Multivariate Time Series Forecasting of Abortion Incidence Rate in New Zealand Livestock Populations Using Deep Learning Models

A thesis presented in partial fulfilment of the requirements
for the degree of

Master of Information Science

in

Computer Science

at Massey University, Albany,
New Zealand.

Xiaotong Zhao

2025

Abstract

Livestock abortion counts are important indicators of population health and productivity. In practice, however, obtaining timely and accurate abortion incidence data and producing reliable forecasts remain challenging due to factors such as reporting delays, incomplete records, and substantial observational noise. Abortion dynamics are influenced by multiple interacting factors, including production history, seasonal variability, climatic conditions, and socio-behavioural effects. As a result, the monthly abortion time series exhibits both long-term trends and short-term fluctuations that reflect complex underlying processes. Modelling such heterogeneous, multi-source time-series data while maintaining appropriate control over model complexity remains a central challenge in applied forecasting research.

This study focuses on livestock abortion incidence data in New Zealand and applies time-series forecasting at both national and regional scales. Multiple data sources are incorporated, including historical abortion records, livestock population statistics, climatic variables, and Google Trends search information. These data are aggregated at a monthly resolution to construct a unified time series for forecasting.

In the national-level analysis, two time-series forecasting models are considered to address the characteristics of multivariate inputs. One model builds on the standard Long Short-Term Memory (LSTM) architecture by incorporating an attention mechanism, whereas the other employs a Mixing-Channel Patch Time Series Transformer (PatchTST) framework with a parallel LSTM branch. These models are evaluated with respect to their architectural design and predictive performance, and their results are compared with those obtained from other baseline deep learning approaches.

For the regional-level analysis, New Zealand's 16 administrative regions are grouped into seven study areas based on the livestock demography and geographical adjacency. Both model configurations are subsequently applied to each region, and differences in predictive performance are analysed alongside regional patterns in livestock abortion counts.

Overall, in national-level prediction tasks, the proposed Mixing-Channel PatchTST parallel LSTM framework achieved superior results compared to several well-established baseline models (RMSE=6.0062, MAE=4.6993, R^2=0.9784). Compared with the strongest baseline model (LSTM with attention mechanism; RMSE=7.4426, MAE=6.0720, R^2=0.9668), this method achieved a 19.3% relative reduction in RMSE, a 22.6% relative reduction in MAE, and an absolute increase of 0.0116 in R^2. Further experiments at the regional level showed that the framework achieved varying levels of error reduction in most regions, demonstrating more robust predictive behaviour and a more consistent capability in trend characterisation. In conclusion, the model proposed in this thesis provides a robust and effective modelling framework for predicting livestock abortion numbers and can provide practical technical support for pasture management and risk-related decision analysis.

Acknowledgements

This research was partially funded by the Massey University Foundation's Morris Trust, the Massey University MURF Foundation, and the SREF Foundation. Data was provided by Jonathan Watts of MPI, NIWA, and the Google Trends data team.

I sincerely thank my supervisors, Dr Bo Liu and Dr Masako Wada, for their continuous guidance, encouragement, and valuable feedback throughout the research process. I also thank my family and friends for their support and understanding during my studies at Massey University.

Table of Contents

List of Figures

List of Tables

1. Introduction

This chapter introduces the background and motivation for the research on time-series prediction of livestock abortion, and also provides the structural framework of the thesis to facilitate readers' understanding of subsequent chapters.

1.1 Research Background

New Zealand is a major agricultural country with livestock farming as its foundation, and livestock farming has long occupied a core position in its national economy and export system. According to relevant reports and official market-overview data, agriculture and related primary industries account for a significant share of New Zealand's GDP, and their export revenue plays a substantial supporting role in the country's overall economy. Dairy products are among the most representative export products [1,2]. Thanks to its temperate maritime climate and relatively well-developed pasture management system, New Zealand has developed into a major global dairy producer and exporter. Its dairy industry not only serves the international market but is also deeply embedded in the domestic agricultural production system and regional economic structure.

Within the livestock industry, the dairy industry has formed a highly integrated industrial chain across production organisation, processing, distribution, and export trade, with long-term implications for farm operational stability, regional employment, and related supporting industries. In recent years, fluctuations in dairy product prices and changes in production costs have also triggered widespread social attention to the sustainable development of the livestock industry, further highlighting the real importance of this industry in New Zealand's economy and society [3]. Therefore, ensuring the stable operation of the livestock production process is of great significance for maintaining the healthy development of agricultural output, farm income, and the export system.

In livestock production, reproductive health is a key factor affecting production efficiency and economic benefits. Livestock abortion not only directly impacts the health of individual animals but can also lead to decreased reproductive rates, shortened milk production cycles, and limited subsequent production capacity, resulting in sustained economic losses for farms. For dairy farms primarily engaged in large-scale farming, the cumulative effect of abortion events can significantly impact overall production efficiency and operational stability. Therefore, a systematic analysis and research on livestock abortion from the perspectives of production management and risk control has clear practical significance and real-world relevance.

1.2 Research Motivations

In practical livestock production and management, attention to livestock abortion has long extended beyond post-abortion recording and statistical reporting. Instead, it has gradually shifted toward data-driven analysis and risk identification based on historical records to anticipate potential risks and support management decision-making. As the scale and complexity of livestock farming continue to increase, farm managers show a growing demand for analytical tools and predictive approaches capable of identifying abortion risks and characterising temporal trends. Under these conditions, systematic data analysis and modelling of livestock abortion counts are strongly driven by practical management requirements.

In earlier studies, analyses of livestock abortion have predominantly relied on traditional statistical modelling approaches. These methods are generally characterised by relatively strong interpretability and are well-suited to examining relationships between individual factors or a limited number of variables and abortion outcomes. However, in real production environments, abortion events are often influenced by the combined effects of multiple interacting factors, including individual physiological status, genetic characteristics, disease risks, as well as external influences such as climatic conditions and management practices. Under the interaction of these factors, temporal variations in abortion counts frequently exhibit pronounced non-linear patterns and seasonal fluctuations, which constrain the capacity of traditional methods to capture complex temporal dependency structures adequately [4].

With the development of artificial intelligence techniques, machine learning and deep learning methods have increasingly been introduced into livestock health-related research and have demonstrated promising modelling capabilities across a range of applications. For example, previous studies have explored the use of genetic information and machine learning-based models to analyse abortion risks and associated disease factors, highlighting the potential of data-driven approaches for analysing complex biological systems [4,5]. However, compared with well-established time series forecasting domains such as electricity load and financial markets, research on the number of livestock abortions remains relatively limited, and existing work has primarily focused on individual-level analysis, specific diseases, or static risk assessment, with insufficient emphasis on group-level abortion sequences exhibiting clear temporal structure.

From the perspective of data characteristics, livestock abortion records usually have strong time series attributes, and their changes are often accompanied by seasonal fluctuations and long-term trends. Simultaneously, climate factors may affect livestock reproductive health through lagged effects, and background information, such as changes in population size, may also indirectly affect abortion numbers. These characteristics make it difficult to comprehensively characterise the dynamic changes in abortion numbers using a single data source or a single modelling approach. Therefore, it is necessary to explore modelling methods that integrate multiple sources of information and simultaneously characterise both long-term trends and short-term fluctuations to improve the stability and reliability of predicting abortion numbers.

1.3 Research Questions

Drawing on the research background and motivation presented above, this thesis investigates the problem of forecasting the number of livestock abortions using multivariate time-series data. The study focuses on three interrelated research questions.

RQ1: Given multivariate time-series data (including historical livestock abortion records, livestock population information, climate variables, and Google Trends search activity), how can a reliable and robust predictive model be constructed to effectively capture long-term trends and seasonal patterns in the time series and deliver accurate forecasts of monthly livestock abortion numbers?

RQ2: In national-level livestock abortion prediction tasks, how can multi-source, heterogeneous, and easily accessible auxiliary information be effectively integrated to further enhance the model's predictive accuracy and forecasting stability?

RQ3: To further assess the predictive stability of the model in complex real-world settings, this thesis extends the research to the regional level. Faced with pronounced regional heterogeneity in livestock abortion data, can the constructed model still effectively learn underlying temporal dynamics and maintain consistently stable predictive performance and strong generalisation capability across different regions?

1.4 Significance of the Research

This thesis explores the modelling and analysis of livestock abortion prediction using multivariate time-series data, with relevance to both methodological development and applied forecasting practice.

At the modelling level, the study examines how different deep learning architectures perform when applied to livestock abortion prediction tasks. An attention-enhanced LSTM model is analysed alongside a hybrid framework that combines a Mixing-Channel Patch Time Series Transformer (PatchTST) structure with a parallel Long Short-Term Memory (LSTM) branch. Through an examination of model architecture choices, training convergence behaviour, and predictive performance, the thesis investigates how hybrid structural design influences forecasting outcomes in a multivariate setting.

The experimental analysis extends beyond aggregate performance metrics. By comparing results across multiple baseline models and conducting ablation studies, input-variable sensitivity analysis, and fusion-parameter experiments, the study provides a detailed examination of model behaviour under different configurations. These experiments contribute to a clearer understanding of model stability and the role of individual components in shaping predictive performance.

From an applied perspective, the analysis is further extended to regional-level forecasting of livestock abortion numbers. Differences in model performance across regions with heterogeneous data distributions are examined to assess the impact of regional characteristics on predictive outcomes. This regional analysis situates the modelling results in a practical context and informs the discussion of model applicability and limitations for livestock management and abortion risk assessment.

1.5 Structure of the Thesis

Based on the above discussion of the problem background and research motivation, the objectives of conducting an in-depth investigation into multivariate time-series data for livestock abortion prediction are clearly defined. Accordingly, the structure and content of this thesis are organised as follows.

Chapter 2 provides a systematic review of existing studies on the application of traditional statistical methods, machine learning algorithms, and deep learning approaches to time-series forecasting and livestock health research, and summarises the main limitations of current work. Chapter 3 introduces the research dataset and data preprocessing procedures, including data sources, data characteristics, dataset partitioning rules, and interpolation methods. Chapter 4 focuses on the methodology, experimental results, and analysis of the Attention-Enhanced LSTM model. Chapter 5 presents the Mixing-Channel PatchTST with Parallel LSTM model, together with its experimental results and corresponding analysis. Chapter 6 applies the two modelling approaches to regional-level data and compares and discusses their predictive effectiveness across regions. Finally, Chapter 7 summarises the main research conclusions and outlines potential directions for future work.

2. Literature Review

This chapter reviews the existing literature on livestock abortion prediction and related time-series modelling, summarising the key concepts and methodological approaches of prior research in this field and establishing a theoretical foundation and empirical basis for subsequent model construction and experimental design. Building upon the research motivation presented in Chapter 1, this chapter focuses on the applicability and limitations of different modelling approaches when handling livestock health-related time series data.

Specifically, this chapter first surveys the typical applications of traditional statistical methods in time series analysis and livestock reproduction and health research, and evaluates their strengths and limitations in characterising long-term dependencies, nonlinear features and seasonal variations. Then, it systematically reviews the development of machine learning and deep learning approaches for time series forecasting, with particular emphasis on their application to livestock health surveillance, risk assessment, and related predictive tasks. Finally, based on a comprehensive comparison of existing research findings, it summarises the remaining shortcomings in current research, thereby clarifying the entry point for this study in terms of methodological selection and research perspective.

2.1 Traditional Statistical Methods for Livestock-Related Time Series

In early studies on livestock abortion prediction and animal epidemiological analysis, researchers mainly used traditional statistical modelling methods to analyse abortion-related influencing factors. These methods are usually based on clear statistical assumptions and construct regression models or survival analysis models to characterise the relationship between variables such as physiological characteristics, environmental factors, and management conditions and the risk of abortion, thereby providing a theoretical basis for livestock risk control and management decisions [6,7,9].

Within this research framework, several studies have conducted empirical analyses of livestock abortion risk based on traditional statistical models. For example, Rafati et al. [6] employed a logistic regression model to estimate the probability of abortion (PPA) and combined it with an accelerated failure time (AFT) model for survival analysis. Their results showed that the probability of abortion decreases with the increase of open days and advancing stages of pregnancy, while a history of abortion significantly elevates the risk of subsequent abortion. Thurmond et al. [7] proposed a hierarchical Bayesian logistic survival model to model the relative risk and prediction probability of abortion in dairy cows, and introduced an epidemiological modelling related to fetal health to verify the effectiveness of the method in improving abortion prediction ability and characterising fetal survival rates. In addition, Mallam et al. [8] applied multivariate logistic regression and decision tree approaches to analyse the influencing factors of abortion and stillbirth in Yankasa sheep flocks and found that seasonal variation exerts a significant effect on the abortion rate. Overall, these studies show that traditional statistical methods provide strong interpretability in risk modelling and factor analysis, and continue to process both research relevance and practical value in specific application contexts.

However, in actual livestock production environments, livestock abortion is often affected by multiple interacting factors, including individual physiological conditions, climatic variability, and human management interventions, and its temporal evolution typically exhibits pronounced nonlinear behaviour. In particular, when handling multi-source, high-dimensional time series data, traditional statistical models often suffer from insufficient prediction accuracy and modelling flexibility in characterising long-term temporal dependencies, seasonal fluctuations, and complex inter-variable interactions[10]. Against this backdrop, researchers have gradually explored the application of machine learning and deep learning methods for livestock abortion prediction to better capture complex temporal structures and improve predictive performance.

2.2 Machine Learning Methods for Livestock-Related Time Series

With the continuous improvement of the informatisation and digitalisation of animal husbandry, the data sources related to livestock health and reproduction have gradually expanded from single statistical records to a multi-source data system covering environmental factors, behavioural monitoring and management information. Against this background, the limitations of traditional statistical models in handling complex nonlinear relationships, multi-factor interactions and high-dimensional features have become increasingly apparent, making it difficult to fully characterise the dynamic formation mechanism of the complex process of livestock abortion. To make up for the above shortcomings, researchers have gradually introduced machine learning methods to improve the predictive ability and adaptability of models in complex scenarios through data-driven modelling [10].

Compared to traditional statistical methods, machine learning models typically do not depend on strict linear assumptions and can automatically learn the mapping relationship between input features and prediction targets from training data. In livestock health-related research, methods such as decision trees, random forests, support vector machines, and XGBoost have been widely applied to classification and prediction tasks. These approaches demonstrate notable advantages in handling nonlinear relationships and high-dimensional variables, and can provide a certain degree of interpretability through feature importance or variable contribution analysis [10,11]. In the field of livestock reproductive risk prediction, existing studies have integrated behavioural surveillance data, individual characteristics and environmental information, and used models such as random forests to predict pregnancy status or identify risk factors associated with abortion and perinatal mortality, verifying the application potential of machine learning methods in complex risk factor modelling [12,13]. In addition, related studies have also shown that machine learning has certain value in revealing the potential biological mechanisms in the pregnancy process and supporting actual production management decisions [14,15].

Although machine learning methods have demonstrated strong nonlinear modelling capabilities in predicting livestock abortion and related health risks, they still have significant limitations in time series modelling. Traditional machine learning models usually rely on artificially constructed features, introducing time information into the model as static variables, which makes it difficult to directly characterise long-term time dependencies and complex seasonal structures. In multi-source, high-dimensional time series scenarios, this feature engineering-driven modelling approach often limits the model's ability to characterise the overall dynamic evolution process, and its predictive stability and generalisation performance are also affected [10]. Based on the above shortcomings, researchers have gradually shifted their focus to deep learning methods in order to more effectively model complex time structures and uncover potential patterns in multi-source time series data.

2.3 Deep Learning Methods for Livestock-Related Time Series

In recent years, with the rapid development of deep learning technology, neural network-based models have been gradually introduced into the field of time series prediction to improve the ability to model complex and high-dimensional data. Compared with traditional machine learning methods, deep learning models can automatically learn feature representations from raw data through multi-layer nonlinear structures, which reduces the dependence on manual feature engineering to a certain extent. This characteristic makes them show obvious advantages in processing time series data with long-term time dependence, multi-scale changes and significant seasonality, and provides a new research direction for modelling complex time series problems [10,11].

In time series modelling, deep learning methods usually start with RNN and its improved structures. Among them, Long Short-Term Memory (LSTM) alleviates the gradient vanishing problem through a gating mechanism, enabling it to capture the sequence change pattern over a longer time span and has gradually become a relatively mature benchmark model [16]. Based on this, attention mechanisms and structures such as the Transformer have been introduced into time series prediction to more effectively model long-term dependencies and complex dynamic changes. Related studies have shown that they have good stability and generalisation ability in multivariate and long-term prediction tasks [19,20]. Deep learning has also been explored in application scenarios such as livestock health and reproductive risk prediction. For example, it has been used for tasks such as predicting the risk of disease outbreaks in dairy cattle, modelling abortion pathogen transmission burden, and predicting and interpreting pregnancy loss, thereby demonstrating its application potential for characterising complex behavioural patterns and time-dependent structures [17,18,21].

Overall, deep learning methods offer more expressive modelling tools for characterising complex nonlinear relationships and long-term time dependencies in time series forecasting, demonstrating significant advantages in multi-factor driven problems such as livestock abortion prediction. However, in practical applications, different deep learning models typically focus on time dependency modelling, feature extraction, or local dynamic characterisation, and their modelling capabilities are still constrained by structural design and task emphasis. When faced with time series systems containing multivariate information, variations across different time scales, and complex interaction mechanisms, a single model often struggles to simultaneously achieve prediction accuracy, stability, and robustness. Therefore, how to more comprehensively characterise multivariate dependencies and the overall evolutionary features of the system, while preserving the advantages of deep learning approaches, has become an open issue that warrants further investigation in current research.

2.4 Multivariate and Hybrid Modelling Strategies

In real-world applications, time series systems are rarely driven by a single factor, but are instead simultaneously influenced by multiple interacting variables. For such complex dynamic systems, a single model finds it difficult to achieve prediction accuracy, stability, and robustness at the same time when characterising the intrinsic relationships among multiple factors and their combined effects on the target variable. Especially in situations involving multivariate information, multi-scale temporal variations, and complex interaction mechanisms, the representational capacity of a single modelling method is often insufficient. Therefore, researchers have increasingly turned to multivariate modelling and hybrid modelling strategies to more comprehensively capture the underlying evolutionary processes of complex systems.

The core idea of hybrid models is not merely to splice or combine multiple models, but to adopt a structured design in which different models or different modelling mechanisms are assigned distinct roles, thereby compensating for the limited representational capacity of single methods. Early studies have investigated hybrid modelling strategies that combine statistical models with neural networks, for example, employing statistical models to capture linear trends and seasonal components in time series, followed by using neural networks to model the nonlinear residuals [22]. Such hybrid modelling ideas have also been widely applied in biomedical and animal science research, including tasks such as miscarriage risk prediction, livestock behaviour and movement modelling, disease and mortality analysis, and individual identification [23-26]. Existing findings indicate that, by integrating the complementary strengths of different models in temporal structure representation and local dynamic characterisation, hybrid models can generally achieve more stable predictive performance in complex application scenarios.

Although multivariate and hybrid modelling approaches provide important tools for forecasting complex time series, existing research still exhibits certain limitations with respect to modelling objectives and information integration. On the one hand, some studies place greater emphasis on specific event detection or risk identification, while paying insufficient attention to the direct modelling of the comprehensive time series of abortion numbers. On the other hand, in the use of multi-source information, different types of variables are often treated separately, lacking a systematic modelling approach that simultaneously characterises long-term structure and short-term fluctuations within a unified time series framework. These shortcomings, to some extent, limit the comprehensive characterisation of livestock abortion patterns in real-world production scenarios, but also provide a clear entry point for further research into time series forecasting methods that integrate multi-source information.

2.5 Summary and Research Gap

In summary, existing research has analysed livestock abortion and related health issues from multiple perspectives and has gradually introduced environmental factors, behavioural monitoring information, and other auxiliary data for risk identification and predictive modelling. Traditional statistical methods, machine learning models, and deep learning combined with hybrid modelling strategies have all demonstrated certain advantages in different research scenarios, providing valuable references for understanding the mechanisms of abortion and improving predictive capabilities. However, these studies still exhibit a relatively fragmented research characteristic in terms of modelling objectives, data integration methods, and temporal structure characterisation.

From the perspective of time series modelling, the number of livestock abortions, as a comprehensive group-level time series, has not yet received sufficient attention. Existing work mainly focuses on the detection of abortion events, analysis of specific diseases, or static risk assessment, with limited systematic modelling of the time structure of abortion numbers in terms of long-term trends, seasonal variations, and short-term fluctuations. Furthermore, in the use of multi-source information, variables such as climate conditions, search behaviour indicators, and livestock population structure are often treated in isolation, lacking research that simultaneously characterises the synergistic effects of multi-source auxiliary information and temporal dynamics within a unified time series framework. Therefore, it is necessary to further explore a prediction method that can integrate historical abortion time series with various auxiliary information, so as to provide a more systematic research foundation for subsequent model construction and experimental analysis.

3. Data Sources and Preprocessing

This chapter systematically introduces the data sources and corresponding data preprocessing procedures used in livestock abortion prediction research. Given that this study involves multiple sources of information, including official abortion records, climate data, livestock population statistics, and search behaviour indicators, and considering the differences in temporal resolution, spatial scale, and statistical calibre among these data, it is necessary to standardise and align the various data types temporally before model construction. Accordingly, this chapter focuses on interpreting data sources and characteristics, data integration and preprocessing workflows, feature selection strategies, and dataset partitioning principles, providing a standardised data foundation for subsequent model training and experimental analysis.

3.1 Overview of Multi-Source Data

This section details the data sources utilised in the multi-source dataset for this study. Specifically, the dataset includes official livestock abortion records released by New Zealand's Ministry for Primary Industries (MPI), together with climate-related data associated with abortion variability, Google Trends search behaviour data, and livestock population statistics.

3.1.1 MPI abortion records

This study aims to predict the number of livestock abortions. Accordingly, livestock disease diagnosis records provided by the New Zealand Ministry for Primary Industries (MPI) constitute the core data source of this research and serve as the target variable for all modelling tasks. The MPI is the national authority responsible for animal health surveillance and disease record management across the country [27]. Its data is an important part of New Zealand’s livestock disease surveillance system and has high authority and real-world representativeness. The data covers three major livestock species: cattle, horses, and sheep, and provides a comprehensive reflection of the abortion-related conditions within New Zealand's livestock production system.

The MPI abortion-related records span from January 2004 to December 2023, covering 16 administrative regions across New Zealand and exhibiting a long continuous time series structure. It should be noted that this type of data is typical passive surveillance data, which is mainly collected based on test samples voluntarily submitted by farmers or veterinarians, rather than through systematic or designed sampling procedures. As a result, the original data may be imbalanced and biased across regions and time periods, and certain areas contain missing records during specific intervals. In addition, as data collection standards have evolved over time, earlier records are less consistent in terms of format and field completeness, whereas more recent data are comparatively standardised and detailed.

To ensure alignment with the research objectives and reduce noise interference, this thesis only retains cases in which "abortion" is clearly identified as the diagnostic result, constructing a dedicated time series of abortion cases. By screening for etiological categories and data cleaning, the interference caused by the mixing of heterogeneous disease types can be effectively reduced, which enhances the stability of the target variable in time series modelling. Following the above processing, a total of 411,039 original records were finally retained for subsequent analysis. Although the MPI abortion records do not completely capture all abortion events occurring in practice, their long-term continuity and nationwide coverage render them highly valuable for characterising seasonal variations, long-term trends, and relative fluctuations in abortion counts, providing a reliable data foundation for constructing and evaluating time series forecasting models.

3.1.2 Climate data sources

The climate data used in this study were obtained from the National Institute of Water and Atmospheric Research (NIWA) of New Zealand and were processed and curated by EpiCentre under a data-sharing agreement to support research purposes [28]. The dataset covers major climate variables across New Zealand and provides contextual background information for analysing the potential association between climate conditions and variations in livestock abortion numbers. Considering that climate factors may have indirect or direct effects on livestock reproductive health by influencing animal stress levels, forage growth conditions, and pathogen transmission environments, incorporating climate information into the prediction model helps to characterise seasonal fluctuations and short-term anomalous changes in abortion numbers.

In this study, climate data primarily consisted of various diurnal observational indicators, including air temperature, precipitation, atmospheric pressure, mean wind speed, soil moisture, water vapor pressure, relative humidity, and 24-hour Penman potential evapotranspiration. These variables were derived from observational records at meteorological stations throughout New Zealand and were unified to a 5-kilometre spatial resolution climate grid using spatial interpolation methods. The climate data covers the period from January 2003 to December 2023, a time span slightly earlier than the MPI aborted records, providing a relatively flexible time window for subsequent multi-source data integration.

It should be noted that climate data were originally recorded at the District level on a monthly basis, whereas the MPI abortion data were analysed at the Region-month level. To ensure spatial and temporal consistency across data sources, abortion records were first aggregated by Region and month during the data integration phase. A mapping procedure was then applied to align District-level climate observations with the corresponding MPI regional units. For Regions comprising multiple Districts, monthly climate variables were aggregated by computing the arithmetic mean across all Districts within each Region. The resulting Region-level climate indicators were subsequently matched to the corresponding Region–month abortion records for modelling purposes.

Meanwhile, in order to characterise the potential lagged effects of climate conditions on livestock health, lagged climate variables were constructed at the Region–month level. Specifically, one-, two-, and three-month lagged values (t-1, t-2, and t-3) were generated for each climate variable to capture short-term delayed climatic impacts. These lagged indicators were incorporated as additional predictors in the modelling framework to account for potential time-delayed influences of climate factors on abortion incidence.

3.1.3 Google Trends search behaviour data

To further capture potential behavioural responses and shifts in information attention related to variations in livestock abortion counts, this study incorporates internet search behaviour data obtained from Google Trends [29]. Google Trends provides relative search intensity indices derived from search engine usage patterns, which reflect changes in public interest over time for specific topics. Such data are commonly used as proxy indicators of risk perception, information dissemination, or behavioural responses, and therefore offer complementary behavioural insights for abortion forecasting.

The search behaviour data used in this study cover regions across New Zealand and span the period from January 2004 to December 2023. Keyword selection focused on livestock reproduction, animal health, and environmental or disease-related factors that may be associated with abortion risk. The selected keywords include stillbirth, abortion, caesarean section, vet, cull, as well as terms related to abnormal weather and farming environments, such as flood, puddle, paddock, and stormwater. Collectively, these search terms are intended to capture variations in information-seeking behaviour among farmers and veterinary practitioners in response to disease risks or environmental anomalies.

It's important to note that Google Trends does not report absolute search volume, but instead provides a relative index obtained by standardising search frequency within a specific time period and geographic region. Therefore, this data is more appropriate for characterising trends in public attention rather than directly measuring the actual magnitude of specific events. In this study, search behaviour data were employed as an external explanatory variable to compensate for the limitations of traditional statistical data in terms of behavioural response and risk perception. By integrating search behaviour indicators with livestock abortion records and climate information, the model is able to characterise potential driving mechanisms of changes in abortion numbers from multiple perspectives, enhancing its ability to capture short-term fluctuations and anomalous variations.

3.1.4 Livestock population statistics

The livestock population data used in this study are obtained from the National Agricultural Census (NAC) website and are compiled and published by Statistics New Zealand (Stats NZ) [30]. This dataset records population sizes of major livestock categories across New Zealand's 16 administrative regions, including the regional distribution of livestock such as beef cattle, dairy cattle, horses and sheep. Since the agricultural census is conducted on a five-year statistical cycle, this thesis selects regional livestock population data from four census years: 2007, 2012, 2017 and 2022 for subsequent analysis.

As livestock population statistics are available only at five-year intervals, a stepwise carry-forward assignment strategy was adopted to align them with the unified monthly time index used in this study. Specifically, values for 2004-2008 were assigned using the 2007 census data; 2009-2013 used the 2012 census; 2014-2018 used the 2017 census; and 2019-2023 used the 2022 census data. This approach preserves official census estimates while avoiding the introduction of artificial temporal trends through interpolation.

Compared with livestock abortion records, climate data, and Google Trends search behaviour data, livestock population statistics are updated at a substantially lower temporal frequency and therefore cannot reflect short-term fluctuations. Nevertheless, these data provide stable, long-term information on regional livestock scale and production structure, which is valuable for modelling persistent regional differences over extended time horizons. Livestock population size not only directly affects the absolute magnitude of potential abortion events but also serves as an indicator of underlying regional variations in farming practices, production systems, and potential disease risk profiles. As such, despite their coarse temporal resolution, livestock population variables offer important contextual background information for time series analysis.

In this study, livestock population data are mainly incorporated into the model as background variables to describe long-term regional production scale and intensity. On one hand, this information assists the model in making reasonable scale adjustments across regions, avoiding the direct attribution of higher abortion numbers to elevated risk when such differences may instead arise from larger livestock populations. On the other hand, livestock population size may also indirectly influence abortion risk by affecting stocking density, animal contact frequency, and disease transmission conditions. Therefore, incorporating livestock population data helps the model more comprehensively characterise potential disease transmission risks and their long-term impact on changes in abortion numbers.

3.1.5 Descriptive Statistics of Data Sources

Prior to model construction, it is essential to examine the basic characteristics of all data sources to understand their temporal properties, data availability, and inherent limitations. The datasets used in this study differ substantially in temporal coverage, resolution, and variability, reflecting the heterogeneous nature of livestock abortion records, climatic observations, syndromic indicators, and census-based population statistics. Table 3.1 provides a consolidated overview of the temporal coverage, resolution, missingness patterns, variability characteristics, and preprocessing strategies applied to each data source before modelling. This summary facilitates transparency in data integration and clarifies the rationale for the subsequent temporal alignment, aggregation, and feature construction procedures described in the following sections.

Table 3.1 Descriptive overview of the temporal properties, missingness patterns, variability characteristics, and preprocessing strategies of all data sources prior to model construction.

Data Source	Temporal Coverage	Temporal Resolution	Missingness Pattern	Variability Characteristics	Notes
MPI abortion records	2004-2023	Daily	Sparse at daily level; no missing after aggregation	Seasonal fluctuation with occasional spikes	Aggregated to monthly counts
Climate variables	2003-2023	Monthly	No missing	Seasonal pattern	Lag 1-3 months used
Google Trends	2004-2023	Monthly	No missing	Exhibits burst-driven spikes associated with episodic search activity.	Provided as 0-100 relative index; further scaled using MinMaxScaler.
Livestock population	2007-2022	Five-year interval	No missing within reporting years	Gradual long-term trend	Five-year census data (2007-2022); stepwise assignment to monthly time index.

3.2 Data Integration and Temporal Processing

This section introduces the integration and time-based processing workflow for multi-source data, focusing on data aggregation methods at the monthly scale, merging rules at the regional level, and strategies for handling missing values and inconsistent records. Through the above processing steps, all types of raw data are uniformly converted into multivariate time series data that are continuous in time and have a consistent structure, providing standardised input for subsequent feature selection and model training.

3.2.1 Monthly aggregation and regional merging

To satisfy the consistency and continuity requirements of time series modelling, this study first standardises all multi-source data along the temporal dimension, adopting a monthly time scale as the unified temporal unit. Aggregating MPI abortion records at the monthly scale helps alleviate the occurrence of consecutive zero values caused by sample sparsity in daily-scale data, and also facilitates the identification of overall trends in abortion numbers. Furthermore, livestock abortion events are typically influenced by seasonal climate variations and production cycles; the monthly resolution preserves key temporal structural characteristics while effectively reducing noise introduced by high-frequency daily fluctuations. Accordingly, all data sources are converted and aligned to a unified "year–month" time index.

In terms of spatial dimension, the original spatial division of different data sources is not entirely consistent. To improve spatial consistency and enhance the stability and interpretability of the analysis results, this thesis merges the national administrative regions based on the original MPI regional divisions, combining geographical proximity and livestock distribution characteristics to form seven analytical units: Northland and Auckland; Waikato; Bay of Plenty, Gisborne, and Hawke's Bay; Taranaki, Manawatu-Whanganui, and Wellington; Tasman, Nelson, and West Coast; Marlborough and Canterbury; Otago and Southland. Although the original MPI records include finer spatial identifiers at the District Council level (e.g., South Wairarapa District), the present study adopts the Region-level classification provided in the MPI dataset for aggregation. For example, Wairarapa districts are administratively recorded under the Wellington Region and are therefore incorporated within the Wellington analytical unit rather than treated as an independent spatial entity. This regional merging effectively alleviates the data sparsity problem caused by overly detailed spatial divisions while preserving key spatial differences.

After unifying the temporal scale and spatial units, the MPI abortion data were merged with climate variables and Google Trends search behaviour data according to temporal alignment and regional correspondence, thereby constructing a continuous regional-level monthly time series spanning from January 2004 to December 2023. For months in which no abortion cases were reported, the corresponding abortion count was assigned a value of zero to preserve the continuity of the time series and avoid the artificial introduction of temporal discontinuities. Based on the above processing procedures, two datasets were ultimately constructed: one set is national-level data to characterise the overall livestock abortion trend in the livestock abortion numbers; the other set is regional-level data, covering seven aggregated regions, each containing 240 months of continuous observations, providing a unified and consistent data foundation for subsequent regional-scale predictive modelling.

3.2.2 Handling missing values and inconsistencies

Due to heterogeneity in data collection practices, reporting cycles, and update frequencies across the different data sources, the original datasets inevitably contain missing values and inconsistencies. To ensure data reliability and enhance the stability of model training, a series of systematic preprocessing procedures was implemented during the data integration stage. These procedures focused on the temporal completion of livestock population variables, the construction of lagged climate features, and the standardisation of continuous variables.

Livestock population statistics are incorporated as structural variables that characterise long-term production scale and baseline regional differences. As the analysis does not differentiate among individual livestock species, population counts for major livestock categories are aggregated at the regional level to represent the overall livestock scale. Given the coarse temporal resolution of census-based population data, which is not directly compatible with monthly time series modelling, a rule-based interpolation approach was adopted to obtain a continuous and stable population sequence over the study period. This treatment provides consistent long-term contextual background information for subsequent modelling efforts.

In addition, considering that the effects of climatic conditions on livestock reproductive health may manifest with temporal delays, lagged climate features are constructed to capture potential delayed effects on abortion outcomes. The inclusion of lagged variables enables the model to more effectively characterise time-dependent relationships between environmental conditions and abortion dynamics.

After data cleaning, interpolation, and feature construction, all continuous variables were normalised using the min-max scaling approach to enhance the numerical stability during model training. During the model evaluation phase, the prediction results are converted back to the original scale through an inverse normalisation operation, thereby ensuring the interpretability of the prediction results in real-world application scenarios.

3.3 Feature Selection and Screening

This section introduces feature selection and screening strategies, focusing on methods for selecting Google Trends search keywords and climate variables under multi-source data conditions. By integrating statistical analysis results with domain knowledge, this study implements targeted input feature selection to reduce redundant information and emphasise variables most relevant to livestock abortion prediction, providing more stable and effective feature inputs for subsequent model training. In addition, this section further discusses the methodological rationale underlying the adopted screening strategy, including a comparison with alternative feature selection approaches and an explanation of its suitability for the regional time-series data structure of this study.

3.3.1 Selection of Google Trends keywords

To identify search behaviour features potentially associated with variations in the number of livestock abortions, this thesis employs the Pearson Correlation Coefficient (PCC) to analyse the linear correlation between candidate Google Trends search keywords and the monthly number of livestock abortions, thereby preliminarily screening search terms that may fluctuate synchronously with the number of abortions in terms of temporal trends. During the screening process, candidate keywords are generally categorised into two groups: one comprises search terms directly related to livestock reproduction and animal health, and the other comprises search terms associated with the pasture environment and meteorological conditions. After multiple rounds of correlation analysis and comparative evaluation, seven representative keywords were ultimately selected as model input features, namely stillbirth [31], abortion [32], caesarean section, flood [33], puddle, paddock and stormwater.

The results of the correlation analysis between the selected keywords and the number of livestock abortions are shown in Figure 3.1. This indicates that search terms directly related to livestock reproduction and health (such as abortion, caesarean section, and stillbirth) exhibit positive correlations with the number of abortions. These search terms typically reflect shifts in the information attention among farmers, veterinarians, and the general public when abnormal reproductive events are observed, and have been widely regarded as an important indirect indicator for characterising the risk of livestock abortion in previous epidemiological studies. In contrast, environment-related keywords, including flood, puddle, paddock, and stormwater, display weaker linear correlations, with relatively small absolute correlation coefficients. Nevertheless, these terms may capture changes in attention related to pasture moisture, water accumulation, or extreme weather events. Existing studies have indicated that increased environmental moisture, flooding, and waterlogging can facilitate pathogen transmission and complicate farm management, thereby indirectly increasing abortion risk. For instance, pathogens such as Leptospira and Listeria, which are known causes of abortion and stillbirth in cattle and sheep, have been associated with such environmental conditions [33,34].

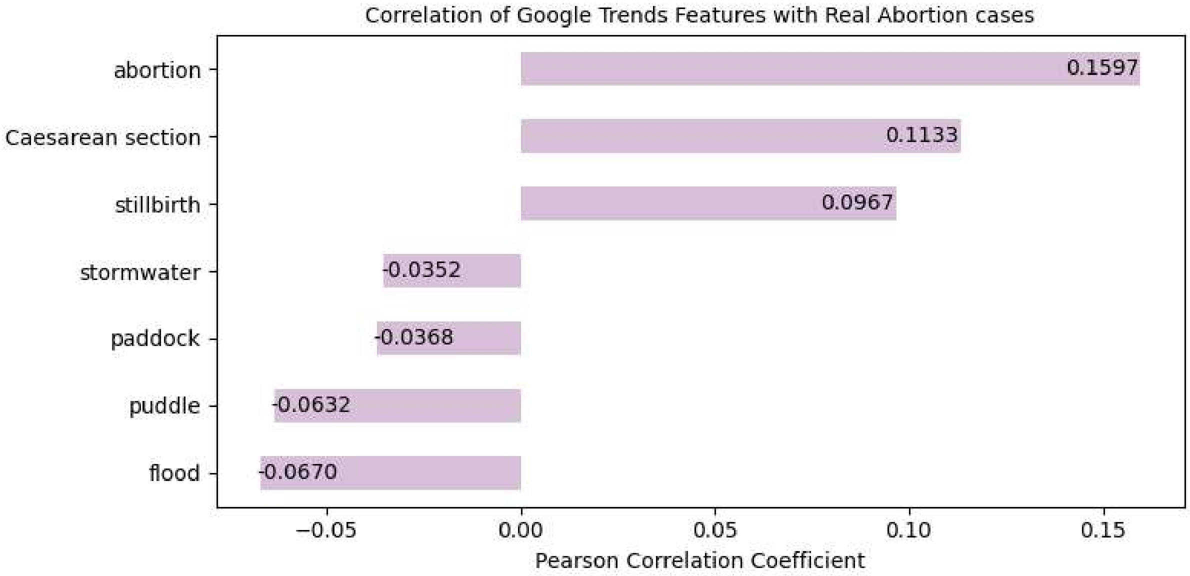


Figure 3.1 Pearson correlation coefficients between seven selected Google Trends search keywords and the actual number of livestock abortions. The results showed that the actual number of livestock abortions was positively correlated with keywords related to livestock reproduction and health, while keywords related to the environment showed a weaker linear correlation.

A further comparative analysis of the time series patterns of the selected search keywords and observed livestock abortion numbers is presented in Figure 3.2. Reproduction-related keywords exhibit periods of temporal alignment with increases in abortion counts, suggesting that search behaviour may partially reflect real-world changes in livestock reproductive health conditions. Based on the combined evidence from correlation analysis and domain knowledge, the seven selected keywords were retained as explanatory input features. These keywords encompass both direct biological relevance and indirect environmental or behavioural signals, thereby providing complementary information to support subsequent time series forecasting models.

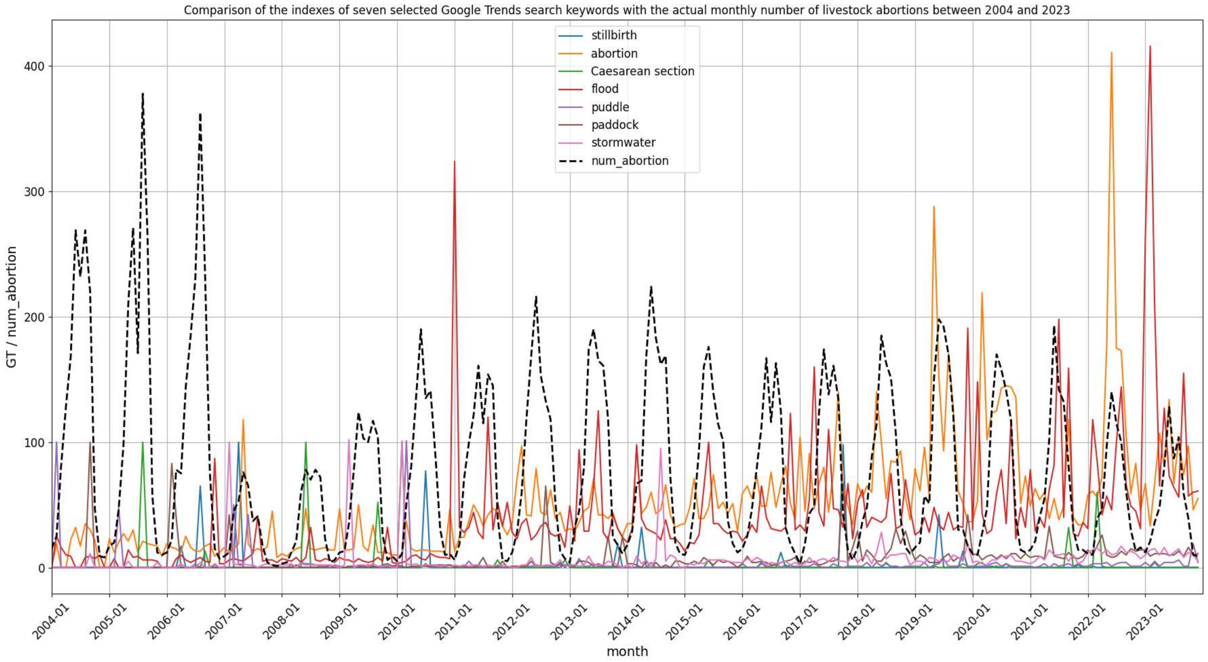
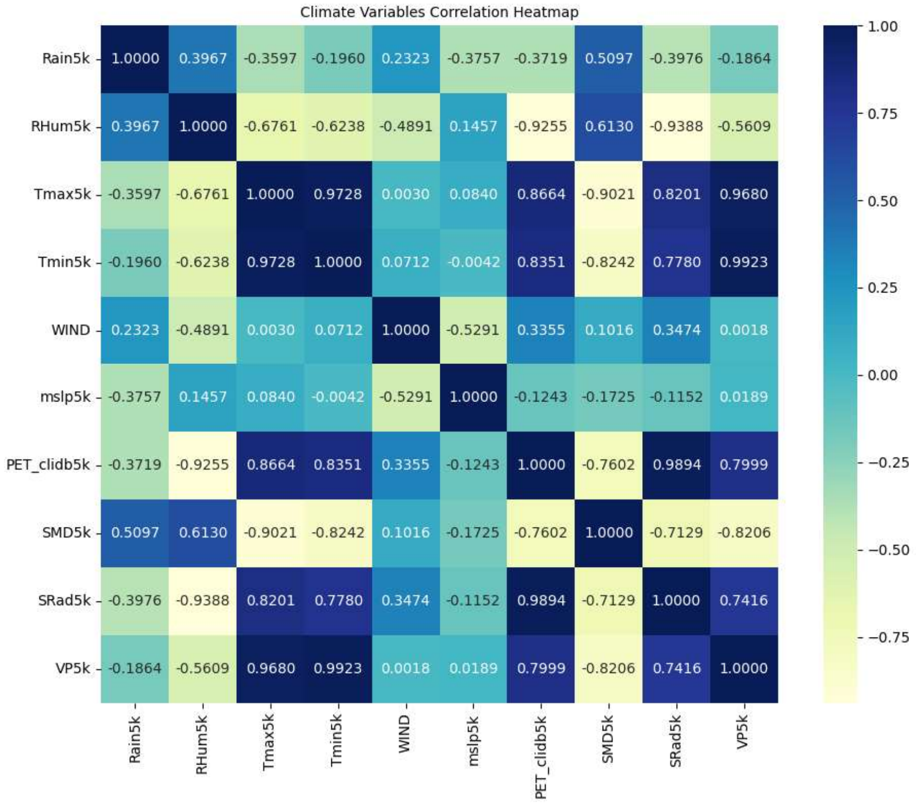


Figure 3.2 Comparison of the indexes of seven selected Google Trends search keywords with the actual monthly number of livestock abortions between 2004 and 2023. The results showed that the search terms related to reproduction coincided with the increase in the number of livestock abortions over time, suggesting a possible consistency between search behaviour and livestock reproductive health outcomes.

3.3.2 Climate variable screening

Climate variables play an important role in livestock health and reproduction research, but strong intercorrelations frequently exist among different climate indicators. If all climate variables are directly incorporated into the model without prior screening, multicollinearity may be introduced, which can adversely affect the stability of the model training and predictive performance. Therefore, this thesis first examines the correlations between various climate variables to identify potential information redundancy during the selection of climate features. Then, it further selects climate features based on their correlation with livestock abortion counts and their physical interpretability. The correlation matrix of candidate climate variables is presented in Figure 3.3. The results indicate that several climate indicators exhibit strong intercorrelations. For example, relative humidity (RHum5k) demonstrates pronounced correlations with potential evapotranspiration (PET_clidb5k) and solar radiation (SRad5k), reflecting overlapping information associated with moisture and radiation conditions. Similarly, maximum temperature (Tmax5k) is highly correlated with minimum temperature (Tmin5k), soil moisture deficit (SMD5k), and vapour pressure (VP5k), suggesting that these variables represent different aspects of thermal or moisture-related states.



Climate Variables Correlation Heatmap

	Rain5k	RHum5k	Tmax5k	Tmin5k	WIND	mslp5k	PET_clidb5k	SMD5k	SRad5k	VP5k
Rain5k	1.0000	0.3967	-0.3597	-0.1960	0.2323	-0.3757	-0.3719	0.5097	-0.3976	-0.1864
RHum5k	0.3967	1.0000	-0.6761	-0.6238	-0.4891	0.1457	-0.9255	0.6130	-0.9388	-0.5609
Tmax5k	-0.3597	-0.6761	1.0000	0.9728	0.0030	0.0840	0.8664	-0.9021	0.8201	0.9680
Tmin5k	-0.1960	-0.6238	0.9728	1.0000	0.0712	-0.0042	0.8351	-0.8242	0.7780	0.9923
WIND	0.2323	-0.4891	0.0030	0.0712	1.0000	-0.5291	0.3355	0.1016	0.3474	0.0018
mslp5k	-0.3757	0.1457	0.0840	-0.0042	-0.5291	1.0000	-0.1243	-0.1725	-0.1152	0.0189
PET_clidb5k	-0.3719	-0.9255	0.8664	0.8351	0.3355	-0.1243	1.0000	-0.7602	0.9894	0.7999
SMD5k	0.5097	0.6130	-0.9021	-0.8242	0.1016	-0.1725	-0.7602	1.0000	-0.7129	-0.8206
SRad5k	-0.3976	-0.9388	0.8201	0.7780	0.3474	-0.1152	0.9894	-0.7129	1.0000	0.7416
VP5k	-0.1864	-0.5609	0.9680	0.9923	0.0018	0.0189	0.7999	-0.8206	0.7416	1.0000

Figure 3.3 Pearson correlation matrices for 10 climate variables. The results show that there is a strong correlation among multiple climate indicators, reflecting an overlap in their information characterising temperature, humidity, and radiation conditions.

Based on this analysis, additional feature selection was performed by jointly considering the correlations between climate variables and monthly livestock abortion counts, as well as their physical interpretability (Figure 3.4). Among variables conveying similar environmental information, those with clearer physical meaning and relatively stronger associations with abortion counts were prioritised. For instance, solar radiation (SRad5k) was retained to characterise radiation conditions, maximum temperature (Tmax5k) was selected to represent thermal stress, and soil moisture deficit (SMD5k) was included to capture pasture moisture conditions. In addition, although certain climate factors, such as rainfall and wind speed, exhibit weak linear correlations with abortion counts, they remain biologically and operationally relevant in livestock production systems. These factors may indirectly influence abortion risk by affecting environmental moisture conditions, pathogen transmission processes, or animal stress levels. Therefore, feature selection was not solely based on correlation thresholds, and rainfall (Rain5k) and wind speed (WIND) were retained on the basis of domain knowledge.

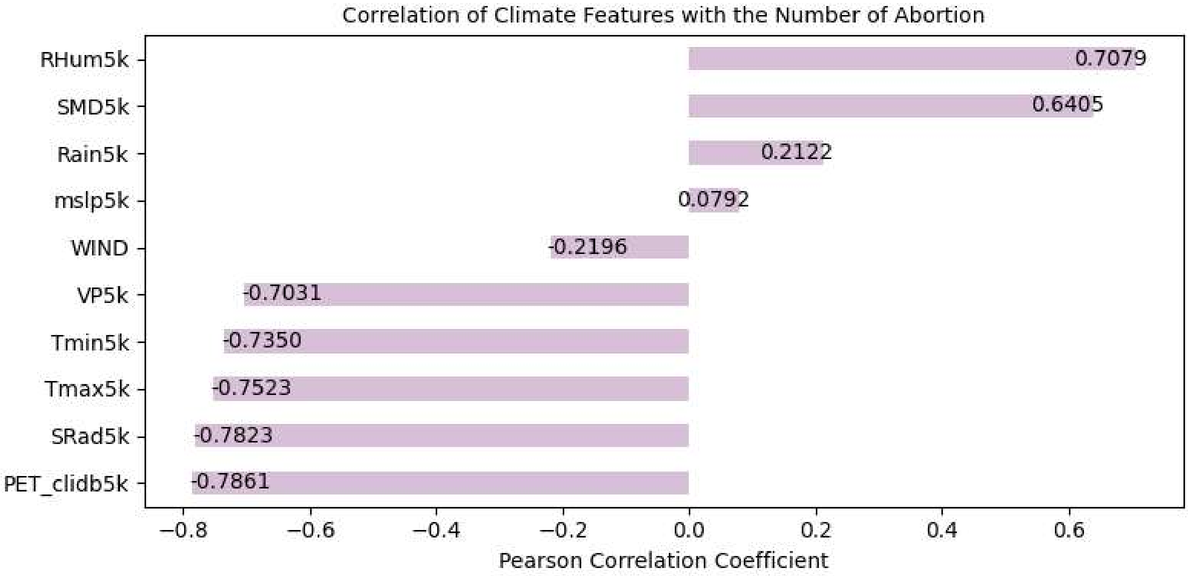


Figure 3.4 Pearson correlation coefficients between 10 climate variables and monthly livestock abortions.

In summary, six climate variables were selected as baseline climate features: RHum5k, Tmax5k, SMD5k, SRad5k, Rain5k, and WIND. To explicitly account for potential delayed effects of climatic conditions on reproductive outcomes, lagged features with delays of one to three months were constructed for each selected variable. This resulted in a total of 18 climate-related input features, which were subsequently employed for model training and evaluation.

3.3.3 Rationale and Methodological Considerations

This study employs a correlation coefficient–based screening approach during the feature selection stage, combined with domain knowledge for comprehensive evaluation, aiming to achieve a balance between statistical relevance and practical interpretability. Considering the heterogeneous nature of the data sources and the relatively limited number of temporal observations at the regional level, correlation analysis provides a transparent, computationally efficient, and readily interpretable variable selection strategy. This approach helps maintain consistency in screening criteria across different regions while avoiding the introduction of unnecessarily complex modelling assumptions under limited-sample conditions.

It should be noted that only reliance on statistical correlation may be influenced by spurious associations or random temporal fluctuations. Therefore, during the screening process, this study further incorporated domain knowledge related to livestock production practices and epidemiological mechanisms of animal diseases to assess the plausibility of candidate variables, ensuring that the selected predictors possess practical significance within biological and agricultural production contexts.

Other commonly adopted feature selection approaches also exist in the literature. For example, regularisation-based feature selection methods (such as L1 Regularisation or Lasso regression) enable automatic coefficient shrinkage in high-dimensional feature spaces; model-based importance ranking approaches (such as feature importance measures derived from tree-based algorithms like Random Forest) are also frequently employed to evaluate the contribution of variables to predictive performance. These methods perform effectively in large-sample or high-dimensional data environments and are particularly suitable for settings characterised by complex nonlinear relationships or variable interaction effects.

However, within the regional time-series data structure of this study, substantial collinearity exists among climate variables, and the temporal sample size is relatively limited. Under such conditions, model-derived importance rankings may be sensitive to sampling variability and influenced by model-specific structural assumptions. In contrast, the screening approach based on correlation coefficients demonstrates greater stability and interpretability, and facilitates comparability across experimental settings in different regions. Therefore, this study adopts correlation-based screening as a preliminary variable filtering step, while the ultimate representation and predictive contribution of variables are subsequently learned through the deep learning models.

It must be acknowledged that correlation analysis primarily captures linear associations and cannot directly characterise complex nonlinear or interaction effects. Consequently, the feature selection results in this study do not constitute causal inference, but instead provide a reasonable and stable set of input variables for subsequent model training.

3.4 Dataset Splitting Strategy

To objectively evaluate the predictive performance of the model and prevent information leakage, this thesis divides the dataset in a time-aware manner after feature construction. Given that the research object is time-series data, the construction of the training, validation, and test sets strictly follows the chronological order, rather than using random partitioning. This section will explain the specific dataset partitioning scheme, including the time range settings for each subset, and the basic principles for preventing future information leakage during the partitioning process, so as to ensure that the model evaluation results can truly reflect its performance in actual prediction scenarios.

3.4.1 Training, validation, and test sets

This thesis arranges the datasets from earliest to latest according to time, and divides them into three parts: training set, validation set and test set. This division method strictly follows the basic principle of "predicting the future with the past" in time series forecasting, ensuring that the model uses only historically available information during training and evaluation, thereby avoiding the leakage of future information.

The specific partitioning scheme is shown in Table 3.2: The training set contains samples with a target time of December 2017 and earlier, used for learning and fitting model parameters; the validation set is used for model structure selection and hyperparameter tuning, with a target prediction time range of January 2020 to December 2021, and the corresponding input sequence is historical observation data from January 2018 to December 2019; the test set is used for final model performance evaluation, with a target prediction covering January 2022 to December 2023, and the model input is historical data from January 2020 to December 2021.

Table 3.2 The table shows the dataset splitting rules. This dataset is split into training, validation, and test sets in chronological order to ensure that only historical information is used for model learning and evaluation, preventing data leakage.

Dataset	**Target Period**	**Input Period (24-month window)**
Training set	Up to Dec 2017	Historical data up to Dec 2017
Validation set	Jan 2020 - Dec 2021	Jan 2018 - Dec 2019
Test set	From Jan 2022 onward	Jan 2020 - Dec 2021

In the process of sample construction, this thesis uses a 24-month historical observation sequence as model input to predict the number of livestock abortions in the following month. This input window setting allows the model to simultaneously include complete annual seasonal information and longer-term trend characteristics, which helps the model learn more effectively the periodic patterns and long-term time dependencies in livestock abortion time series.

3.4.2 Avoidance of information leakage

To avoid potential information leakage during time series modelling, this thesis strictly adheres to the principle of temporal order in all stages of dataset partitioning, sample construction, and model evaluation. All input data used for model training, validation, and testing comes from historical observations prior to the target prediction time, ensuring that the model is not exposed to future information at any stage and guaranteeing the rationality of the experimental design from a temporal perspective.

In choosing a model evaluation method, this thesis does not employ the traditional k-fold cross-validation strategy. This is because time series data usually have a significant time-dependent structure. Randomly dividing samples and performing cross-validation will disrupt the original time order, which may lead to future information being inappropriately introduced into the model training or validation process, thus causing systematic bias in the evaluation results. Therefore, in the time series prediction scenario discussed in this study, traditional cross-validation methods are not applicable.

To make the model evaluation process more closely resemble real-world prediction scenarios, this thesis employs a time-series-based strategy for partitioning the training, validation, and test sets. A separate validation set is used for model structure selection and hyperparameter tuning. Once the model structure and parameters are fully determined, the test set is used only for final performance evaluation and does not participate in any form of model adjustment. Through the above design, this thesis effectively avoids information leakage while ensuring the reproducibility of the experimental process, enabling the model evaluation results to more realistically reflect its potential application performance in actual livestock abortion prediction tasks.

4. LSTM Model with an Attention Mechanism Framework

This chapter discusses Long Short-Term Memory (LSTM) time series prediction models that incorporate attention mechanisms. First, the overall modelling framework and key design considerations are presented, and the background and motivation for introducing the attention mechanism in the task of monthly livestock abortion prediction are systematically explained. Then, the specific structural components of the model are introduced in detail, including the input regularisation strategy, the integration of the attention mechanism with LSTM architecture, and the design of the prediction output layer. On this basis, the training process and prediction performance of the model on national-level data are further analysed, and the model's behaviour is evaluated and discussed in terms of performance metrics and sensitivity to input variables.

4.1 Model Overview

Long-term, multivariate time series forecasting is a core problem in many practical applications. The main challenge lies in simultaneously characterising the long-term trend evolution and short-term local fluctuations of the series. In the scenario of predicting monthly abortions in livestock farming, time series typically exhibit a significant seasonal structure, while also being affected by factors such as abnormal weather, disease outbreaks, or changes in management measures at certain times, resulting in substantial short-term disturbances. These complex time-series characteristics, which combine periodicity and suddenness, place higher demands on the ability of prediction models to model information across different time scales.

Long Short-Term Memory (LSTM) networks effectively alleviate the gradient vanishing problem that traditional recurrent neural networks are prone to during long sequence modelling by introducing a gating mechanism. Since its inception, it has been widely used in various time series prediction tasks [16]. However, when faced with complex time-series data with high noise and multiple variable inputs, standard LSTM tends to model historical time steps relatively uniformly, making it difficult to explicitly distinguish the importance of different moments to the prediction results, thus limiting its ability to characterise key time dependencies. To address this issue, recent studies have gradually introduced attention mechanisms into sequence modelling frameworks, enabling models to adaptively allocate weights in the time dimension and more effectively focus on historical segments that contribute significantly to the prediction target [46]. Meanwhile, as network structures become deeper and more complex, the introduction of residual connections and normalisation strategies has been shown to stabilise the optimisation process and alleviate gradient degradation issues that may arise during deep model training [47,48].

Based on the above research background, this study constructs an LSTM-based time series prediction model that integrates with attention mechanisms. While retaining LSTM as the main structure for time series modelling, a multi-head attention mechanism and a residual normalisation module are introduced to enhance the model's capacity to represent complex temporal dependencies. At the same time, considering the inevitable noise perturbation and abnormal fluctuations in real-world observational data, a regularisation strategy based on random perturbation is applied at the input stage of the model. Existing research has demonstrated that injecting moderate Gaussian noise into the input features during training can improve the generalisation capability of the model and reduce the risk of overfitting to local abnormal patterns [49]. Through the above structural design, this model serves as a systematic time series modelling framework for examining the effects of attention mechanisms and structural regularisation strategies in multivariate prediction tasks.

4.2 Training and Evaluation Setup

To ensure fairness, stability, and reproducibility in comparing different models for predicting the number of livestock abortions, this thesis adopts a unified experimental setup during the model training and evaluation phases. This section focuses on the loss function, optimisation strategy, and evaluation metric system used in the training process. This training and evaluation protocol applies not only to the prediction model proposed in this chapter but also to the Mixing-Channel PatchTST with parallel LSTM branch model introduced in Chapter 5, to ensure that different methods can be compared and analysed under consistent experimental conditions. The loss function and evaluation metrics used will be explained in detail below.

4.2.1 Loss Function

During the model training phase, the choice of loss function has a significant impact on the stability and robustness of the prediction results. For the time series prediction task of livestock abortion numbers, this thesis adopts Huber Loss as the optimisation objective function of the model to replace the traditional mean squared error (MSE). Huber Loss was first proposed by Huber (1964) [42] in robust statistical estimation research. Its design goal is to achieve a balance between fitting accuracy and outlier robustness.

Huber Loss has a piecewise definition across different error ranges: when the prediction error is small, its loss form is the same as MSE, providing smooth gradient information, which is beneficial for the stable convergence of the model; when the prediction error exceeds a preset threshold, the loss function transforms into a linear form, similar to Mean Absolute Error (MAE), thereby effectively suppressing the excessive influence of extreme errors on model parameter updates. This feature makes Huber Loss more robust to outliers while maintaining its overall fitting ability.

For livestock abortion time series data, the series typically contains both relatively stable long-term trends and short-term abnormal fluctuations caused by factors such as extreme weather conditions and disease outbreaks. In this case, using MSE alone can easily lead to oversensitivity to outlier samples, while using MAE alone may weaken the model's ability to fit the overall trend. Considering the above, this thesis selects Huber Loss as the objective function for the training phase, and its mathematical form is defined as follows:

$$L_\delta(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta \end{cases}, \tag{31}$$

Where e represents the error between the predicted and actual values, and δ is the threshold parameter controlling the switching between the quadratic and linear loss terms.

By introducing Huber Loss, the model can effectively reduce the impact of abnormal fluctuations and outliers on parameter updates while maintaining a stable fit to the overall long-term trend, thus achieving a balance between robustness and fitting ability, making it more suitable for the time series prediction task of livestock abortion numbers involved in this study.

4.2.2 Evaluation Metrics

This thesis selects a variety of evaluation metrics that are widely used in time series regression tasks to measure the predictive ability of the model from different perspectives, including root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) [43-45].

1) Root Mean Squared Error (RMSE)

RMSE measures the overall deviation between predicted and actual values. It is more sensitive to larger prediction errors and reflects the model's overall fit quality. Its definition is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{32}$$

where n represents the total number of predicted samples, and y_i and $\hat{y}_i$ represent the true value and predicted value of the i-th sample, respectively. The smaller the RMSE value, the lower the overall prediction error of the model.

2) Mean Absolute Error (MAE)

MAE measures model performance by calculating the average absolute value of prediction errors. Compared to RMSE, it is less sensitive to outliers and better reflects the stability of prediction results. Its definition is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \tag{33}$$

A smaller MAE indicates that the model has a lower average bias, which generally corresponds to more stable prediction performance across the overall prediction process.

3) Coefficient of Determination (R^2)

The coefficient of determination (R^2) quantifies a model's ability to explain the variance of the target variable. It is a widely used comprehensive performance metric in regression tasks, and its definition is:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{34}$$

where $\bar{y}$ represents the sample mean of the actual values, the value of R^2 reflects the extent to which the model explains the variance of the target variable: when R^2 is close to 1, it indicates that the model can well characterize the variation features of the target sequence; when R^2 is close to 0, the model's predictive performance is comparable to that of using the sample mean; if R^2 is negative, it indicates that the model has failed to adequately capture the dominant variation patterns in the data.

In the task of predicting the number of livestock abortions, R^2 can reflect the model's ability to capture long-term trends and regional differences from the perspective of explained variance. When combined with RMSE and MAE, this evaluation metric system enables a more comprehensive and reliable assessment of predictive performance across multiple dimensions, such as error magnitude, predictive stability, and overall goodness of fit.

4.3 Architecture of the LSTM with Attention Mechanism

The overall structure of the attention-enhanced LSTM model is illustrated in Figure 4.1. Based on the conventional stacked LSTM framework, this model integrates multiple structural components commonly used in sequence modelling to enhance representational capacity and training stability in multivariate time-series prediction tasks.

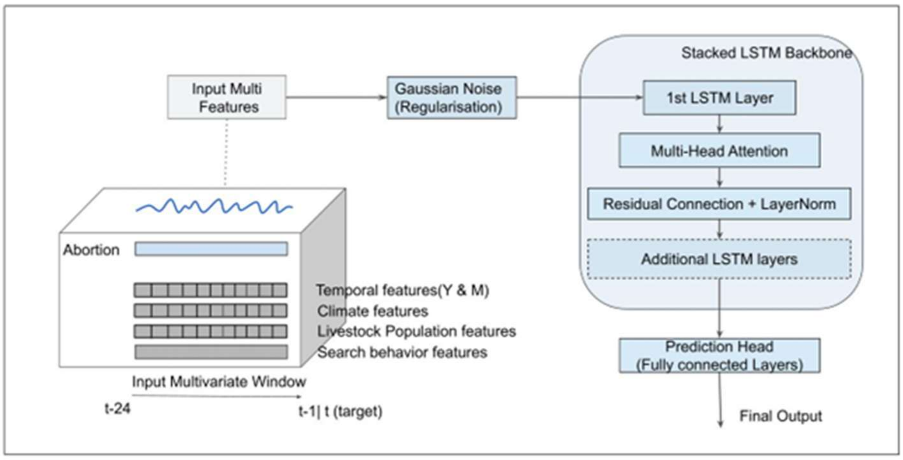


Figure 4.1 The architecture of an LSTM model with an attention mechanism. The model receives a multivariate input window containing historical abortion counts, temporal features, climate variables, livestock population data, and search behaviour indicators. Gaussian noise is injected into the input stage for regularisation, followed by a stacked LSTM backbone network that enhances the multi-head attention module and residual normalisation, and finally, the final output is generated through a fully connected prediction head.

The model consists of three principal components: an input regularisation module, an LSTM backbone network with an attention mechanism, and a fully connected output layer for generating prediction results. The model takes a fixed-length multivariate time window as input, which includes historical abortion number sequences and related auxiliary features such as time characteristics, climate variables, livestock population information, and search behaviour indicators. During the input phase, the input is regularised by introducing random perturbations into the feature space to reduce the model's sensitivity to local noise or abnormal patterns in the training samples, thereby improving the model's generalisation performance.

After input regularisation, the sequence features are fed into a stacked LSTM network to model temporal dependencies. Unlike the standard LSTM structure, this model introduces a multi-head attention mechanism after the output of the first LSTM layer, and combines residual connections and normalisation operations to enhance the model's ability to model information at key time steps, while improving the numerical stability of the deep recurrent structure during training. Depending on the specific experimental setup, the LSTM backbone can be stacked in multiple layers to gradually extract higher-level temporal representations. Finally, the high-dimensional temporal features are mapped to predicted values of the target variable through a prediction head composed of fully connected layers. These structural modules work synergistically within the overall framework, enabling the model to maintain the advantages of LSTM in modelling temporal dependencies while possessing stronger feature selection capabilities and more robust training performance.

4.3.1 Input Regularisation with Gaussian Noise

In multivariate time series prediction tasks, input features are inevitably affected by measurement errors, statistical fluctuations, and uncertainties in the external environment. For predicting monthly abortions in livestock, auxiliary features such as climate variables, livestock population statistics, and search behaviour indicators inherently possess strong noise characteristics and time-varying properties. If the model is overly sensitive to local fluctuations or random perturbations in the input sequence during training, it is prone to overfitting, thereby weakening its generalisation ability at future time points.

Following the idea of introducing Gaussian noise as a regularisation method in similar LSTM model structures [50], this study introduces a regularisation strategy based on random perturbation in the model input stage. During training, a noisy input sequence is constructed by adding a random perturbation drawn from a zero-mean normal distribution to the original input sequence, which can be expressed mathematically as follows:

$$\tilde{x}_t = x_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{35}$$

where ε_t denotes the random noise term injected at time step t, and σ is a parameter controlling the magnitude of the input perturbation. This form of input-level regularisation reduces the model's reliance on local anomalous patterns during training and leads to improved generalisation on unseen data.

4.3.2 Attention Mechanism and Residual Normalisation in LSTM

After input regularisation, the sequential features are passed to the subsequent LSTM backbone for temporal dependency modelling. After the first LSTM layer, a multi-head attention module is added together with a residual connection and layer normalisation to improve the model's representation of informative time steps and long-range temporal dependencies. As illustrated in Figure 4.1, this design retains the inherent advantage of LSTM in preserving sequential order, while enabling adaptive weighting of different time steps through the attention mechanism.

Specifically, let the output of the first LSTM layer at time step be denoted as $h_t \in \mathbb{R}^d$. By stacking these outputs along the temporal dimension, a hidden state sequence representation is obtained:

$$H = [h_1, h_2, \cdots, h_T] \in \mathbb{R}^{T \times d}. \tag{36}$$

Building upon this, the multi-head attention mechanism constructs query, key, and value vectors respectively through linear mapping:

$$Q = HW^Q, K = HW^K, V = HW^V, \tag{37}$$

where W^Q, W^K, and $W^V$are learnable parameter matrices. Each attention head models the correlation between different time steps through a scaled dot product attention mechanism [19] and performs weighted aggregation of features to highlight the temporal information that is more critical to the prediction target. Then, to further stabilise the training process of the deep recurrent architecture, this study introduces residual connections, in which the attention output is added to the original LSTM hidden state representation, thereby preserving the fundamental temporal information learned during sequence modelling. Its form can be expressed as:

$$H' = H + MHA(H). \tag{38}$$

Subsequently, layer normalisation is introduced to standardise the fused features:

$$H'' = LayerNorm(H'). \tag{39}$$

The introduction of residual connections and layer normalisation in this manner helps to alleviate the gradient propagation problem and accelerate model convergence [47,48], thereby enhancing numerical stability during the training process.

4.3.3 Prediction Head and Activation Function

After completing sequence feature modelling and attention enhancement, the model needs to map the high-dimensional temporal representation to the final prediction output. To achieve this, this study introduces a prediction head composed of fully connected layers after the LSTM backbone network. This prediction head is responsible for compressing temporal features and performing regression mapping, thereby generating the predicted value for the target time point.

Let the hidden representation obtained from the final LSTM layer be denoted as:

$$h \in \mathbb{R}^{B \times d},$$

Where B represents the batch size and d is the hidden dimension of the top-level LSTM. This hidden representation is first input to a fully connected transform layer, where feature recombination is achieved through nonlinear mapping, and can be expressed as follows:

$$z = \sigma(W_1 h + b_1), \tag{40}$$

with W_1 and b_1 represent the weight matrix and bias term, respectively, and $\sigma(\cdot)$ is the activation function. By introducing nonlinear transformations, the model's ability to express complex feature relationships can be enhanced.

In this study, the Gaussian Error Linear Unit (GELU) was used as the activation function. Unlike the traditional ReLU activation function, the GELU function introduces the probability characteristics of the input distribution, which makes it perform a smooth weighting operation on the input features, thereby avoiding the gradient discontinuity problem caused by hard thresholding [51]. Its mathematical definition is as follows:

$$GELU(x) = x \cdot \Phi(x), \tag{41}$$

where $\Phi(x)$ represents the cumulative distribution function of the standard normal distribution, and the final result used to generate the single-step prediction is:

$$\hat{y} = W_2 z + b_2, \tag{42}$$

where $\hat{y}$ represents the model's predicted value at the target time point.

4.4 Model Performance and Analysis

This chapter evaluates the training behaviour and predictive performance of the LSTM model with an attention mechanism on national-level data. This analysis mainly focuses on the training convergence characteristic, the agreement between predicted values and actual values, the evaluation of quantitative performance metrics, and a sensitivity analysis examining the model’s response to variations in input features.

4.4.1 Training Convergence and National-Level Prediction Performance

As shown in Figure 4.2, the figure illustrates the evolution of training and validation losses of the attention-enhanced LSTM model during the training process as the number of training iterations increases. It can be observed that, as training proceeds, the loss values on both the training and validation sets exhibit an overall downward trend, with relatively smooth and stable convergence behaviour.

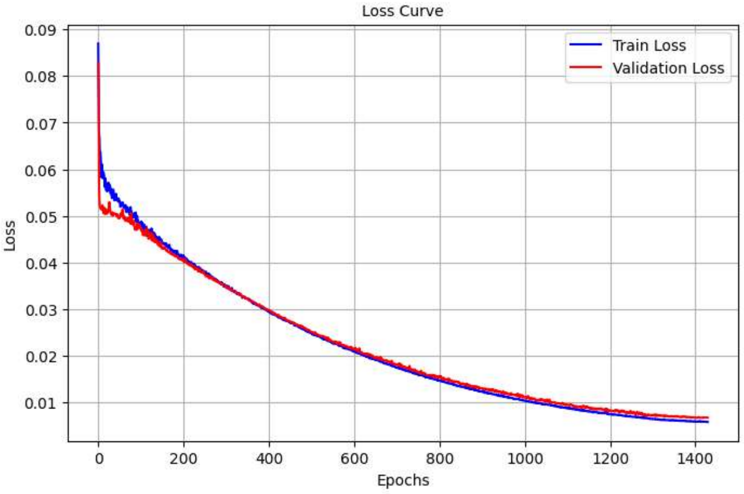


Figure 4.2 Training and validation loss curves of the LSTM model with an attention mechanism on national-level data, illustrating the convergence behaviour of the model over training epochs.

In the initial stage of training, the loss function decreases rapidly, indicating that the model can effectively capture the main temporal structure of the input time series within a relatively small number of training iterations. As training continues, the rate of loss reduction gradually slows, and the curves exhibit a smooth convergence pattern in later stages. This change characteristic is consistent with the convergence behaviour of deep time series models. Moreover, throughout the entire training process, the training and validation loss curves remain closely aligned, indicating good consistency between learning on the training and validation sets, with no evident signs of overfitting or underfitting.

As shown in Figure 4.3, the figure shows a comparison between the predicted and observed monthly livestock abortion counts generated by the attention-enhanced LSTM model at the national level.

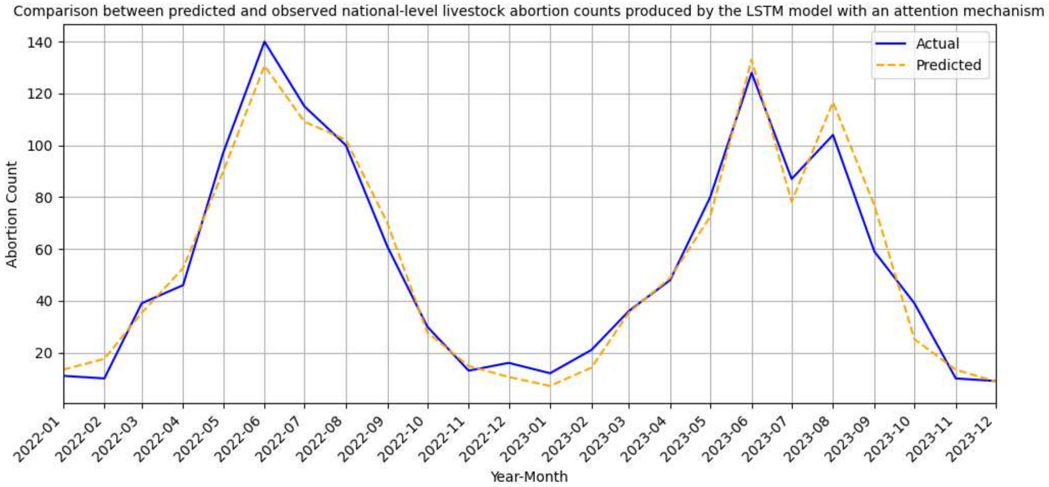


Figure 4.3 Comparison between predicted and observed national-level livestock abortion counts produced by the LSTM model with an attention mechanism over the evaluation period.

From an overall trend perspective, the predicted and actual value curves exhibit consistent upward and downward movement across most time points, indicating that the model can effectively capture the overall variation pattern of abortion numbers at the national scale. From a seasonal perspective, during the relatively high seasonal peaks and troughs in abortion numbers, the predicted and actual values are largely consistent in time, with only minor deviations. This demonstrates that the model possesses a certain degree of stability in learning long-term seasonal structures and periodic patterns, and can capture data trends well.

4.4.2 Quantitative Performance Evaluation

Building on the preceding analyses of training convergence and prediction visualisation, this section further assesses the predictive performance of the LSTM model incorporating an attention mechanism on the test dataset using multiple regression evaluation metrics. Table 4.1 presents a summary of the quantitative results for the national-level livestock abortion prediction task, including the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

Table 4.1 RMSE, MAE, and R^2 results of the LSTM model with an attention mechanism on the national-level test dataset.

Model	**RMSE**	**MAE**	$\boldsymbol{R^2}$
LSTM model with an attention mechanism	7.4426	6.072	0.9668

In terms of error magnitude, the model achieves an RMSE of 7.4426 and an MAE of 6.072 on the test set. While RMSE is more sensitive to larger prediction deviations, MAE reflects the average absolute error across all time points. The relatively close values of these two metrics suggest that the distribution of prediction errors is reasonably balanced over time, without overall performance being dominated by a small number of extreme deviations. This observation is consistent with the overall trend alignment observed in the prediction curves discussed earlier.

In terms of goodness of fit, the model achieves an R^2 value of 0.9668, indicating that it can explain the majority of the variance in the target time series and demonstrates strong overall fitting ability.

4.4.3 Sensitivity Analysis of Input Variables

To examine the influence of different categories of input variables on predictive performance, this study developed six progressively expanded feature sets. These feature sets were organized in a hierarchical sequence, progressing from time-dependent information to scale-structural factors, followed by environmental driving variables, and finally behavioral signaling variables. By introducing variables incrementally, the marginal contribution of each information category was assessed under conditions where other experimental settings remained largely consistent.

The historical abortion counts, serving as the primary time-dependent variable of the prediction target, constitute the fundamental input across all experimental configurations. The time series itself exhibits evident autocorrelation and seasonal periodicity, forming the core informational basis upon which the model learns short-term temporal dynamics and cyclical patterns.

In this framework, livestock population size is treated as a structural scale variable. Abortion counts represent a non-negative count-type outcome whose absolute magnitude is inherently bounded by the overall livestock inventory. Under a relatively stable abortion rate, fluctuations in inventory size directly influence the total number of observed abortion events. Accordingly, the population variable is incorporated not as a direct causal determinant, but as a structural constraint characterizing the exposure base of potential abortion occurrences, enabling the model to distinguish between scale-driven and risk-driven variations.

In modeling climate variables, this study adopts a 1-3 month lag structure rather than directly utilizing contemporaneous climate data. Climatic influences on livestock abortion are generally not instantaneous, but are transmitted gradually through intermediary physiological and environmental processes. The use of lagged climate variables aims to capture this delayed transmission pathway within a biologically plausible time window.

Following environmental variables, Google Trends search intensity features are incorporated as behavioral signal variables. These variables do not directly participate in physiological mechanisms but may reflect farmers' risk perception or management responses, and are therefore introduced at a later stage to evaluate the incremental contribution of behavioral information under a multi-source framework.

Furthermore, a control configuration excluding Time2Vec encoding is constructed based on the complete feature combination to independently assess whether continuous temporal encoding provides additional representational capacity beyond conventional month one-hot encoding.

Through this hierarchical and progressive design, the study evaluates the marginal impact of different information categories under relatively controlled conditions, while avoiding interpretability challenges that may arise when all variables are introduced simultaneously. The detailed experimental configurations are presented below.

The experimental conditions for the feature increments used in this section are shown in Table 4.2. Set 1 contains only one-hot encoded information of historical abortion numbers and months as the most basic time series input. Based on this baseline, Set 2 incorporates Time2Vec encoding for the month variable to enhance the representational capacity of periodic temporal features. Set 3 further introduces total livestock population size information, in addition to temporal features, to capture the influence of macro-scale production changes on prediction outcomes. Set 4 expands the feature space by adding multiple climate variables and their corresponding lagged terms. Finally, based on the preceding feature sets, Set 5 additionally incorporates Google Trends search behaviour features to characterise potential variations in external information attention.

Table 4.2 The incremental feature set configuration for input variables gradually incorporates features such as month, livestock population statistics, climate, and search behaviour.

Feature Set	**Description**
Set 1	Historical abortion counts with month one-hot encoding
Set 2	Set 1 with Time2Vec-based month encoding
Set 3	Set 2 with aggregated livestock population features
Set 4	Set 3 with lagged climate variables (six features with 1-, 2-, and 3-month lags)
Set 5	Set 4 with Google Trends search behaviour features (seven keywords)
Set 6	Set 5 without Time2Vec-based month encoding

The results are reported in Table 4.3. It can be observed that when the model relies solely on basic temporal features (Set 1 and Set 2), the overall prediction error remains relatively large, and the R^2 values are comparatively low, indicating that temporal information alone is insufficient to fully capture the complex variation patterns in monthly livestock abortion data. Notably, the overall performance of Set 2 is inferior to that of Set 1. This suggests that the additional Time2Vec encoding of the month variable did not lead to more informative feature learning, but instead introduced additional noise into the model. This result may also imply that Time2Vec-based temporal encoding requires complementary auxiliary features to function effectively as a supportive representation. When livestock population size information is incorporated (Set 3), the model exhibits improvements across RMSE, MAE, and R^2, indicating that livestock population scale contributes positively to predictive performance by providing important macro-level contextual information.

Table 4.3 The performance of the LSTM model with an attention mechanism on different input features was evaluated using RMSE, MAE, and R^2. Bold values indicate relatively better results.

Feature Set	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Set 1	10.8633	7.2158	0.9293
Set 2	11.0949	8.0018	0.9263
Set 3	10.4742	7.9723	0.9343
Set 4	9.3566	6.6244	0.9476
Set 5	**7.4426**	**6.072**	**0.9668**
Set 6	7.6308	5.8730	0.9651

When multiple climate variables with different lag structures are introduced in Set 4, the model's predictive performance improves further, with substantial reductions in error metrics and an increase in R^2 to 0.9476. This indicates that climate factors have a significant impact on changes in the number of livestock abortions, and their effects often have a certain time lag. By explicitly modelling multi-lag climate features, the model can more effectively capture medium-term fluctuation characteristics, thereby improving overall prediction accuracy. Finally, after further incorporating Google Trends search behaviour features in Set 5, the model achieved optimal performance across all evaluation metrics. Similarly, to explore the effect of Time2Vec encoding of months on the model, we removed the Time2Vec encoding of months from the input parameters under the experimental conditions of Set 5, as in Set 6. The results show that adding the Time2Vec encoding feature of the month has an auxiliary effect on the model's predictive performance. From the results of Set 6, we can infer that the added Time2Vec encoding feature of months, in combination with other input features, allows the model to learn the changing trends, thus positively promoting the model's fitting effect.

5. Mixing-Channel PatchTST with Parallel LSTM Framework

This chapter focuses on the mixed-channel time series prediction model proposed in this thesis, with an emphasis on the modelling framework that combines the Mixing-Channel Patch Time Series Transformer (PatchTST) backbone structure with parallel Long Short-Term Memory (LSTM) channels. In view of the characteristics of multivariate information coupling and the coexistence of long-term trends and short-term fluctuations in livestock abortion time series, this chapter first explains the design motivation and overall modelling idea of the model, and explains the necessity of introducing a hybrid channel structure and parallel sequence modelling mechanism. Subsequently, the core structural components, information flow patterns, and key module functions of the model are systematically described. Finally, the training process and predictive performance of the model are analysed and discussed on national-level data to evaluate its performance in overall time series forecasting tasks, laying the foundation for regional-level experiments and model comparison analyses in subsequent chapters.

5.1 Method Overview

Monthly time series of livestock abortion numbers typically exhibit pronounced seasonal variations alongside periodic short-term fluctuations. On the one hand, abortion numbers follow cyclical patterns associated with annual production cycles and seasonal factors; on the other hand, abnormal climatic conditions and disease outbreaks may trigger substantial short-term fluctuations during specific periods. This temporal characteristic of coexisting long-term structure and localised changes places greater demands on the modelling capabilities of time series forecasting models across multiple time scales.

Existing time series prediction models tend to emphasise different aspects when addressing these characteristics. For example, the Long Short-Term Memory (LSTM) Network is a classic sequence modelling approach that is effective in capturing local sequential dependencies in time series. However, due to its recursive structure, LSTM models are susceptible to information attenuation over long time horizons and complex periodic patterns, and therefore often produce overly smooth prediction outputs [35]. In contrast, Transformer-based models relying on self-attention mechanisms demonstrate strong capabilities in modelling global temporal structures and long-term dependencies, but their modelling focus is more on the overall pattern, and they are relatively less sensitive to local fluctuations in short time scales [36]. Therefore, a single model architecture is often insufficient to simultaneously satisfy modelling requirements across different temporal scales.

Based on the above considerations discussed above, this thesis adopts a parallel dual-channel modelling strategy to process the same set of multivariate time series inputs simultaneously. Both the PatchTST branch and the LSTM branch receive input features with an identical structural format, and the differences in temporal modelling across multiple time scales primarily arise from the inherent characteristics of the model architectures, rather than from any explicit partitioning of input features. Through parallel modelling, the framework is able to extract complementary temporal representations from different sequence-modelling mechanisms, while sharing the same underlying input information. Furthermore, considering the potential correlations between multiple input variables, using only channel-independent modelling may not be sufficient to fully integrate multivariate information. Therefore, a hybrid channel mechanism is introduced in the PatchTST branch, enabling the model to learn the interaction relationships between variables in the channel dimension, thereby improving the overall representation capability of multivariate time series.

5.2 Model Architecture

Based on the characteristics of the dataset and the prediction objectives of this study, this thesis proposes a parallel time series prediction framework. The framework consists of two parallel modelling branches, which are based on PatchTST and LSTM architectures, respectively, to model multivariate time series. By sharing the same set of multivariate input features, parallel modelling and fusion of branch outputs enable the model to simultaneously learn and capture long-term trends, short-term fluctuations, and potential correlations between variables in the input series, thereby improving the stability and robustness of the overall prediction results.

The overall model structure is shown in Figure 5.1. Given a multivariate time series input of length T, which includes historical livestock abortion data, temporal features, climate variables, search behaviour indicators, and livestock population statistics, all input features are first organised into a multivariate sliding-window representation and simultaneously fed into both the PatchTST branch and the LSTM branch for parallel modelling. The PatchTST branch focuses on capturing medium- to long-term time-dependent features and cross-variable correlations through patch-level representation learning and self-attention mechanisms, while the parallel LSTM branch complements this process by modelling the smooth evolutionary dynamics of the sequence and long-term temporal dependencies from a recursive perspective. The two branches produce their respective prediction outputs, which are subsequently combined via a weighted fusion strategy to generate the final prediction result.

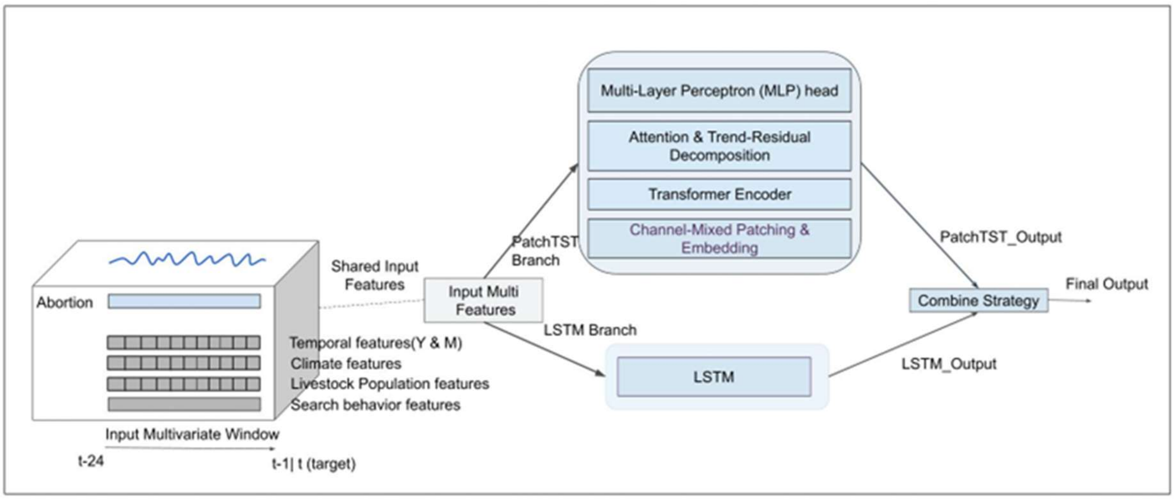


Figure 5.1 The architecture of the parallel PatchTST-LSTM prediction framework is such that the two branches share the same multivariate input, and the prediction results of the two branches are finally fused together for output.

It should be noted that the PatchTST branch adopted in this thesis employs a channel-mixed modelling strategy, rather than the channel-independent approach used in the original PatchTST [36], in order to better accommodate the strong correlations among multi-source heterogeneous variables inherent in the livestock abortion prediction task. At the same time, the parallel LSTM branch is not intended as an alternative or replacement model, but rather serves as a complementary modelling component that operates in conjunction with the PatchTST branch, compensating for the potential limitations of a single modelling paradigm in characterising multi-scale temporal dynamics. The subsequent sections provide a detailed introduction to the specific structure and implementation details of the PatchTST branch, the parallel LSTM branch, and the output fusion strategy used to generate the final prediction results.

5.2.1 PatchTST Branch

In the parallel time series forecasting framework proposed in this study, the PatchTST branch serves as a core modelling component, primarily responsible for learning inter-variable interactions in multivariate time series and characterising medium- to long-term temporal dependencies. Compared to traditional time series methods based on recursive modelling, this branch provides more flexible representation capabilities for modelling complex temporal dynamics and variable correlation patterns through patch-level representation learning and self-attention mechanisms.

This section will systematically explain the internal structure and key modules of the PatchTST branch according to the actual data processing flow. Specifically, it includes the patch construction and embedding method of mixed channels, the Transformer encoder structure, the attention mechanism and trend-residual decomposition module, and the Multi-Layer Prediction Head used to generate prediction results. By introducing each component module in turn, the aim is to clearly present the functional positioning of the PatchTST branch in the overall parallel framework and its role in time-dependent modelling.

5.2.1.1 Channel-Mixed Patching and Embedding Strategy

The original PatchTST model adopts a channel-independent modelling strategy [36]. This means that in the multivariate time series prediction task, each input variable is regarded as an independent channel and is input into an independent Transformer encoder for modelling. Finally, the output results of each channel are combined. While this design offers advantages in computational efficiency and structural simplicity, it may limit the model's ability to effectively model cross-variable interactions in application scenarios where there is a strong correlation among multiple variables.

In multivariate time series prediction tasks, input features are often not independent of each other, but rather form a complex correlation structure through underlying environmental factors or behavioural patterns. Taking livestock abortion prediction as an example, pronounced correlations and co-variation characteristics generally exist among historical abortion records, climate variables, livestock population statistics, and search behaviour indicators. If each variable is treated completely independently in the early stages of modelling, the model may struggle to capture cross-variable dependencies in a timely fashion, thereby limiting its ability to characterise the overall temporal dynamics.

Based on the above considerations, this thesis introduces a "channel-mixed" patch construction strategy into the PatchTST backbone structure, replacing the original channel-independent patching design. Specifically, prior to entering the Transformer encoder, cross-channel information fusion is first performed at the input stage through a one-dimensional convolution operation, allowing interactions among different variables to be explicitly modelled before temporal patching. In this way, richer cross-variable contextual information is provided for subsequent representation learning.

Let the input multivariate time series be represented as:

$$\mathrm{X} \in \mathbb{R}^{B\times C\times T},$$

Where B, C, and T represent the batch size, the number of input variables, and the length of the time series, respectively. To avoid information truncation at the end of the time series due to convolution operations, this thesis uses duplicate-right padding to expand the input sequence, denoted as:

$$\tilde{\mathrm{X}} = Pad_{replicate_{right}}(\mathrm{X}, \mathrm{s}), \tag{1}$$

Here, 's' represents the convolution stride, corresponding to copying the value of the last time step to the right of the time axis. This padding method effectively avoids the loss of boundary information while maintaining the local continuity of the time series. A one-dimensional convolution operation is performed on the expanded input sequence $\tilde{\mathrm{X}}$ to achieve cross-variable local feature fusion:

$$\mathrm{Z} = Conv1d\left(\tilde{\mathrm{X}}; W_c, b_c\right), \qquad W_c \in \mathbb{R}^{d_{model}\times C\times P_{conv}}, b_c \in \mathbb{R}^{d_{model}}, \tag{2}$$

The convolutional kernel slides along the time dimension, but covers all input variables along the channel dimension. Specifically, for the element at time position j in the o-th channel of the output tensor, its computational form is:

$$\mathrm{Z}[b, o, j] = b_c[o] + \sum_{c=0}^{C-1}\sum_{p=0}^{P-1} W_c[o, c, p] \cdot \tilde{\mathrm{X}}[b, c, j \cdot s + p], \tag{3}$$

This operation performs a weighted aggregation of all input variables within a local temporal window, thereby simultaneously modelling temporal dependencies and intervariate interactions at the feature generation stage. The dimensionality of the convolution output tensor Z can be expressed as:

$$\mathrm{Z} \in \mathbb{R}^{B\times d_{model}\times \mathrm{L}_{conv}}, \qquad \mathrm{L}_{conv} = \left\lfloor \frac{\mathrm{T} + s - P_{conv}}{s} \right\rfloor + 1. \tag{4}$$

After cross-channel feature fusion is completed, the convolution output is reorganised along the temporal dimension into a set of non-overlapping temporal patches. Following the concept of temporal slicing [37], let the patch length be denoted as P_{patch}; then the i-th patch can be expressed as:

$$z_p^{(i)} = Z\left[:, :, i \cdot S : i \cdot S + P_{patch}\right], \tag{5}$$

and is projected into a fixed-dimensional embedding space through flattening followed by a linear mapping:

$$u^{(i)} = W_p \cdot Flatten\left(z_p^{(i)}\right) + E_{pos}^{(i)}, \quad W_p \in \mathbb{R}^{d_{model}\times (d_{model}\cdot P)}, \tag{6}$$

where $E_{pos}^{(i)}$ denotes the positional embedding of the i-th patch, which is used to preserve temporal order information. The resulting sequence of patch embeddings is then fed as token inputs to the Transformer encoder. The number of temporal patches is computed as follows:

$$\mathrm{L}_{patch} = \left\lfloor \frac{L_{conv} - P_{patch}}{S} \right\rfloor + 1, \tag{7}$$

here L_{patc} , L_{conv}, P_{pat} , and S represent the number of tokens fed into the Transformer encoder, the length of the convolution output sequence, the length of the patch, and the sliding stride. It should be noted that the convolution kernel length P_{conv} and the patch length P_{patch} are functionally different: the former is used to achieve cross-variable feature fusion within a local time window, while the latter is used for block modelling of time series. By introducing channel-mixed convolution operations before patching, the model retains PatchTST's advantage in capturing local temporal patterns while significantly enhancing its ability to express cross-variable correlation structures in multivariate time series, laying a more robust feature foundation for the subsequent Transformer encoder to model long-term temporal dependencies.

5.2.1.2 Transformer Encoder

After channel-mixed convolution and temporal patching, the model produces a set of temporal patch representations of fixed length. The primary objective of this stage is to model the dependencies between different temporal patches, thereby characterising the long-term dynamic evolution of the time series over longer time scales. To this end, this thesis introduces a Transformer-based encoder module into the PatchTST backbone structure to process the patch representation sequence, fully leveraging the advantages of self-attention mechanisms in capturing long-range temporal dependencies.

Suppose that the patch representation sequence obtained after channel-mixed Patching and embedding mapping is as follows:

$$P = \left[p^{(0)}, p^{(1)}, \dots\ p^{(L_{patch}-1)}\right] \in \mathbb{R}^{B \times L_{patch} \times d_{model}} \tag{8}$$

Each patch vector $p^{(i)}$ is treated as a token in the input sequence and fed into the Transformer Encoder for processing in chronological order.

The Transformer Encoder adopts the standard multi-head attention (MHA) architecture, and its basic form follows the framework proposed by Vaswani et al. [19] in *Attention Is All You Need*. For the h-th attention head, first construct the query, key, and value vectors through linear mapping:

$$Q_h = UW_h^Q,\ K_h = UW_h^K,\ V_h = UW_h^V, \tag{9}$$

where W_h^Q, W_h^K, $W_h^V \in \mathbb{R}^{d_{model} \times d_k}$ are learnable projection matrices, $d_k = d_{\mathrm{model}}/H$, and H denotes the number of attention heads. The output of a single attention head is computed using the scaled dot-product attention mechanism:

$$Attention(Q_h, K_h, V_h) = Softmax(\frac{Q_h K_h^T}{\sqrt{d_k}})V_h. \tag{10}$$

This mechanism adaptively assigns attention weights according to the correlation between different temporal patches, thereby modelling the dependencies between temporal patches. The outputs of all attention heads are then concatenated along the feature dimension, and the final output of the multi-head attention layer is obtained through a linear projection as follows:

$$MHA(P) = Concat(O_1, O_2, \dots, O_H)W^O, \tag{11}$$

where $W^O \in \mathbb{R}^{Hd_k \times d_{model}}$ is the output projection matrix.

Following the attention layer, the model introduces a position-forward network (FFN) to enhance the non-linear representational capacity of features. The FFN consists of two fully connected layers and a GELU activation function, and its form is as follows:

$$FFN(x) = \ \sigma(xW_1 + b_1)W_2 + b_2, \tag{12}$$

where $\sigma(\cdot)$ presents the GELU activation function.

Unlike the pre-norm structure based on Layer Normalisation in the original PatchTST [36], the Transformer Encoder adopts a post-norm design based on Batch Normalisation in this study, that is, normalisation is performed after the residual connection:

$$P' = BatchNorm1d(P + MHA(P)), \tag{13}$$

$$P'' = BatchNorm1d(P' + MHA(P')). \tag{14}$$

This design is primarily based on the characteristics of the dataset used in this study. The task of predicting livestock abortion involves multi-source heterogeneous variables (including the number of abortions, climate indicators, search behaviour, and livestock population statistics). The distribution differences between different features are quite significant, and the model training process usually uses a relatively small batch size. Against this backdrop, the post-norm structure based on Batch Normalisation has shown more robust practical results in stabilising the training process and mitigating gradient fluctuations.

Through the Transformer Encoder module described above, the model can effectively model temporal dependencies at the patch level and integrate the correlation information between different local time segments into a richer intermediate representation. The final output P'' will serve as the input to the subsequent channel attention mechanism and trend-residual decomposition module, providing a basic representation for multi-scale temporal dynamic modelling.

5.2.1.3 Attention Mechanism and Trend–Residual Decomposition

After modelling the patch-level temporal representations using a Transformer encoder, the model can effectively capture the dependencies between different temporal blocks. However, in multivariate time series prediction tasks, the contributions of individual input features to the prediction target are rarely uniform. Especially in livestock abortion prediction scenarios, while features such as historical abortion records, climate lag variables, search behaviour indicators, and livestock population statistics are correlated, their informational relevance varies significantly across different time periods and feature dimensions. If all feature channels are treated equally in the subsequent modelling stages, the representational capacity of key variables may be diluted, thereby limiting the model's ability to characterise complex temporal dynamics.

Based on the above considerations, this thesis introduces the Efficient Channel Attention (ECA) mechanism [38] into the PatchTST backbone structure to perform channel-level recalibration on the features produced by the Transformer encoder. This mechanism adaptively adjusts the weights of different channels, enabling the model to focus more on feature representations that have high informational value for the prediction task.

Specifically, the features are first subjected to global average pooling in the temporal dimension to obtain the overall statistical description of each channel:

$$z_c = \frac{1}{L_{patch}} \sum_{i=1}^{L_{patch}} P''_{i,c}. \tag{15}$$

Subsequently, a one-dimensional convolution is applied to model local inter-channel dependencies along the channel dimension. This process can be represented as follows:

$$w = \sigma\big(Conv1D_k(z_c)\big),$$

where the kernel size k is adaptively determined based on the number of channels, thereby enhancing cross-channel interaction capabilities while maintaining computational efficiency. The channel weight vector w obtained after the sigmoid activation function is applied to the original feature representation on a per-channel basis as follows:

$$P^{ECA} = P'' \odot w, \tag{16}$$

here, $\odot$ represents an element-wise weighted operation. Through the channel attention recalibration process described above, the model can adaptively emphasise input feature channels that have high information relevance to the prediction task, while suppressing redundant or noisy variables, thereby enhancing the discriminative power and stability of multivariate time series representation.

After completing the channel attention adjustment, the model further accounts for the potential effects of jointly modelling long-term trends and short-term fluctuations in multivariate time series. In livestock abortion prediction tasks, time series typically exhibit a relatively stable long-term trend, while also having pronounced seasonal variations and short-term abnormal volatility. If a single unified model is applied to components operating at different temporal scales, the dominant trend component may obscure local variation characteristics, thereby reducing the model's sensitivity to short-term disturbances. Based on this consideration, a Trend-Residual Decomposition module is incorporated into the PatchTST backbone to explicitly decompose the feature sequences processed by the ECA mechanism. Specifically, a moving average operator is first applied to smooth the feature sequence, in order to extract the long-term trend component:

$$P_{trend} = MovingAvg(P^{ECA}). \tag{17}$$

Subsequently, the residual component characterising short-term fluctuations and local disturbances is obtained by subtracting the trend component from the original sequence:

$$P_{residual} = P^{ECA} - P_{trend}, \tag{18}$$

among them, P_{trend} mainly reflects the long-term trend of livestock abortion numbers over time, while $P_{residual}$ characterizes the short-term fluctuations in the time series, including seasonal peaks, local anomalies, and transient disturbances that are closely associated with lagged climate variables or variations in search behaviour.

By introducing a channel attention mechanism and integrating it with trend-residual decomposition, the model can adaptively emphasise variable-level information that is more critical to the prediction tasks, while explicitly separating long-term evolutionary trends from short-term fluctuation components along the temporal dimension. This hierarchical modelling strategy enhances the model's ability to characterise multi-scale temporal dynamics and provides clearer and more stable feature representations for subsequent modelling of different temporal components, thereby improving the stability and robustness of the overall prediction.

5.2.1.4 Multi-Layer Prediction Head

After completing the trend-residual decomposition, the model obtains feature representations corresponding to the long-term trend component P_{trend} and the short-term residual component $P_{residual}$. Since these two components differ substantially in temporal scale and statistical properties, it may be difficult to simultaneously and effectively model both stable long-term changes and local short-term fluctuations if a single unified prediction mapping method is used. Based on this consideration, this thesis designs independent Multi-Layer Prediction Heads for the trend component and the residual component, respectively, so as to enable differentiated mapping and prediction modelling for distinct temporal components.

For any component P (where P can represent P_{trend} or $P_{residual}$), its input feature tensor can be expressed as:

$$P \in \mathbb{R}^{B \times C \times d_{model}},$$

where B represents the batch size, C is the number of feature channels, and d_{model} is the feature embedding dimension. To facilitate subsequent fully connected mapping, this tensor is first flattened, converting it into a one-dimensional feature vector:

$$P^{(1)} = Flatten(P). \tag{19}$$

Subsequently, the flattened feature vectors are passed through the first fully connected layer, and the GELU activation function is applied to enhance the model's nonlinear modelling capability:

$$H = \sigma\left(W_1 P^{(1)} + b_1\right), \tag{20}$$

where W_1 and b_1 represent the weight matrix and bias term of the first fully connected layer, respectively, and $\sigma(\cdot)$ represents the GELU activation function. The purpose of this layer is to map high-dimensional features into a latent space, thereby extracting more discriminative combined features.

Based on this representation, a second fully connected layer is applied to further project the hidden features into the prediction space, producing the predicted value for the corresponding component:

$$\hat{y} = W_2 H + b_2, \tag{21}$$

where W_2 and b_2 are the learnable parameters of the second fully connected layer.

Accordingly, the trend component and the residual component generate their respective prediction outputs through their dedicated Multi-Layer Prediction Heads:

$$\hat{y}_{trend} = MLP(P_{trend}), \tag{22}$$

$$\hat{y}_{residual} = MLP(P_{residual}), \tag{23}$$

Finally, the output of the PatchTST backbone branch is obtained by summing the trend prediction and the residual prediction:

$$\hat{y}_P = \hat{y}_{trend} + \hat{y}_{residual}. \tag{24}$$

This separated prediction structure enables the model to independently model long-term smooth trends and short-term disturbance information. While preserving a relatively simple network structure and manageable computational complexity, it enhances the model’s capacity to characterize features across different temporal scales and improves the stability of the prediction results as well as the interpretability of the learned representations.

5.2.2 Parallel LSTM Branch

Although the PatchTST backbone demonstrates strong capability in modelling local temporal patterns and cross-variable interactions in multivariate time series, its patch-based modelling strategy may, to some extent, weaken the representation of continuous and smoothly evolving long-term temporal dependencies. To address this potential limitation, a parallel LSTM branch is incorporated into the overall forecasting framework, providing complementary modelling of global temporal structures and long-term evolutionary trends from a recurrent perspective.

Long Short-Term Memory (LSTM) networks effectively mitigate the vanishing gradient problem commonly encountered in traditional recurrent neural networks when learning long sequences through gating mechanisms, enabling the retention of critical information over extended time horizons. As a result, LSTM has been widely adopted in time series forecasting applications [16,39,40]. In addition, LSTM has shown strong stability in characterising long-term dependencies and low-frequency variation trends [35], making it an effective complement to Transformer-based modelling structures and forming a synergistic relationship with the PatchTST branch.

In this study, the LSTM branch shares the same raw multivariate time series inputs as the PatchTST branch. The input tensor to the LSTM branch can be expressed as:

$$\mathrm{X} \in \mathbb{R}^{B\times C\times T},$$

where B, C, and T denote the batch size, the number of input variables, and the length of the time series, respectively. Since LSTM follows a sequence modelling paradigm that is centred on the temporal dimension, the input tensor is first rearranged along its dimensions and transformed into a time-first input format:

$$X_{LSTM} = \mathrm{Permute}(\mathrm{X}) \in \mathbb{R}^{B\times T\times C}. \tag{25}$$

The rearranged input sequence is fed into a stacked LSTM network for temporal modelling. Each LSTM layer characterises the temporal dependencies in the sequence through a time-stepwise state update mechanism. Let the hidden state at time step t be:

$$H_t = StackedLSTM(X_{LSTM})_t, \qquad \mathrm{t} = 1, \cdots, \mathrm{T}, \tag{26}$$

During the prediction phase, the hidden state H_T corresponding to the final time step of the sequence is selected as the global representation of the entire input time window by the LSTM branch:

$$U^L = H_T \in \mathbb{R}^{B \times d_L}, \tag{27}$$

Where d_L represents the dimensionality of the top-level hidden state of the LSTM, which aggregates temporal contextual information from the past observations to the current time step, thereby characterising the long-term evolutionary behaviour of the sequence.

To project the hidden representation produced by the LSTM branch into the prediction space, a fully connected layer is applied at its output stage to generate the numerical prediction corresponding to the target forecasting horizon:

$$z^L = W_L U^L + b_L \in \mathbb{R}^{B \times pred_{len}}. \tag{28}$$

Since the LSTM branch does not explicitly differentiate among input variable channels, this thesis replicates and expands the prediction along the channel dimension to ensure structural consistency with the output of the PatchTST branch. Therefore, we obtain:

$$\hat{y}_L = RepeatChannel(z^L) \in \mathbb{R}^{B \times C \times pred_{len}}. \tag{29}$$

Overall, the LSTM branch, through recursive state updates over continuous time, focuses on characterising the global evolution of the sequence and smooth, low-frequency temporal dynamics. In contrast, the PatchTST branch, based on temporal patch structures, is better suited to capturing longer-range temporal dependencies and modelling interaction patterns and non-stationary variations among multiple variables. These two modelling approaches complement each other from different time modelling perspectives, jointly enhancing the model's overall representational capacity for complex multivariate time series.

5.2.3 Fusion Strategy

After the parallel modelling phase is completed, the prediction outputs from different modelling branches must be integrated to produce the final model output. Since the branches differ in their modelling perspectives and temporal feature characterisation mechanisms, their prediction results are complementary to a certain extent. Therefore, introducing a reasonable fusion strategy helps to fully utilise the information learned by different branches.

considering both stability and interpretability, this thesis adopts a linear weighted fusion strategy to combine the prediction results from different branches. Let the predicted output of the PatchTST branch be $\hat{y}_P$ and the predicted output of the LSTM branch be $\hat{y}_L$, then the final prediction result of the model is defined as:

$$\hat{y} = \alpha \hat{y}_P + (1 - \alpha) \hat{y}_L, \tag{30}$$

Here, $\alpha \in [0,1]$ is the fusion weight parameter, used to control the relative contribution of the two branches in the final prediction. When α is large, the model output is more biased towards the prediction results of the PatchTST branch; conversely, when α is small, the LSTM branch contributes more substantially to the overall prediction. Linear weighted fusion strategy is widely used in ensemble learning and time series prediction research due to its simple structure and strong interpretability. This method can effectively integrate complementary information learned by different models, improving the robustness and generalisation ability of the prediction results while maintaining the stability of the model structure [41].

In this study, the fusion weight α was not fixed in advance, but was determined by experimental tuning on the validation set to determine the optimal branch combination ratio. The final fusion prediction result $\hat{y}$ is used as the output of the model for subsequent loss-function calculation and performance evaluation.

5.3 Performance Evaluation and Analysis

This chapter evaluates the training behaviour and predictive performance of the Mixing-Channel PatchTST with Parallel LSTM model using national-level livestock abortion data. The analysis considers multiple aspects of model behaviour, including convergence characteristics and optimisation dynamics during training, which are examined to assess training stability. Model predictive capability is then evaluated by comparing predicted outputs with observed values and by analysing standard quantitative performance metrics. In addition, the effects of model structure, input feature composition, and fusion parameter settings on prediction outcomes are investigated through module-level analysis and sensitivity experiments. Together, these analyses provide insight into the key factors influencing model performance and robustness.

5.3.1 Training Convergence and National-Level Prediction Performance

Figure 5.2 shows the evolution of training and validation loss during the training of the Mixing-Channel PatchTST with the Parallel LSTM model. Both loss curves decrease rapidly in the initial training stage, suggesting that the model learns the dominant temporal structure of the input time series within a limited number of training iterations.

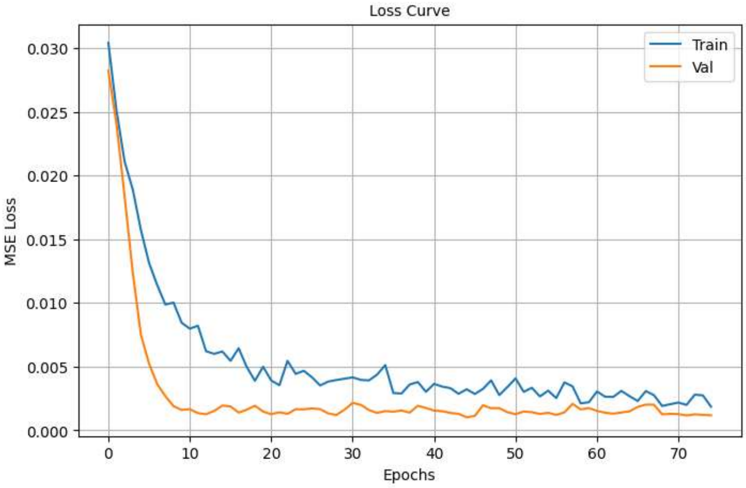


Figure 5.2 Training and validation loss curves of the Mixing-Channel PatchTST with Parallel LSTM model on national-level data, illustrating the convergence behaviour of the model over training epochs.

As the number of training iterations increases, the training loss and validation loss continue to decrease and gradually stabilise in the early stages, showing good convergence characteristics overall. It is worth noting that throughout the training process, the two loss curves maintained a relatively close trend, without any significant separation or drastic fluctuations. This phenomenon indicates that the model did not experience significant overfitting during the optimisation process and was able to maintain good generalisation consistency between the training and validation sets.

From the perspective of optimisation stability, the loss curve smoothly across iterations, with no noticeable oscillations or signs of gradient instability, suggesting that the current model structure and training strategy are numerically optimisation level. The collaborative modelling approach of the PatchTST branch and the parallel LSTM branch enables the model to simultaneously characterise long-term temporal dependencies and local dynamic changes, which, to some extent, alleviates the optimisation instability risk that may be caused by a single modelling mechanism. This thesis compares the predicted results on the test set with the actual monthly livestock abortion numbers to further analyse the model's predictive ability at the national scale. The results are shown in Figure 5.3, The prediction time range covers 2022 to 2023 to reflect the model's predictive performance on the overall time scale.

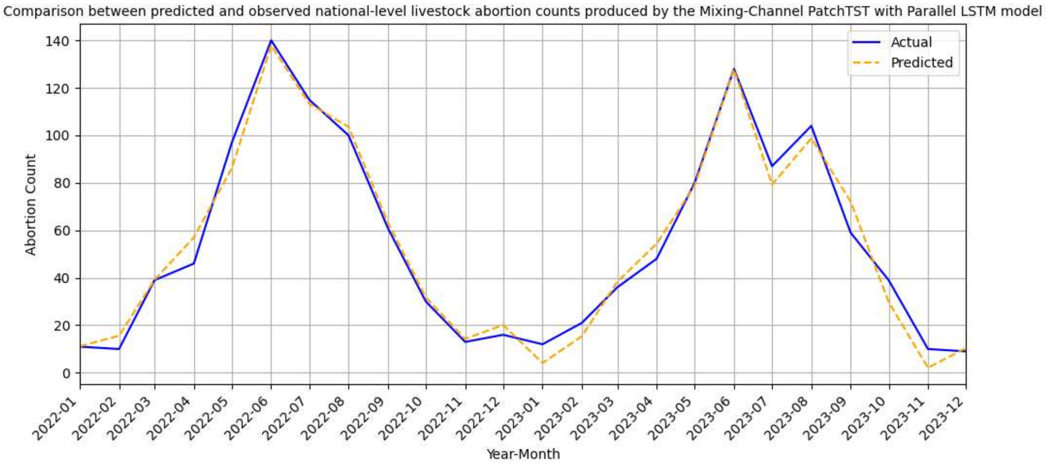


Figure 5.3 Comparison between predicted and observed national-level livestock abortion counts produced by the Mixing-Channel PatchTST with Parallel LSTM model over the evaluation period.

From an overall trend perspective, the model-predicted and actual sequences are highly consistent in their temporal evolution and can track the periodic fluctuations in the number of livestock abortions associated with seasonal changes. During both the annual rising and declining phases, the prediction results accurately capture the direction of change without noticeable systematic bias, indicating that the model possesses a stable capability to characterise long-term temporal dependencies at the national scale.

In addition, the predicted and actual sequences are largely aligned in timing at seasonal peaks and troughs, indicating that the model can effectively identify critical temporal turning points. Although there are still some deviations between the predicted and actual values in certain peak months, these deviations are mainly reflected in the peak intensity rather than the peak timing, indicating that the model maintains robust performance in seasonal structure modelling. Moreover, during intervals with relatively low abortion counts and gradual temporal variation, the predicted curve closely overlaps with the observed curve, demonstrating the model's high stability in representing background trends and low-volatility periods. This characteristic contributes to reducing overall prediction error and enhances the consistency of long-term forecasting results

Therefore, the national-level visualisation results show that the model can simultaneously account for long-term trends and seasonal fluctuations within a unified prediction framework, and is capable of producing stable and continuous predictions of the overall evolution of monthly livestock abortion numbers.

5.3.2 Model Comparison and Quantitative Performance Evaluation

To more comprehensively evaluate model performance, this study selects seven mature and representative models as the baseline models, including the standard RNN [52], the standard LSTM [35], the channel-independent PatchTST model [36], CNN-enhanced LSTM [53], SARIMA-LSTM hybrid model [54], Seq2Seq-based LSTM model [55], and the LSTM Model with an attention mechanism. Under identical experimental settings, these models are applied to national-level livestock abortion data, and their predictive performance is systematically compared with that of the Mixing-Channel PatchTST with Parallel LSTM model. These selected baselines cover both fundamental deep learning time series models and hybrid approaches combining deep learning with traditional statistical techniques, and are representative in terms of model complexity, modelling mechanisms, and nonlinear learning capability.

The quantitative performance results of all models are summarised in Table 5.1. The Mixing-Channel PatchTST with Parallel LSTM model achieved low RMSE and MAE values of 6.0062 and 4.6993 respectively over the full evaluation period, while attaining the highest coefficient of determination with R^2 values of 0.9784 among all compared models. These results indicate that the prediction deviations are effectively controlled and that the model exhibits a high degree of consistency with the observed temporal fluctuations in livestock abortion numbers. Among the baseline models, the SARIMA-LSTM hybrid model shows relatively weak fitting performance, which is likely attributable to its limited ability to capture complex nonlinear dynamics in the input series. The channel-independent PatchTST model performs slightly worse than the standard LSTM, indicating that ignoring cross-variable interactions constrains its ability to represent seasonal variation patterns in multivariate time series data. Moreover, among models incorporating sequence enhancement mechanisms, the Seq2Seq-based LSTM model outperforms the CNN-enhanced LSTM and standard RNN models, but still exhibits noticeable biases during periods of pronounced seasonal variation. Among all baseline models, the LSTM model with an attention mechanism achieves a relatively high R^2 value of 0.9668, indicating its ability to capture the overall temporal pattern of livestock abortion numbers and maintain close agreement between predicted and observed sequences. However, its performance is still marginally lower than that of the Mixing-Channel PatchTST with Parallel LSTM model.

Table 5.1 RMSE, MAE, and R^2 results of eight models on the national-level test dataset. Bold values indicate relatively better results.

Model	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Standard RNN	12.4517	9.6038	0.9071
Standard LSTM	11.1951	9.5328	0.9249
Channel-Independent PatchTST	14.2942	11.4462	0.8776
CNN-enhanced LSTM	11.9669	9.2342	0.9142
SARIMA-LSTM hybrid	26.399	20.5643	0.5825
Seq2Seq-based LSTM	10.6394	8.3285	0.9322
LSTM model with an attention mechanism	7.4426	6.072	0.9668
Mixing-Channel PatchTST with Parallel LSTM	**6.0062**	**4.6993**	**0.9784**

As shown in Figure 5.4, the LSTM model with an attention mechanism tends to underestimate livestock abortion counts during the two peak months (around June) when compared with the Mixing-Channel PatchTST with

Parallel LSTM model. This underestimation indicates limitations in its ability to represent pronounced seasonal fluctuations. In contrast, most other baseline models show an opposite pattern, with peak values being systematically overestimated and predictions decreasing too slowly during the subsequent decline period. Such behaviour reflects a high sensitivity to dominant seasonal trends, which leads to reduced stability in their forecasts. The Mixing-Channel PatchTST with Parallel LSTM model exhibits a different predictive pattern, tracking the temporal evolution of livestock abortion counts in New Zealand more consistently and remaining closely aligned with observed seasonal variations. This result further demonstrates that the proposed dual-channel framework, relative to single-channel modelling approaches, is better suited to learning complex seasonal structures in multivariate time-series data.

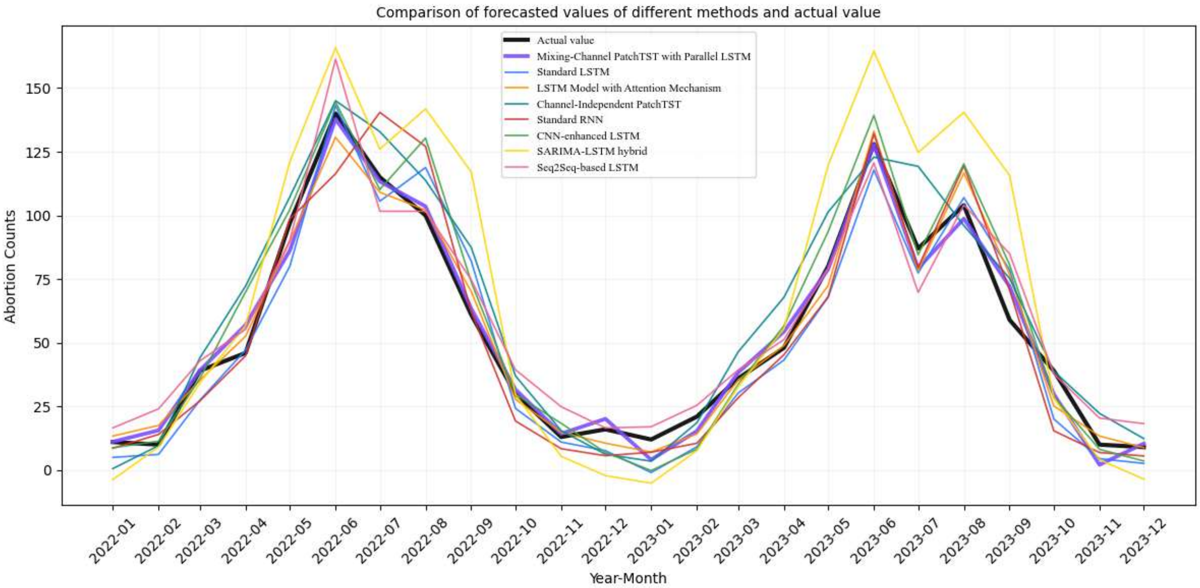


Figure 5.4 Comparison between predicted and observed national-level livestock abortion counts produced by the Mixing-Channel PatchTST with Parallel LSTM and other models over the evaluation period.

In summary, the Mixing-Channel PatchTST with Parallel LSTM model demonstrates relatively robust predictive characteristics in national-level prediction tasks, achieving consistent and well-balanced performance across quantitative error metrics as well as faithful trend representation.

5.3.3 Ablation Analysis of Model Components

To more explicitly analyse the individual contribution of each module to the model, this section conducts experiments on the progressive integration process of the main components of the model, under the premise of consistent input features, training strategies, and evaluation metrics. The experimental configuration is shown in Table 5.2. Based on the Mixing-Channel PatchTST backbone structure, this thesis incrementally introduces channel attention, trend-residual decomposition, multi-layer prediction heads, and parallel LSTM branches to construct a set of structural variants with progressively increasing model complexity, in order to observe how prediction performance evolves with module integration.

Table 5.2 Incremental module configuration.

Module Set	**Model Structure Description**
M1	Mixing-Channel PatchTST backbone
M2	M1 with Efficient Channel Attention (ECA)
M3	M2 with Trend-Residual Decomposition
M4	M3 with Multi-Layer Prediction Head
M5	M4 with Parallel LSTM branch

The experimental results are shown in Table 5.3. Based on the results, we can observe that as functional modules are gradually introduced, the model shows a steady improvement trend in indicators such as RMSE, MAE, and R^2, which suggests that the progressive structural enhancement process is reasonable in the overall direction.

Table 5.3 The performance of the Mixing-Channel PatchTST with Parallel LSTM model on module configurations was evaluated using RMSE, MAE, and R^2. Bold values indicate relatively better results.

Module Set	**RMSE**	**MAE**	$\boldsymbol{R^2}$
M1	11.8792	10.3934	0.9155
M2	9.6888	8.5762	0.9438
M3	9.3231	8.1414	0.9479
M4	8.9864	7.2229	0.9516
M5	**6.0062**	**4.6993**	**0.9784**

Furthermore, the basic model (M1) is able to capture the main trends in time series, but it still has certain limitations in error robustness. Building upon this, model M2 incorporates a channel attention mechanism and achieves consistent improvements across all evaluation metrics. These results suggest that the ECA module contributes to more effective differentiation among input variables, which is reflected in improved predictive performance. When the trend-residual decomposition module (M3) is introduced, the model performance is modestly but consistently improved. This demonstrates that, with the support of the existing channel attention mechanism, explicitly separating long-term trends and short-term fluctuations in the time series helps the model to more precisely characterise the sequence structural features. Subsequently, a multi-layer prediction head (M4) with stronger nonlinear representation capacity was introduced, and the model prediction error was further reduced, indicating that enhancing the output mapping structure based on high-dimensional time series representation helps to mitigate prediction bias. Finally, these introduction of a parallel LSTM branch (M5) results in a marked improvement in model performance, with a substantial reduction in RMSE and MAE, and an R^2 increase to 0.9784. This indicates that the parallel sequence modelling branch plays a role in capturing local temporal dynamics and short-term dependencies, effectively complementing the backbone model.

Therefore, the progressive experimental results demonstrate the rationality of the structural design to a certain extent, and provide the necessary experimental basis for subsequent sensitivity analysis of input variables and fusion parameters under a fixed model structure.

5.3.4 Sensitivity Analysis of Input Variables

The six feature sets adopted in this section follow the same hierarchical expansion logic as described in Section 4.4.3, progressing from time-dependent information to structural, environmental, and behavioral variables. Historical abortion counts serve as the fundamental time-series input, followed by livestock population size as a structural scale constraint, lagged climate variables to capture delayed environmental effects, and Google Trends indicators as external behavioral signals. This staged design enables evaluation of the incremental contribution of different information categories under a controlled feature expansion framework.

While maintaining consistency in model structure, training strategy, and evaluation metrics, we further designed a series of sensitivity analysis experiments based on stepwise feature introduction of input variables. The experimental settings are shown in Table 5.4. This set of experiments is to further examine the impact of different input variables on the model's predictive performance.

Table 5.4 The incremental feature set configuration for input variables gradually incorporates features such as month, livestock population statistics, climate, and public search behaviour.

Feature Set	**Description**
Set 1	Historical abortion counts with month one-hot encoding
Set 2	Set 1 with Time2Vec-based month encoding
Set 3	Set 2 with aggregated livestock population features
Set 4	Set 3 with lagged climate variables (six features with 1-, 2-, and 3-month lags)
Set 5	Set 4 with Google Trends search behaviour features (seven keywords)
Set 6	Set 5 without Time2Vec-based month encoding

The results are shown in Table 5.5. Our observations show that when using only historical abortion numbers and monthly information (Set 1), the model is able to capture the overall trend of the time series well and achieve relatively robust predictive performance. This indicates that the livestock abortion time series itself has strong temporal correlation and seasonal structure, which is an important source of basic information for the model to make predictions. However, RMSE and MAE remain at comparatively high levels, indicating that relying solely on historical information and time-location characteristics is insufficient to fully explain the detailed changes contained in monthly fluctuations. When adding Time2Vec encoding to the monthly representation (Set 2) did not significantly improve the overall model performance; some error indicators even slightly increased. This may indicate that the newly added Time2Vec encoding of the month did not enable the model to learn more informative features, but instead introduced noise. It may also indicate that this encoded feature may need to be complemented by introducing more input features. Therefore, when the overall livestock size information was further incorporated (Set 3), the model prediction performance improved substantially, with both RMSE and MAE decreasing and R^2 increasing to 0.9489. This indicates that population size has a positive effect on the model.

Table 5.5 The performance of the Mixing-Channel PatchTST with Parallel LSTM model on different input features was evaluated using RMSE, MAE, and R^2. Bold values indicate relatively better results.

Feature Set	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Set 1	10.2283	7.8853	0.9373
Set 2	10.5223	8.2522	0.9337
Set 3	9.2341	7.7376	0.9489
Set 4	8.0885	6.0072	0.9608
Set 5	**6.0062**	**4.6993**	**0.9784**
Set 6	8.9173	7.2202	0.9524

When multiple climate variables with different time lags were introduced into Set 4, the model's predictive performance further improved, the error index decreased significantly, and R^2 increased to 0.9608. This suggests that climate factors have a notable impact on changes in livestock abortion numbers, and their effects often have a certain time lag. By explicitly modelling multi-lag climate characteristics, the model can more effectively capture medium-term fluctuation characteristics, thereby improving overall prediction accuracy. Finally, after further introducing Google Trends search behaviour features in Set 5, the model achieved optimal performance on all evaluation metrics. These results indicate that search behaviour features can reflect potential social attention or risk changes to a certain extent, providing valuable supplements to the model and effectively complementing traditional structural and environmental variables. To further explore the effect of Time2Vec encoding of months on the model, we removed the Time2Vec encoding of months from the input parameters under the experimental conditions of Set 5. The results of Set 6 show that adding the Time2Vec encoding feature of months has an auxiliary effect on the model's predictive performance.

The analysis results of this experiment show that the improvement in model performance does not simply depend on increasing the number of features, but is closely related to the type of information contained in the introduced features. By rationally selecting and combining multi-source features, the model's ability to characterise complex time series dynamics has been significantly enhanced.

5.3.5 Sensitivity Analysis of Fusion Parameters

In the Mixing-Channel PatchTST with Parallel LSTM model, the fusion parameter α controls the relative contribution ratio of the PatchTST branch and the parallel LSTM branch in the final prediction stage. This parameter is not fixed in advance but is an adjustable hyperparameter that is fine-tuned through experimental analysis. This parameter is not predefined but is used as an adjustable hyperparameter, optimised through experimental analysis. To evaluate the impact of the fusion parameter α on the model's predictive performance and analyse the stability of the two-branch structure under different weight configurations, this section presents a systematic sensitivity analysis experiment focusing on this parameter.

Based on the functional positioning of each branch in the model structure design, the PatchTST branch is mainly used to characterise long-term time-dependent structures, while the LSTM branch focuses on capturing local dynamic change features. Based on this consideration, this thesis limits the value range of the fusion parameter α to [0.39,0.51]. This range covers typical configurations with relatively balanced branch weights and slight bias towards a single branch, while avoiding the impact of extreme weight settings on the integrity of the model structure and the interpretability of the fusion. Within this range, multiple representative fusion weights were selected for experiments.

Table 5.6 The performance of the Mixing-Channel PatchTST with Parallel LSTM model under different α fusion weight settings was evaluated using RMSE, MAE, and R^2. Bold values indicate relatively better results.

Fusion Value	**RMSE**	**MAE**	R^2
0.51	6.0785	4.7222	0.9779
0.50	6.0555	4.6813	0.9780
0.47	6.0116	4.6673	0.9783
0.45	**6.0062**	**4.6993**	**0.9784**
0.42	6.0448	4.8152	0.9781
0.39	6.1485	4.912	0.9774

The results are shown in Table 5.6. The experimental results show that the fusion parameter α has a certain impact on the model's prediction performance. When α=0.45, the model achieves optimal or near-optimal performance in terms of RMSE, MAE, and R^2, with RMSE decreasing to 6.0062 and R^2 increasing to 0.9784. These results indicate that, under the current experimental settings, appropriately increasing the weight of the LSTM branch in the fusion stage helps the model more effectively characterise short-term fluctuations and local changes, thereby improving overall prediction accuracy.

Further observation revealed that within the range of $\alpha \in [0.45,0.47]$, the model performance exhibited relatively small fluctuations, with all evaluation metrics remaining at a high level. This indicates that the two-branch model fusion structure possesses strong stability and robustness within this parameter range. In contrast, when the fusion weights were further skewed towards the PatchTST branch (e.g., $\alpha = 0.51$), or when the LSTM branch weights were excessively reduced (e.g., $\alpha = 0.42$ and $\alpha = 0.39$), the prediction error showed a moderate upward trend.

In summary, the analysis shows that the fusion parameter α has a measurable impact on the model's predictive performance within a reasonable range, but the overall performance changes exhibit a relatively smooth trend without pronounced performance oscillations. With different weight configurations, the model shows varying degrees of emphasis between long-term trend modelling and short-term dynamic characterisation. When the contribution ratios of the two branches are relatively balanced, the overall predictive performance of the model is more stable.

6. Comparison and Discussion

Building on the national-level experimental analysis presented in the previous chapter, this chapter further extends the scope of investigation to the regional level in order to examine differences in the temporal evolution characteristics and predictive modelling performance of livestock abortion counts across regions. Compared with the aggregated national-level series, regional-level data exhibit more pronounced heterogeneity in terms of sample size, fluctuation magnitude, seasonal intensity, and the distribution of extreme values. Such heterogeneity reflects underlying differences in climatic conditions, livestock production structures, and management practices across regions, and poses greater challenges to the generalisation capability and robustness of time series forecasting models.

Specifically, this chapter, based on the cleaned original dataset, partitions the national data into 7 independent regional subsets according to administrative regions. Prediction experiments are then conducted on each regional dataset using two modelling methods: one is the Mixing-Channel PatchTST with Parallel LSTM model, the other is the LSTM model with an attention mechanism, which performs better in the baseline models, to observe and analyse the model stability.

The purpose of conducting regional experiments is not only to focus on the predictive performance of the models, but also to analyse the differences between regional data through model results. This chapter first provides a descriptive analysis of the temporal distribution characteristics of livestock abortion data at the regional level, focusing on comparing the differences in long-term trend evolution, seasonal fluctuation amplitude, and frequency of extreme peaks in different regions. Then, it analyses the index results, compares the predictive performance of different models in various regions, and discusses the potential impact of regional heterogeneity on model performance differences, aiming to provide more interpretable experimental evidence for the applicability of multiple models under complex, multi-source time-series data conditions.

Furthermore, following the regional-level comparative analysis of predictive performance and data heterogeneity, the forecasting horizon, feature availability conditions, and the model’s generalisability under varying data environments are further examined. From the perspective of operational boundaries and generalisation prerequisites, the practical applicability of the proposed framework is subsequently clarified.

6.1 Characteristics of Regional Data

Compared to national-level data, regional time series exhibit more significant differences in sample size, volatility intensity, and outlier distribution. These differences not only reflect objective variations in climate conditions, livestock structure, and farming scale across different regions but may also directly impact the training stability and predictive performance of forecasting models.

As shown in Figure 6.1, the figure illustrates the temporal variation of actual livestock abortions in various regions from 2004 to 2023. Overall, most regions exhibit a relatively clear annual cyclical fluctuation, with abortion numbers typically peaking in specific months each year, indicating that seasonal factors remain a key driver of abortion variation at the regional scale. However, there are notable differences in the amplitude of fluctuations and the frequency of peak occurrences in different regions, reflecting the heterogeneity between regions in terms of natural environment and production conditions.

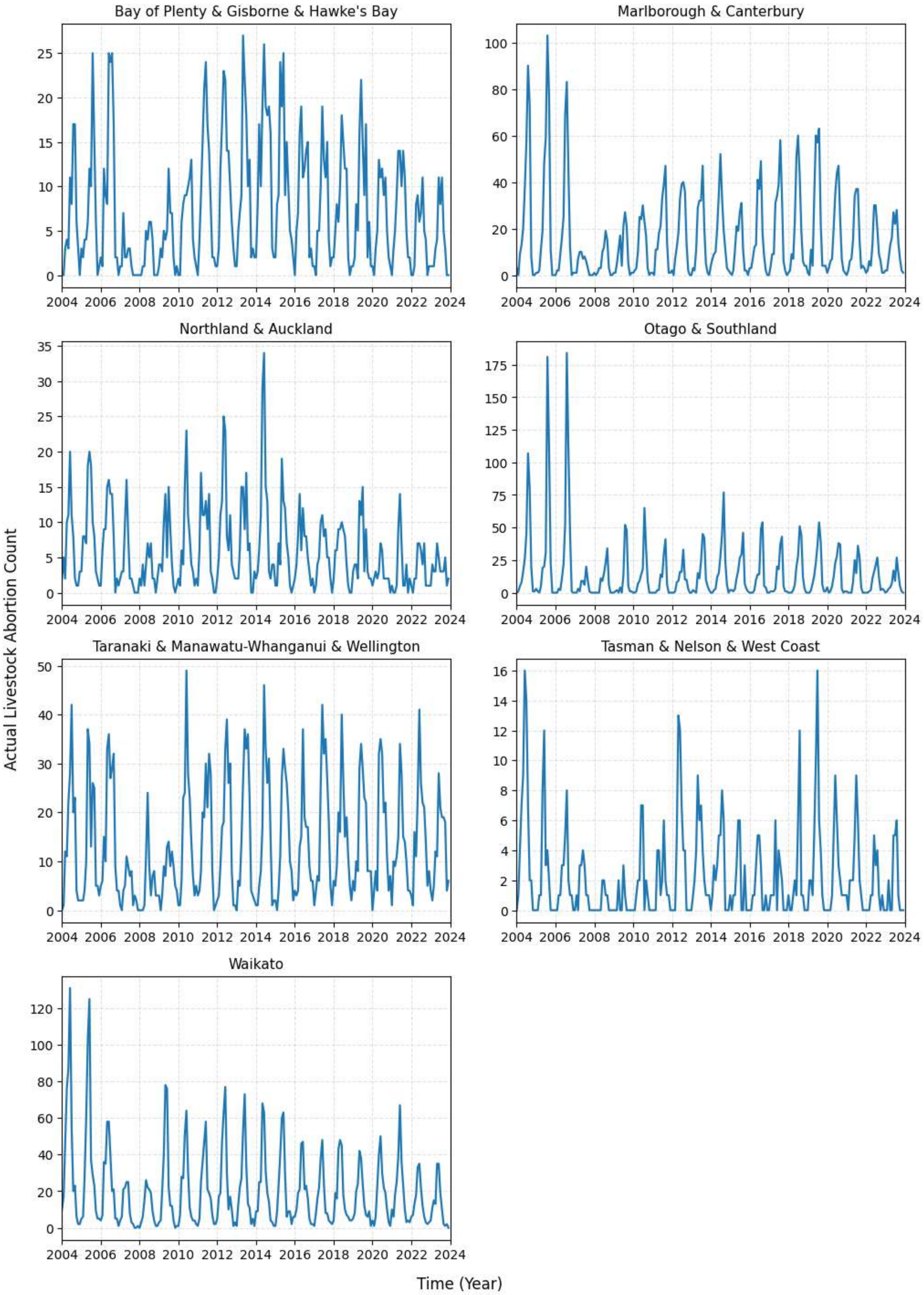


Figure 6.1 The data shows the number of livestock abortions per month in different regions between 2004 and 2023.

Further analysis revealed that the peak abortion numbers in regions such as Otago, Southland and Waikato were substantially higher than in other areas, with some years exhibiting unusually large peaks. This indicates that while the time series data in these regions exhibit a significant seasonal structure, they are also accompanied by irregular extreme variations, placing higher demands on the robustness and anomaly handling capabilities of the predictive models. In contrast, regions such as Tasman, Nelson, and the West Coast have lower overall abortion numbers, relatively flat sequence fluctuations, fewer peak frequencies, and more concentrated data distribution, but the effective signals available for model learning are comparatively limited.

In regions with a moderate number of livestock abortions (such as Marlborough, Canterbury, Taranaki, Manawatu-Whanganui, and Wellington), the number of livestock abortions exhibits a relatively stable cyclical pattern. These regions maintain a clear seasonal structure while having a relatively low frequency of extreme outliers and more stable overall changes, providing comparatively balanced training sample conditions for the model. We can observe that the trend changes in different regions do not show a fully consistent direction. Some regions exhibited high peak levels in the early stages of the study, but gradually stabilised in the later stages; other regions have demonstrated an increasing trend in volatility in recent years. This suggests that aggregating data at the national level may obscure regional patterns of change, while regional-scale analysis would be more valuable for studying livestock abortion numbers and risk mitigation.

Furthermore, from a modelling perspective, due to substantial differences in sample size, fluctuation structure, and outlier distribution across different regions, the predictive performance of the model at the regional level is more sensitive to changes in data distribution. Therefore, regional-scale experiments can more clearly reflect the model's adaptability under different data conditions. If the model maintains relatively stable predictive performance in most regions, it indicates that the learned time-series representation has a certain degree of universality; conversely, if the model's performance shows marked differences across different regions, it may suggest that the model has a degree of dependence on the data characteristics of specific regions.

6.2 Overview of Regional-level Prediction Results

This study applied an LSTM Model with an attention mechanism and a mixing-channel PatchTST with a Parallel LSTM to regional datasets and summarised the prediction results for each region in Table 6.1. Overall, the prediction performance of the two models varied across different regions, and neither achieved performance comparable to the national-level results. This further indicates that regional-level forecasting tasks are inherently more challenging than national-level prediction and impose higher demands on model representation capacity and generalisation capability.

Table 6.1 Under the same experimental conditions, the RMSE, MAE, and R^2 results of the Mixing-Channel PatchTST with Parallel LSTM and the LSTM model with an attention mechanism on regional test data were compared. Bold values indicate relatively better results.

Merged Regions	**Model**	**RMSE**	**MAE**	$\boldsymbol{R^2}$
Bay of Plenty, Gisborne, and Hawke's Bay	LSTM model with an attention mechanism	2.3717	1.8838	0.6091
	Mixing-Channel PatchTST with Parallel LSTM	**2.1738**	**1.586**	**0.6716**
Marlborough and Canterbury	LSTM model with an attention mechanism	6.453	4.7001	0.5783
	Mixing-Channel PatchTST with Parallel LSTM	**3.5908**	**2.5456**	**0.8694**
Northland and Auckland	LSTM model with an attention mechanism	1.4566	1.2357	0.5646
	Mixing-Channel PatchTST with Parallel LSTM	**1.4334**	**1.0094**	**0.5784**
Otago and Southland	LSTM model with an attention mechanism	5.9868	4.0877	0.493
	Mixing-Channel PatchTST with Parallel LSTM	**4.157**	**3.1345**	**0.7555**
Taranaki, Manawatu-Whanganui and Wellington	LSTM model with an attention mechanism	4.4762	3.1191	0.7909
	Mixing-Channel PatchTST with Parallel LSTM	**3.3571**	**2.5471**	**0.8824**
Tasman, Nelson, and West Coast	LSTM model with an attention mechanism	1.2678	1.0472	0.5894
	Mixing-Channel PatchTST with Parallel LSTM	**1.2244**	**0.8341**	**0.6171**
Waikato	LSTM model with an attention mechanism	3.774	3.0444	0.8898
	Mixing-Channel PatchTST with Parallel LSTM	**3.3931**	**2.5579**	**0.9109**

From the distribution of error-based metrics, the Mixing-Channel PatchTST with Parallel LSTM achieves lower RMSE and MAE values in most regions, accompanied by generally higher R^2 scores compared with the LSTM model with an attention mechanism. This suggests that the former model exhibits more stable predictive performance at the regional level. However, it should be noted that absolute error magnitudes such as RMSE and MAE are partially influenced by the overall scale of abortion counts in each region. As a result, these metrics alone are insufficient for directly comparing model performance across regions with substantially different data distributions.

It is worth noting that when RMSE is low, R^2 is often also low. However, since regional datasets are generally small in size, R^2 is more appropriate as a metric when evaluating the prediction results of data models at the regional level. This is because R^2 focuses more on the model's ability to explain the structural changes in abortion numbers within a region, more effectively reflecting whether the model captures the main patterns of change in the time series. Taking the Marlborough and Canterbury region as an example, although its RMSE value under the Mixing-Channel PatchTST with Parallel LSTM model is higher than that of some low abortion regions, the model still achieved a high R^2 level, indicating that it can relatively stably track the seasonal fluctuations and peak changes in the region, demonstrating strong structural modelling capabilities. In regions like Northland and Auckland, the overall number of livestock abortions is low, and the seasonality is relatively weak. Although the RMSE is small, the R^2 is comparatively low, indicating that the model learns limited information about the time-series structure from the datasets for these regions. Therefore, the prediction results are more susceptible to random fluctuations.

Based on the above analysis, we can infer that the difference in model performance depends not only on the model structure itself, but also on the statistical characteristics of the regional data. In regions with pronounced seasonality and fluctuations, Mixing-Channel PatchTST with Parallel LSTM is more likely to leverage its advantages in multi-channel feature fusion and long-term dependency modelling; while in regions with weaker structures and limited signal strength, the performance gap between the two models is relatively narrowed.

6.3 Impact of Regional Heterogeneity on Model Performance

Further analysis of the model prediction results in Table 6.1 reveals significant differences in the predictive performance of the two models across different regions. The reason for this difference is not simply determined by the model structure itself, but is also closely related to the data characteristics of the region. Compared to aggregated sequence data at the national level, time series data in regional data exhibit greater heterogeneity in terms of sample size, seasonality, and completeness, making it more challenging for models to predict regional data.

First, the overall scale of livestock abortions varies significantly across different regions, which directly impacts the amount of effective time-series information that can be learned in the model. In regions with a higher number and larger fluctuations in livestock abortion cases, the time series typically exhibits a clearer fluctuation structure, providing the model with a relatively clear learning objective. For example, in the Waikato region, models are more likely to capture the relationship between long-term trends and seasonal cycles, and their predictions tend to show higher stability and consistency. Second, in regions with a generally low number of livestock abortions, the overall sample size is smaller, and some months may have zero values. Although data padding is applied to make the time series continuous, the seasonal signals of the time series remain weak, thus limiting the structural information available to the model and restricting further improvements in its predictive performance.

In addition, differences in climate conditions, livestock structure, and livestock breeding and ranch management methods between different regions can indirectly affect the temporal distribution characteristics of abortion numbers. For example, the number of abortions in some regions shows relatively regular seasonal fluctuations, while in other regions it shows irregular fluctuations or localised surges, such as when an epidemic of infectious disease breaks out in a certain region at a certain time of year, resulting in a large number of livestock abortions. This means that the model faces different types of prediction problems in different regions: in regions with clearer temporal structures, the model focuses more on characterising periodic and trend changes, while in regions with weaker temporal structures, the prediction task is closer to short-term estimation against a noisy background, making it more challenging to learn long-term stable trends.

Moreover, the reduction in the number of samples in a regional-scale dataset will further amplify the impact of noise on the model training process. With a limited sample size, model parameter estimation is more susceptible to interference from outliers or short-term perturbations, thus affecting the stability of prediction results. This factor partly explains why, in some regions, even with more complex model structures, performance improvements remain relatively limited.

In summary, the differences in model prediction performance at the regional level reflect objective differences in the structural strength and information density of regional time series data. These differences do not imply that the model fails in certain regions, but rather highlight the crucial role that data characteristics play in regional prediction tasks.

6.4 Summary of Regional-level Performance Differences

In regional-level model performance prediction experiments, both the LSTM Model with Attention Mechanism and the Mixing-Channel PatchTST with Parallel LSTM were able to demonstrate the temporal variation characteristics of livestock abortions in a region to some extent. However, the two models still showed systematic performance differences under different regional conditions.

As shown in Table 6.1, the Mixing-Channel PatchTST with Parallel LSTM exhibits relatively more stable predictive performance in most regions. Its advantage lies primarily in its ability to jointly model multivariate input information. Due to its structural advantages, it can comprehensively learn long-term trends and seasonal structures during the prediction process. Especially in regions with pronounced seasonal characteristics and relatively clear temporal structures, it often achieves better trend fitting and higher interpretability (such as the Waikato and Wellington regions). While the LSTM model with an attention mechanism also shows stable predictive performance in regions with pronounced seasonal characteristics, it is more prone to under-capturing peak changes or exhibiting lag in response in regions with large seasonal fluctuations, thus negatively impacting overall predictive performance. It is also worth mentioning that in some regions with weaker temporal structures and higher noise levels (such as the Northland and Auckland regions), the performance gap between the two models has narrowed significantly. This phenomenon indicates that in regional data prediction tasks, the structural strength and information density inherent in the regional data itself largely limit the potential for model performance improvement.

In summary, these two models each have their own characteristics under different regional conditions, and their performance differences may stem more from the structural strength and information density of the regional time series. These aspects warrant close attention in future regional data research.

6.5 Prediction Scope and Feature Availability

This study adopts a rolling one-step-ahead forecasting framework based on a 24-month historical data window. This implies that, at each prediction point, only the observed data from the preceding 24 consecutive months are utilised to forecast the number of abortions in the subsequent month. Within this forecasting framework, all input variables are available as observed information at the time of prediction. Climate variables are formulated using a 1-3 month lag structure, population size variables represent low-frequency structural indicators, and Google Trends search behaviour indicators correspond to contemporaneous statistics, all of which are accessible under a rolling forecast scheme. Therefore, under the current single-step forecasting configuration, the model does not encounter issues related to unavailable future information in terms of input accessibility and remains feasible for practical implementation.

It is important to further distinguish that, if the forecasting task is extended to multi-step ahead prediction (e.g., projecting abortion count trends over the subsequent six months or one year), certain exogenous variables would no longer be directly observable at future time points. For instance, climate observations and search behaviour indicators corresponding to future months are not available at the time of prediction. Under such circumstances, these exogenous variables would need to be independently forecasted or replaced with scenario-based inputs. In the absence of an exogenous variable forecasting module, the model implicitly depends on the assumption that future exogenous information will be accessible, an assumption that does not hold under long-term forecasting conditions.

Therefore, in its current formulation, the framework proposed in this study is more appropriately suited for short-term rolling forecasting and monthly risk monitoring applications. When medium- to long-term trend projection or scenario-based planning is required, the model architecture may be extended by incorporating climate forecasting models, behavioural variable scenario simulations, or statistical calibration components to strengthen its robustness and applicability in long-term decision-support contexts.

6.6 Generalisability

The multi-source time series forecasting framework proposed in this study has been systematically validated and examined using both national-level and regional-level datasets. The empirical findings indicate that the model achieved relatively stable predictive performance across different regional subsamples. This suggests that, under conditions in which the statistical definitions of the data remain consistent and the variable structure is preserved, the framework demonstrates a degree of cross-regional generalisability. Nevertheless, it should be acknowledged that the model's applicability remains contingent upon the specific characteristics of the data environment.

From a methodological standpoint, the dual-channel fusion framework developed in this study is not inherently tied to any specific region or livestock category. Its fundamental principle lies in independently modelling time-dependent characteristics and exogenous driving factors, followed by their integration through a fusion mechanism. Accordingly, under data conditions involving comparable multi-source time series inputs, this framework can, in principle, be extended to other agricultural disease surveillance or public health forecasting contexts.

At the same time, several underlying assumptions persist at the data level. First, the model assumes that the target variable exhibits some temporal dependence and seasonal structure. Second, climate variables and search behaviour indicators are expected to maintain a statistically learnable association with the target variable. If, in the intended application context, the data exhibit greater volatility, the observation window is shorter, or substantial discrepancies exist in statistical definitions, the model's predictive performance may be adversely affected. Therefore, these foundational conditions should be carefully assessed prior to transferring the framework to other countries or distinct production systems.

Furthermore, the comparative experiments conducted across the seven merged regions revealed pronounced differences in sample size, variability patterns, and extreme value distributions among regions, with model performance correspondingly varying as the regional data structure evolved. This observation further suggests that, when deploying the model at a broader scale or in cross-country contexts, it is necessary to account for the influence of regional differences in production structures, climatic conditions, and management practices on the model's generalisability.

Overall, the framework proposed in this study demonstrates a certain degree of methodological transferability; however, its predictive stability remains contingent upon specific data characteristics and variable availability conditions. In future work, further efforts may be directed toward enhancing the model's adaptability and robustness under complex data environments, for example through cross-regional joint training strategies, domain adaptation techniques, or scenario-based extension approaches.

7. Conclusion

This study centres on model research and heterogeneity analysis for predicting the number of livestock abortions in New Zealand. It integrates historical abortion records, livestock population size, lagged climate variables, and Google Trends search behaviour data. Two deep learning methods, Mixing-Channel PatchTST with Parallel LSTM and the LSTM model with an attention mechanism, are employed to model and predict livestock abortion counts.

Through comparative experiments with national and regional data, the performance differences of two models based on different principles in capturing long-term time dependence, seasonal fluctuations, and multi-source feature fusion were examined, and the performance of their prediction models was evaluated. In experiments with national-level data, both models showed relatively stable consistency across evaluation indicators such as RMSE, MAE, and R^2 and were able to effectively capture the long-term trend of abortion numbers, while maintaining reasonable predictive performance during seasonal peaks and troughs. In the experiment with the progressive introduction of features, the association between selected multi-source information and the number of livestock abortions was further verified: (1) Livestock population characteristics help to characterize the baseline differences between different years; (2) Lagged climate variables can represent the cumulative impact of environmental factors on the number of livestock abortions; (3) Google Trends search behavior features provide the model with forward-looking supplementary information to a certain extent, which enhances prediction performance. In the modelling analysis of regional data, it was shown that the model performance varied across different regions. In regions with distinct seasonal structures and relatively high abortion numbers, the Mixing-Channel PatchTST with Parallel LSTM typically exhibits more stable fitting capability. However, in regions with smaller sample sizes or weaker seasonal signals, the performance gap between the two models narrows significantly, and the prediction results are more constrained by the structural characteristics of the data itself.

By synthesising national- and regional-level experimental results, this study shows that fusion-based modelling strategies with a hybrid architecture exhibit strong adaptability and robustness in multi-source time-series forecasting tasks. Compared with single-stream temporal modelling approaches, such models demonstrate clear advantages in capturing complex seasonal fluctuations and cross-feature relationships, providing a practical pathway for quantitative modelling of livestock abortion counts.

Despite the encouraging results obtained in this study, several limitations remain. First, the current work focuses primarily on single-step (monthly) forecasting, and future studies may extend to multi-step prediction or longer-term risk assessment tasks. Second, regional-level experiments employ a uniform model structure and parameter configuration; future research could consider region-adaptive strategies to better accommodate differences in data scale and structural characteristics across regions. In addition, further expansion of input features remains possible, such as incorporating higher-resolution climate indicators or disease surveillance data to strengthen the modelling of abnormal fluctuations and rare events.

In summary, while the proposed framework demonstrates stable predictive capability under both national and regional settings, its practical deployment is most appropriate for short-term rolling forecasting scenarios where exogenous inputs are observable at the time of prediction. The generalisability of the framework is supported at the structural level; however, its predictive stability remains dependent on data characteristics and feature availability conditions. These findings highlight the importance of aligning modelling strategies with data environments when extending the framework to broader contexts.

Bibliography

[1] New Zealand scraps clean, green policies to boost economy. https://www.reuters.com/world/asia-pacific/new-zealand-scraps-clean-green-policies-boost-economy-2024-08-04/

[2] New Zealand Country Commercial Guide. https://www.trade.gov/country-commercial-guides/new-zealand-market-overview

[3] Butter madness: New Zealanders turn to churning as price of dairy staple soars. https://www.theguardian.com/world/2025/jun/06/butter-madness-new-zealanders-turn-to-churning-as-price-of-dairy-staple-soars

[4] Wijma, R., Weigel, D.J., Vukasinovic, N., Gonzalez-Peña, D., McGovern, S.P., Fessenden, B.C., McNeel, A.K. and Di Croce, F.A., 2022. Genomic prediction for abortion in lactating Holstein dairy cows. Animals, 12(16), p.2079.

[5] Villa-Cox, G., Van Loo, H., Speelman, S., Ribbens, S., Hooyberghs, J., Pardon, B., Opsomer, G. and Pascottini, O.B., 2024. Machine learning modeling to predict causes of infectious abortions and perinatal mortalities in cattle. Theriogenology, 226, pp.20-28.

[6] Rafati, N., Mehrabani-Yeganeh, H. and Hanson, T.E., 2010. Risk factors for abortion in dairy cows from commercial Holstein dairy herds in the Tehran region. Preventive Veterinary Medicine, 96(3-4), pp.170-178.

[7] Thurmond, M.C., Branscum, A.J., Johnson, W.O., Bedrick, E.J. and Hanson, T.E., 2005. Predicting the probability of abortion in dairy cows: a hierarchical Bayesian logistic-survival model using sequential pregnancy data. Preventive Veterinary Medicine, 68(2-4), pp.223-239.

[8] Mallam, I., Abubakar, I. and Yakubu, A., 2023. Application of multivariate logistic regression and decision trees to assess factors influencing prevalence of abortion and stillbirth in Yankasa sheep. Animal Science and Genetics, 19(4), pp.17-27.

[9] Dohoo, I.R., Martin, S.W. and Stryhn, H., 2009. Veterinary epidemiologic research (Vol. 2). Charlottetown: VER, Incorporated.

[10] Spiliotis, E., 2023. Time series forecasting with statistical, machine learning, and deep learning methods: Past, present, and future. In Forecasting with Artificial Intelligence: Theory and Applications (pp. 49-75). Cham: Springer Nature Switzerland.

[11] Makridakis, S., Spiliotis, E. and Assimakopoulos, V., 2018. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PloS one, 13(3), p.e0194889.

[12] Marques, T.C., Marques, L.R., Fernandes, P.B., de Lima, F.S., do Prado Paim, T. and Leão, K.M., 2024. Machine learning to predict pregnancy in dairy cows: an approach integrating automated activity monitoring and on-farm data. Animals, 14(11), p.1567.

[13] Villa-Cox, G., Van Loo, H., Speelman, S., Ribbens, S., Hooyberghs, J., Pardon, B., Opsomer, G. and Pascottini, O.B., 2024. Machine learning modeling to predict causes of infectious abortions and perinatal mortalities in cattle. Theriogenology, 226, pp.20-28.

[14] van Rumpt, M. and Rabaglino, M.B., 2024. The role of machine learning in decoding the molecular complexity of bovine pregnancy: a review. Reproduction, Fertility and Development, 37(1).

[15] Perez, M.M., Cabrera, E., Rial, C., You, Y., Wang, Y., Weinberger, K., Nydam, D.V. and Giordano, J.O., 2025. Screening and selection of a machine learning algorithm for development of a model to select cows for clinical examination using data from automated health monitoring technologies and other predictors of cow health. Journal of Dairy Science.

[16] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.

[17] Tito, M.H., Rahman, M.S., Hussaini, S.M.A.K., Nasrin, A., Asaduzzaman, M., Ashrafuzzaman, M., Khan, S., Chohan, M.S., Asif, A.H. and Arifuzzaman, M., 2024, December. Predicting Brucellosis outbreaks in dairy cattle using deep learning techniques. In IET Conference Proceedings CP913 (Vol. 2024, No. 37, pp. 325-329). Stevenage, UK: The Institution of Engineering and Technology.

[18] Arteaga-Troncoso, G., Luna-Alvarez, M., Hernández-Andrade, L., Jiménez-Estrada, J.M., Sánchez-Cordero, V., Botello, F., Montes de Oca-Jiménez, R., López-Hurtado, M. and Guerra-Infante, F.M., 2023. Modelling the unidentified abortion burden from four infectious pathogenic microorganisms (Leptospira interrogans, Brucella abortus, Brucella ovis, and Chlamydia abortus) in ewes based on artificial neural networks approach: The epidemiological basis for a control policy. Animals, 13(18), p.2955.

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems, 30.

[20] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H. and Zhang, W., 2021, May. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI conference on artificial intelligence (Vol. 35, No. 12, pp. 11106-11115).

[21] Lin, D., Kenéz, Á., McArt, J.A. and Li, J., 2023. Transformer neural network to predict and interpret pregnancy loss from activity data in Holstein dairy cows. Computers and Electronics in Agriculture, 205, p.107638.

[22] Zeng, D., Xu, J., Gu, J., Liu, L. and Xu, G., 2008, August. Short term traffic flow prediction using hybrid ARIMA and ANN models. In 2008 workshop on power electronics and intelligent transportation system (pp. 621-625). IEEE.

[23] Javed, A., Javaid, N., Hasnain, M., Sarfraz, U., Ahmed, I., Shafiq, M. and Choi, J.G., 2024. Applying advanced data analytics on pregnancy complications to predict miscarriage with explainable AI. IEEE Access.

[24] Bokani, A., Yadegaridehkordi, E. and Kanhere, S.S., 2025. LSTM-H: A Hybrid Deep Learning Model for Accurate Livestock Movement Prediction in UAV-Based Monitoring Systems. Drones, 9(5), p.346.

[25] Jebari, H., Rekiek, S. and Reklaoui, K., 2025. Advancing Precision Livestock Farming: Integrating Hybrid AI, IoT, Cloud and Edge Computing for Enhanced Welfare and Efficiency. International Journal of Advanced Computer Science & Applications, 16(7).

[26] Bello, R.W. and Moradeyo, O.M., 2023. Features-based individual cattle instance identification method using hybrid deep learning models for sustainable livestock management. World Scientific News, 180, pp.119-131.

[27] MPI data https://www.mpi.govt.nz/

[28] NZ climate data https://live-niwad9.pantheonsite.io/climate-and-weather/overview-new-zealands-climate

[29] Google Trends Internet Search Terms data https://trends.google.com/trends/

[30] Population data https://www.stats.govt.nz/topics/population/

[31] Mee, J.F., 2020. Investigation of bovine abortion and stillbirth/perinatal mortality-similar diagnostic challenges, different approaches. Irish Veterinary Journal, 73(1), p.20.

[32] Mee, J.F., 2023. Invited review: Bovine abortion—Incidence, risk factors and causes. Reproduction in Domestic Animals, 58, pp.23-33.

[33] Gaviglio, A., Corradini, A., Marescotti, M.E., Demartini, E. and Filippini, R., 2021. A theoretical framework to assess the impact of flooding on dairy cattle farms: Identification of direct damage from an animal welfare perspective. Animals, 11(6), p.1586.

[34] Tana, T., Wada, M., Benschop, J. and Vallee, E., 2025. The association between rainfall and human leptospirosis in Aotearoa New Zealand. Epidemiology & Infection, 153, p.e112.

[35] Graves, A., 2012. Long short-term memory. Supervised sequence labelling with recurrent neural networks, pp.37-45.

[36] Nie, Y., 2022. A Time Series is Worth 64Words: Long-term Forecasting with Transformers. arXiv preprint arXiv:2211.14730.

[37] Dosovitskiy, A., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[38] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q., 2020. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11534-11542).

[39] Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y., 2016. Deep learning (Vol. 1, No. 2). Cambridge: MIT press.

[40] Hewamalage, H., Bergmeir, C. and Bandara, K., 2021. Recurrent neural networks for time series forecasting: Current status and future directions. International Journal of Forecasting, 37(1), pp.388-427.

[41] Hyndman, R.J. and Athanasopoulos, G., 2018. Forecasting: principles and practice. OTexts.

[42] Huber, P.J., 1992. Robust estimation of a location parameter. In Breakthroughs in statistics: Methodology and distribution (pp. 492-518). New York, NY: Springer New York.

[43] Willmott, C.J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30(1), pp.79-82.

[44] Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscientific model development, 7(3), pp.1247-1250.

[45] Nagelkerke, N.J., 1991. A note on a general definition of the coefficient of determination. biometrika, 78(3), pp.691-692.

[46] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[47] He, K., Zhang, X., Ren, S. and Sun, J., 2016. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[48] Ba, J.L., Kiros, J.R. and Hinton, G.E., 2016. Layer normalization. arXiv preprint arXiv:1607.06450.

[49] Bishop, C.M., 1995. Training with noise is equivalent to Tikhonov regularization. Neural computation, 7(1), pp.108-116.

[50] Abbasi, M., Kargar, M., Ahmadian, F., NoormohammadZadehMaleki, D., Arandan, A. and Hosseini, N.S., 2024, October. GN-CNN-LSTM: Financial market prediction with Gaussian noise embedded CNN LSTM. In 2024 11th International Symposium on Telecommunications (IST) (pp. 287-294). IEEE.

[51] Hendrycks, D., 2016. Gaussian Error Linear Units (Gelus). arXiv preprint arXiv:1606.08415.

[52] Elman, J.L., 1990. Finding structure in time. Cognitive science, 14(2), pp.179-211.

[53] Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Advances in neural information processing systems, 28.

[54] Jain, S., Agrawal, S., Mohapatra, E. and Srinivasan, K., 2024. A novel ensemble ARIMA-LSTM approach for evaluating COVID-19 cases and future outbreak preparedness. Health Care Science, 3(6), pp.409-425.

[55] Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. Advances in neural information processing systems, 27.