Optuna	**162.94**	**214.20**	**0.9862**	**1.40**	**221.14**	**266.96**
	TFT Optuna	286.22	337.76	0.9658	2.45	411.30	450.87
3*Amprion	LSTM Optuna	793.12	1117.89	0.8772	4.19	1044.64	1287.37
	Hybrid TFT-LSTM Optuna	**295.74**	**380.62**	**0.9858**	**1.56**	**327.97**	**411.29**
	TFT Optuna	533.93	618.69	0.9624	2.81	659.05	733.61
3*TenneT	LSTM Optuna	605.21	945.85	0.8878	3.90	690.27	873.61
	Hybrid TFT-LSTM Optuna	**191.02**	**248.16**	**0.9923**	**1.24**	**158.01**	**207.68**
	TFT Optuna	334.62	392.57	0.9807	2.15	353.10	403.37
3*TransnetBW	LSTM Optuna	284.25	415.58	0.9052	4.39	394.09	475.93
	Hybrid TFT-LSTM Optuna	**116.87**	**156.86**	**0.9865**	**1.79**	**158.49**	**190.42**
	TFT Optuna	237.15	270.00	0.9600	3.66	314.23	336.38

Table 8 gives per-TSO performance of the Optuna-tuned models to complement the global evaluation metrics. The Hybrid TFT-LSTM is significantly superior in all four system operators, especially the accuracy of the peak-load.

6.5 Discussion

The present work aimed to determine the effectiveness with which three deep learning systems, including LSTM, Temporal Fusion Transformer (TFT), and a hybrid TFT-LSTM residual system, can predict short-term TSO-level power demand in Germany using a multi-source approach. The results show that there is a clear performance hierarchy of the models and offer information on the importance of multi-source features and hybrid residual learning.

With respect to Objective 1 (building multi-source datasets), the findings show that a combination of operational, meteorological, socioeconomic, and calendar characteristics provides a strong basis for predicting demand. The homogeneous performance of TSOs demonstrates that both the temporal variability and the regional structural variability are sufficiently captured by the dataset, which is also a strong indication that the modelling framework is effective.

As to Objective 2 (optimization and optimization of LSTM, TFT, and hybrid models), the experiments demonstrate that the Optuna-tuned TFT has the state-of-the-art point-forecast performance, which is consistent with the current results in transformer-based energy forecasting. However, the LSTM can be tuned to be significantly better and it still cannot model long-range and multi-factor interactions as well. The hybrid TFT-LSTM model performs effectively: it learns systematic patterns of underestimation in the TFT residuals and hence predicts peak loads better. This is a direct response to the requirement of having architectures that deal much better with extreme events, which is a serious operation requirement for TSOs.

On Objective 3 (model comparison on various metrics), the analysis establishes that transformer-based models perform better in terms of point-forecast performance and early-warning, and the hybrid architecture provides additional value when it comes to peak-demand problems. TFT-based models are generally under-calibrated in uncertainty intervals; nevertheless, their probabilistic sharpness of performance, combined with early-warning, is an operational benefit.

In relation to Objective 4 (critical evaluation of implications and limitations), findings indicate that there are a number of important considerations. First the attention-based architectures are most effective at the feature interaction modelling, however it underestimates the peak events unless additional residual learning is added- hence the hybrid design. Second, the fact that probabilistic intervals are slightly under-covered provides a chance to develop better uncertainty modelling, including the Bayesian and ensemble-based methods. Third, there is a difference in the performance of TSO regions, which points to the possibility of adding additional spatial or behavioural features that would make the model more robust.

Lastly, Objective 5 (development of operational insight Power BI dashboard) is supported by the study, which shows that forecast results can be converted into feasible visual aids. The dashboard would allow operators to interactively explore the model behaviour, peak-risk patterns, and uncertainty and improve their interpretability and functional application in the real world.

Altogether, the results prove that transformer-based and hybrid residual architectures are well adapted to peak-sensitive multi-source TSO-level forecasting. The enhanced argumentation and objective-associated insights highlight the generalizability of the work to the operational forecasting settings.

7 Conclusion and Future Work

This study examined the problem of short-term electricity demand forecasting for German TSOs with the help of multi-source data and advanced deep learning architectures. The study prepared a solid background through harmonisation of a dataset comprising operational, weather, socioeconomic and calendar variables to evaluate various modelling strategies. Three architectures, LSTM, TFT and an integrative TFT-LSTM residual model were created and optimised to offer a detailed comparison regarding forecasting behaviour with point-forecast, probabilistic and operational measures.

Optuna-tuned TFT was found to be the most powerful overall point-forecasting model, and the hybrid structure was found to have obvious benefits in correcting a significant issue of underestimation of peak-load, a major operational issue faced by TSOs. Such results confirm the robustness of the use of transformers to multi-source forecasting and depict the fact that the use of residual learning can improve the performance in extreme-demand scenarios. The interactive Power BI dashboard is another feature that is beneficial to the decision-making processes of the operations department, as it allows exploring the forecasts and patterns of uncertainty in a transparent way.

Though the models have worked well, there are still limitations in calibration of probabilistic intervals and depiction of rare peak events. These problems imply the possibilities for the future work, such as the Bayesian estimation of uncertainty, regime-specific modelling strategies, and the integration of more specific spatial or behavioural data. It would also be useful to test the robustness of its operation with the extension of the framework to the real-time deployment.

Altogether, the research shows that the integration of multi-source feature engineering and the contemporary deep learning models are realistic and efficient in TSO-level load prediction, adding to the knowledge of the methodology and practical usefulness.

7.1 Future Work

Future research could involve exploring the shortcomings of this research and the modelling framework to more operational settings. To begin with, the calibration of the uncertainty should be enhanced which is also a good direction. Bayesian deep learning, quantile ensembling, distribution-aware loss functions might be useful to fix the under-dispersed prediction intervals noted in the TFT-based models. Second, regime-switching architectures, extreme-value-sensitive loss functions, or the use of more refined meteorological and generation data may be useful in the modelling of extreme peak events. Third, by applying the framework to spatiotemporal architectures, smaller TSOs could better achieve their performance, in which the region-based heterogeneity would be more powerful. They may also be more responsive by adding behavioural or demand-side variables. Lastly, the models would be deployed in a real-time operational setting so that one can constantly monitor the performance, test the early-warning utility, and determine the long-term stability of the models. These additions would enhance even more the empirical usefulness and effectiveness of deep learning based models on TSO-level predictions.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T. and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework, *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631.

Behm, C., Nolting, L. and Praktiknjo, A. (2020). How to model european electricity load profiles using artificial neural networks, *Applied Energy* **277**: 115564.

Chen, K., Chen, K., Wang, Q., He, Z., Hu, J. and He, J. (2018). Short-term load forecasting with deep residual networks, *IEEE Transactions on Smart Grid* **10**(4): 3943–3952.

Gao, X., Li, X., Zhao, B., Ji, W., Jing, X. and He, Y. (2019). Short-term electricity load forecasting model based on emd-gru with feature selection, *Energies* **12**(6): 1140.

Giacomazzi, E., Haag, F. and Hopf, K. (2023). Short-term electricity load forecasting using the temporal fusion transformer: Effect of grid hierarchies and data sources, *Proceedings of the 14th ACM international conference on future energy systems*, pp. 353–360.

Guo, Y., Du, Y., Wang, P., Tian, X., Xu, Z., Yang, F., Chen, L. and Wan, J. (2024). A hybrid forecasting method considering the long-term dependence of day-ahead electricity price series, *Electric Power Systems Research* **235**: 110841.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. and Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*, OTexts.

Khan, S., Aslam, S., Mustafa, I. and Aslam, S. (2021). Short-term electricity price forecasting by employing ensemble empirical mode decomposition and extreme learning machine, *Forecasting* **3**(3): 28.

Lago, J., De Ridder, F. and De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms, *Applied Energy* **221**: 386–405.

Lu, W. and Chen, X. (2024). Short-term load forecasting for power systems with high-penetration renewables based on multivariate data slicing transformer neural network, *Frontiers in Energy Research* **12**: 1355222.

Maciejowska, K., Nitka, W. and Weron, T. (2021). Enhancing load, wind and solar generation for day-ahead forecasting of electricity prices, *Energy Economics* **99**: 105273.

Marino, D. L., Amarasinghe, K. and Manic, M. (2016). Building energy load forecasting using deep neural networks, *IECON 2016-42nd annual conference of the IEEE industrial electronics society*, IEEE, pp. 7046–7051.

Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Oueslati, F. S. and Abu-Rub, H. (2021). A novel stacked generalization ensemble-based hybrid lgbm-xgb-mlp model for short-term load forecasting, *Energy* **214**: 118874.

Moesmann, M., Khanal, S. and Pedersen, T. B. (2025). Energy-efficient short-term load forecasting for responsive demand-side analytics for the european electricity market, *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, pp. 177–190.

Nabavi, S. A., Motlagh, N. H., Zaidan, M. A., Aslani, A. and Zakeri, B. (2021). Deep learning in energy modeling: Application in smart buildings with distributed energy generation, *IEEE Access* **9**: 125439–125461.

Narajewski, M. (2022). Probabilistic forecasting of german electricity imbalance prices, *Energies* **15**(14): 4976.

Ouyang, T., He, Y., Li, H., Sun, Z. and Baek, S. (2019). Modeling and forecasting short-term power load with copula model and deep belief network, *IEEE Transactions on Emerging Topics in Computational Intelligence* **3**(2): 127–136.

Qin, B., Gao, X., Ding, T., Li, F., Liu, D., Zhang, Z. and Huang, R. (2024). A hybrid deep learning model for short-term load forecasting of distribution networks integrating the channel attention mechanism, *IET Generation, Transmission & Distribution* **18**(9): 1770–1784.

Raza, M. Q. and Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings, *Renewable and Sustainable Energy Reviews* **50**: 1352–1372.

Snoek, J., Larochelle, H. and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms, *Advances in neural information processing systems* **25**.

Wan, R., Mei, S., Wang, J., Liu, M. and Yang, F. (2019). Multivariate temporal convolutional network: A deep neural networks approach for multivariate time series forecasting, *Electronics* **8**(8): 876.

Wu, H., Xu, J., Wang, J. and Long, M. (2021). Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting, *Advances in neural information processing systems* **34**: 22419–22430.

Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H. and Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting, *Proceedings of the AAAI conference on artificial intelligence*, Vol. 35, pp. 11106–11115.

Ziel, F. and Weron, R. (2018). Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks, *Energy Economics* **70**: 396–420.