### ****Project Title:**** Echocardiography Anomaly Detection Using ConvLSTM Beta Variational Autoencoder (VAE)

**Description:**  
This project focuses on detecting anomalies in echocardiography videos using a **Convolutional LSTM Variational Autoencoder (ConvLSTM VAE)**. The proposed ConvLSTM VAE architecture combines the strengths of convolutional layers for spatial feature extraction and LSTM layers for capturing temporal dynamics.

#### Key Components:

1. **Encoder:**  
   The encoder processes echocardiographic sequences, compressing them into a low-dimensional latent representation. It uses convolutional layers to capture spatial details (e.g., heart chamber boundaries and textures) and LSTM layers to model temporal variations (e.g., rhythmic heart motion). The output of the encoder represents the mean and variance of a latent Gaussian distribution.
2. **Sampling and Latent Space:**  
   The encoder employs the **reparameterization trick** to sample latent variables z from a learned Gaussian distribution. This approach ensures smooth latent space representations while enabling gradient-based optimization. The latent space is regularized to balance reconstruction quality and anomaly detection sensitivity (disentanglement) using a beta value .
3. **Decoder:**  
   The decoder reconstructs echocardiographic sequences from the latent variables z. By using LSTM layers for temporal modeling and deconvolutional layers for spatial reconstruction, the decoder aims to produce high-fidelity outputs resembling the original input.
4. **Loss Function:**  
   The training process minimizes a composite loss function:
   * **Reconstruction Loss**: Ensures accurate reconstruction of normal heart motion sequences.
   * **KL Divergence Loss**: Regularizes the latent space by aligning the learned distribution with a standard Gaussian prior, promoting generalization.

Results

### ****Time Series Prediction of Disease Outbreak Among Farmed Salmon in Norway****

#### **Project Title**

**"Time Series Prediction of Disease Outbreak Among Farmed Salmon in Norway, using SARIMAX and Exogenous Variables.**

#### **Introduction**

This project explores predictive modeling techniques for identifying disease outbreaks in farmed salmon within Norway, focusing on time series methods. It contrasts the effectiveness of traditional SARIMAX models with Temporal Convolutional Networks (TCNs), emphasizing the integration of exogenous variables and the complexities of capturing non-linear patterns in disease outbreak data.

Dataset

The primary datasets were retrieved from the Norwegian Food Safety Authority (Norwegian Food Safety Authority, 2024). The API allows selection of the 2 main datasets proposed in this study: Salmon diseases, Salmon lice. The datasets are recorded from 2012 to 2019 with approximately 75000 disease events and the salmon lice dataset has approximately 750000 rows. 2nd dataset (salmon Lice) tracks weekly records of salmon lice infestation levels and environmental conditions across various salmon farms in Norway. This project involved integrating multiple datasets in different formats to create a comprehensive framework for predicting disease outbreaks in farmed salmon. Additionally, geospatial data in formats like GEOjson The data sources ranged from outbreak records in CSV files to environmental metrics stored in structured databases and farm-specific metrics in Excel files. A key challenge was harmonizing these diverse formats and aligning them temporally, as the data were collected at varying frequencies (e.g., daily, weekly, and monthly). Temporal resampling and interpolation were employed to create a unified timeline, while geospatial data were transformed into tabular formats to integrate with time series models. The variability in scales and units across datasets required robust scaling and normalization techniques to ensure consistency. Missing data were addressed using imputation methods, while dimensionality reduction helped manage the high dimensionality introduced by numerous exogenous variables. The resulting integrated dataset enabled the inclusion of critical factors like temperature, lice counts, and farm density into the modelling process, using both SARIMAX and TCN models. This comprehensive data preparation facilitated a deeper exploration of temporal and spatial patterns, ultimately improving the predictive accuracy of disease outbreak models and providing a goodfoundation for model comparison and evaluation.

Objectives

* **Evaluate SARIMAX:** Analyze the capability of SARIMAX in integrating exogenous variables for disease prediction.
* **Explore Alternatives:** Test the hypothesis that TCNs provide a betterframework for this task, given their ability to handle non-linear dependencies and long-term patterns.
* **Benchmark Accuracy:** Compare model performance using metrics like RMSE and R-squared, with insights into each model's predictive strengths and limitations.

#### **Methodology**

1. **Data Analysis & Preprocessing**:
   * Utilized data from Norwegian salmon farms, focusing on disease outbreak counts, sea temperature, lice counts, and outbreak duration.
   * Performed descriptive statistics and correlation analysis to understand relationships among variables.
2. **SARIMAX Modeling**:
   * Constructed seasonal ARIMA models incorporating exogenous variables (SARIMAX).
   * Evaluated model residuals for autocorrelation, skewness, and kurtosis to identify areas of improvement.
3. **Temporal Convolutional Networks (TCN)**:
   * Implemented a TCN using Keras, with dilated convolutions to capture long-term dependencies.
   * Trained and validated the model with a sequence-based approach using TimeSeriesGenerator for input transformation.
4. **Performance Metrics**:
   * Benchmarked both models using metrics such as RMSE and R-squared.
   * Visualized predictions to evaluate alignment with actual data and detect under/overfitting.

The project reveals the limitations of SARIMAX for modeling complex, multi-dimensional time series data in the context of disease outbreaks among salmon. Temporal Convolutional Networks emerged as a more effective approach, demonstrating improved accuracy and robustness in handling non-linear relationships and long-term dependencies.

