A Mini Project Report On

**Predictive Anomaly Detection in Human Vital Signs Using LSTM for Health Risk Analysis**

**MASTER OF TECHNOLOGY**

In

**EMBEDDED SYSTEMS AND MACHINE LEARNING**



Submitted by

**Sai Prasad E (24ECM1R18)**

**Vamsi Krishna Reddy S (24ECM1R25)**

Under the guidance of

**Prof. Mohammad Farukh Hashmi**

Department of Electronics and Communication Engineering

NATIONAL INSTITUTE OF TECHNOLOGY, WARANGAL

TELANGANA – 506004

(2024-2026)

**ABSTRACT**

In this project, we present a predictive anomaly detection framework for analyzing human vital signs to identify potential health risks. Our method employs Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), well-suited for processing sequential data like time-series records of physiological parameters. The dataset includes comprehensive health metrics such as heart rate, respiratory rate, oxygen saturation, blood pressure, body temperature, and derived indices like BMI, HRV, and MAP, along with lifestyle factors such as sleep patterns, exercise frequency, and dietary habits.

The primary objective is to identify anomalies in vital signs and classify individuals into risk categories, enabling early detection of health complications. The data preprocessing stage involves cleaning, normalizing, and engineering features to capture meaningful patterns across multiple variables. The LSTM model is trained to analyze temporal dependencies in the data, distinguishing between normal and abnormal trends.

Our framework integrates statistical thresholds and domain knowledge to refine anomaly detection. The system’s predictive capabilities are enhanced by incorporating derived features, such as pulse pressure and sleep efficiency, which provide deeper insights into an individual’s health status. The results highlight the model’s accuracy in detecting subtle deviations that may precede critical health events.

This approach represents a significant step forward in leveraging advanced machine learning techniques for health monitoring. By combining LSTM networks with comprehensive health data, we provide a scalable and efficient tool for personalized health risk analysis. This system has the potential to improve clinical decision-making and promote proactive healthcare interventions.

.

**PROBLEM STATEMENT**

Monitoring vital signs is crucial for assessing an individual's health and detecting potential risks at an early stage. Vital signs such as heart rate, respiratory rate, oxygen saturation, blood pressure, and body temperature provide key indicators of a person’s physical state. However, with the growing prevalence of wearable devices and health trackers, the sheer volume of time-series data generated presents significant challenges in analysis and interpretation.

Healthcare professionals often rely on manual examination or rule-based systems to detect anomalies in vital signs, which can be time-consuming, subjective, and prone to error. Furthermore, these methods may fail to identify subtle patterns that could signal the onset of critical conditions, such as cardiovascular diseases or respiratory disorders.

This project aims to address these challenges by developing a predictive anomaly detection system that leverages Long Short-Term Memory (LSTM) networks to analyze vital signs data. The dataset includes a comprehensive set of features, including derived health metrics like BMI, MAP, and HRV, as well as lifestyle factors such as sleep patterns, exercise frequency, and dietary habits. By modeling the temporal dependencies in these data, the LSTM-based system seeks to identify deviations from normal patterns, classify health risks, and provide actionable insights.

The ultimate goal is to create a robust, scalable, and automated solution for health risk analysis. This system has the potential to assist healthcare providers in proactively addressing health issues, improving patient outcomes through early intervention, and reducing the burden on medical resources. By integrating advanced machine learning with physiological and lifestyle data, the project contributes to the broader field of personalized healthcare and preventive medicine

**Table of Contents**

**Abstract**…………………………………………………………………………………….……...i

**Problem statement**…...…………………………………………………………………….…….ii

**CHAPTER 1**

Introduction……………………………………………………………………..………………....1

**CHAPTER 2**

2.1 Background…………………………………………………………………………………....2

2.1.1 Sensor-based anomaly detection ..…………………………………..………………………3

2.1.2 Essential Metrics: The Role of Vitals in Wellness Monitoring……………………….....…...4

**CHAPTER 3**

Existing system………………………………………………………………………………..…..5

**CHAPTER 4**

Algoritm used………………………………………………………………………………...…....6

**CHAPTER 5**

Block diagram………………………………………………………………………………,,,…...9

**CHAPTER 6**

6.1 Data collection…………………………………………………………………………...…...13

6.2 Data pre-processing………………………………………………………………………......14

**CHAPTER 7**

7.1 Details of the dataset……………………………………………………………….……...…17

**CHAPTER 8**

Result and Analysis…………………………………………………………………………...….19

**CHAPTER 9**

Future scope……………………………………………………………………………………...25

**CHAPTER 10**

Conclusions……………………………………………………………………………………....27

**References**…………………………………………………………………………………..…...28

**CHAPTER 1**

**1.1Introduction**

Health monitoring is a critical aspect of preventive healthcare, enabling timely identification of potential health risks and improving overall patient outcomes. Vital signs, including heart rate, respiratory rate, oxygen saturation, blood pressure, and body temperature, serve as fundamental indicators of an individual’s physiological condition. These metrics, alongside derived indices like Body Mass Index (BMI), Mean Arterial Pressure (MAP), and Heart Rate Variability (HRV), provide valuable insights into a person’s health status. Coupled with lifestyle factors such as sleep patterns, exercise frequency, and dietary habits, this data forms a comprehensive picture of an individual’s well-being.The growing availability of wearable devices and health tracking tools has led to an exponential increase in the volume of vital signs data. However, analyzing this data for early detection of anomalies remains a significant challenge. Traditional methods rely on manual interpretation or basic rule-based systems, which are often time-intensive, subjective, and limited in detecting complex patterns indicative of health risks. Subtle anomalies in vital signs that could signal the early onset of conditions such as cardiovascular disease, respiratory disorders, or metabolic issues may go unnoticed, delaying critical interventions.To address these challenges, this project leverages advanced machine learning techniques, specifically Long Short-Term Memory (LSTM) networks, to develop a predictive anomaly detection framework for vital signs data. LSTM networks, with their ability to model temporal dependencies, are well-suited for analyzing time-series data and detecting deviations from normal patterns. By incorporating comprehensive physiological and lifestyle metrics, the system aims to identify health risks proactively and accurately.

This project seeks to bridge the gap between raw data and actionable insights, empowering healthcare providers with tools to enhance early diagnosis, reduce unnecessary interventions, and improve patient outcomes. By combining advanced deep learning methodologies with real-world health data, this work contributes to the broader goal of personalized healthcare and underscores the importance of timely, data-driven clinical decision-making.

**CHAPTER 2**

**2.1 Background**

Analyzing health risks with particular reference to monitoring vital signs is increasingly important in today's health care world. Vital signs, which include, among other things, heart rate, respiratory rate, oxygen saturation, blood pressure, and body temperature, are crucial markers of an individual's health status. When combined with other health information such as age, gender, weight, and lifestyle factors such as sleep pattern, exercise habits, and diet, these parameters affect the assessment of a person's health. The sheer body and complexity of data generated by modern wearables and health trackers pose great challenges in identifying subtle, possibly dangerous anomalies.Traditionally, these vital signs have had to be monitored and manually interpreted by clinical staff for assessing associated health risks. All these factors make it very time-consuming and risky in errors, particularly when large datasets must be derived over long periods. This is one of the several factors that have stimulated interest in recent times towards the automated analysis and detection of anomalies in vital signs data through advanced machine learning techniques. Long Short-Term Memory (LSTM) networks are among the chief among these techniques as they are powerful in handling sequential and time-series data. LSTM is well suited to modeling of temporal dependencies in health data, thus showing patterns that are abnormal that may be indicative of health problems, including cardiovascular diseases, respiratory diseases, or metabolic derangements.For this project, it will only be focusing on the possible application of LSTM networks for time-series anomaly detection because there are also many machine learning methods like CNNs and statistical methods explored in different health applications. This hybrid system proposes the construction of LSTM-based anomaly detection and correlating traditional health monitoring metrics with a multi-faceted dataset consisting of physiological measurements and lifestyle factors to build a health risk assessment tool. The present approach would exemplify early detection of health risks in otherwise healthy individuals through deep learning models as implemented in analyzing vital-signs data, thus bringing applicable insights to healthcare providers that can make for speedy intervention that translates to better patient outcomes. This tool could facilitate bridging the gap between data collection and clinical decision-making; making health management more proactive and personalized.

**2.1.1 Sensor-based anomaly detection**

Anomaly detection based on sensors remains the most important method of screening for possible health risk, since it can identify variations that occur as vital signs before there is clinical evidence of a symptom. The system monitors several variables on a continuous basis, including heart rate, respiratory activity, blood pressure, and oxygen saturation, understood through a network of sensors that regularly relays the above parameters. This real-time data forms the bedrock for anomaly detection, as it identifies deviations from the normal patterns in the healthy individuals with no symptoms.

It thus specifically aims to screen asymptomatic individuals for the early detection of irregularities at a time possible to institute the most effective preventive action or treatment. This means that with clear and continuous measurements of vital signs by sensors, healthcare providers could pick up abnormalities or trends that may suggest a significant underlying health problem even before there is any noticeable symptom.

Sensor monitoring for early detection is a vital determinant of enhanced treatment outcomes and reduced risk of diseases. In summary, sensor-based vital sign monitoring acts as a critical tool for proactive detection of health anomalies for the identification of potential issues before they manifest into signs and symptoms. Therefore, this emphasizes the need to monitor them regularly through sensors for early detection and management of health risks.

**2.1.2** **Essential Metrics: The Role of Vitals in Wellness Monitoring**

Human vital signs are critical indicators of an individual's overall health and wellness. These metrics offer valuable insights into the functioning of essential bodily systems and are often the first signs of potential health issues. The primary vital signs—**heart rate, respiratory rate, body temperature, blood pressure, and oxygen saturation**—are key physiological parameters used to monitor and assess a person’s health status.

* **Heart rate** (the number of heartbeats per minute) provides insights into the cardiovascular health and can indicate issues like arrhythmias, heart disease, or stress.
* **Respiratory rate** (the number of breaths taken per minute) reflects the efficiency of the respiratory system, helping to detect issues like respiratory infections, lung diseases, or breathing difficulties.
* **Body temperature** is a core indicator of overall health, with deviations signaling the presence of infections, inflammations, or other medical conditions.
* **Blood pressure** (both systolic and diastolic) measures the force of blood against the artery walls and is crucial for detecting hypertension, a risk factor for stroke, heart disease, and kidney problems.
* **Oxygen saturation** measures the percentage of oxygen in the blood, which is vital for cellular function. Low oxygen levels can indicate respiratory or cardiovascular issues.

The continuous monitoring of these vital signs is crucial for early detection of health anomalies, particularly in patients with chronic conditions, elderly individuals, or those recovering from surgeries. By tracking these vital parameters, healthcare providers can detect trends or deviations that may suggest emerging health risks. Moreover, advancements in wearable devices and smart technologies have made it easier to continuously track these metrics, providing real-time health data that can be used to detect potential issues before they become severe.

Incorporating these metrics into wellness monitoring systems allows for proactive health management, enabling timely intervention, improving patient outcomes, and preventing complications. As a result, human vitals are integral to the early detection and ongoing management of various medical conditions, making them essential for comprehensive health monitoring

**CHAPTER 3**

**Existing system**

Most vital signs health risk assessment systems are based on conventional monitoring, which usually includes manual interpretation by healthcare professionals. Such systems take time, are prone to human errors, and are inefficient when dealing with big data. In several cases, medical practitioners interpret the data individually, which is subjective and inconsistent, especially concerning time-series data produced by a wearable device or health tracker. Most anomaly detection techniques that can be used with vital signs data are mainly rule-based systems or simplistic statistical thresholds. Such methods prove limited in detection of subtle and complex patterns in data of large volumes. For instance, a traditional threshold-based system would consider a vital sign outside a certain range as anomalous; however, this does not relate to temporal relationships or patterns in data. This scenario ends up leaving room for false positives or sometimes misses anomalies that could indicate emerging health risks. Yet, most of the current systems also do not include effects such as lifestyle habits (like sleep, exercise, diet) or other derived health metrics such as BMI, HRV, and MAP, which are primarily important for a comprehensive health risk assessment. Now sets out to show that the need for solutions that could do much more sophisticated and automated processing of large volumes of time-series data, which could yield better quality anomaly detection, is clearly underscored. Machine learning approaches, particularly those that employ Long Short-Term Memory (LSTM) networks, have shown a lot of significant potential in such a case. Such networks excel in handling time-series data with long-term dependencies and temporal pattern detection that are mostly missed by conventional techniques.Such models necessitate high-quality datasets, labeled and well-curated training, to be effective in any clinical scenario. While the current systems have many benefits to offer, they do not embrace whole health metrics about evidence in making early detection and intervention more efficient. When combined with machine learning techniques such as LSTM networks into health risk analysis, it can give rise to much more precise, reliable, and efficient systems that can detect possible health issues sooner and, eventually, improve the results with proactive management.

**CHAPTER 4**

**Algorithm used:**

* **LSTM Autoencoder Networks:**

#### ****How It Is Used****:

* LSTMs are applied to model time-series health data, involving sequences like heart rate, respiratory rate, temperature, oxygen saturation, and blood pressure.
* The model is trained to recognize normal and anomalous health patterns, allowing it to detect potential health risks.

##### ****How It Works****:

1. **Data Representation**:
   * Input data consists of multiple time steps (e.g., 30 days) for each feature (e.g., vital signs).
2. **LSTM Network Dynamics**:
   * The LSTM network uses memory cells to store critical information over time, helping to retain important patterns while discarding unnecessary details.
   * These memory mechanisms (gates) help LSTMs overcome the challenges faced by traditional RNNs, making them suitable for learning long-term dependencies.
3. **Training Process**:
   * The LSTM model is trained to minimize prediction errors over multiple time steps.
   * The training process involves optimizing the model weights so that it can effectively learn from sequential health data.
4. **Anomaly Detection Mechanism**:
   * During inference, the LSTM compares new sequences to the learned normal patterns. If the difference exceeds a certain threshold, it flags the sequence as anomalous, indicating potential health risks.

#### ****CNN+LSTM Auto encoder Networks:****

##### ****How It Is Used****:

* The CNN+LSTM autoencoder is a hybrid model that combines convolutional layers for extracting features and LSTM units for temporal modeling.
* This model is used for reconstructing sequences and identifying anomalies in the health data.

##### ****How It Works****:

1. **Input Data and Structure**:
   * Input data includes time-series sequences of vital signs over a specific period (e.g., 30 days).
   * A masking layer is used to handle any missing values in the sequence.
2. **Feature Extraction with CNN Layers**:
   * Convolutional layers are used to capture local patterns in the data for each time step.
   * Max pooling layers reduce the dimensionality while preserving important features.
3. **Latent Representation**:
   * The flattened features are encoded into a lower-dimensional latent representation, which is repeated for each time step to prepare for the decoding process.
4. **LSTM Decoder for Sequence Reconstruction**:
   * The LSTM decoder processes the latent representation to reconstruct the input sequence, capturing temporal patterns.
5. **Training Objective**:
   * The goal is to minimize the reconstruction error between the original and reconstructed sequences.
   * The model learns to identify normal patterns and detects any deviations as anomalies.
6. **Anomaly Detection**:
   * During inference, any sequence with a high reconstruction error is flagged as anomalous.
   * This allows the model to differentiate between normal and risky health conditions, leading to early detection of potential issues.

In essence, the CNN-LSTM model combines the strengths of both architectures:

1. CNNs handle spatial correlations, such as patterns in the data at each time step.
2. LSTMs handle temporal relationships, such as trends or sequences across multiple time steps.

This makes CNN-LSTM particularly effective in applications like health monitoring, anomaly detection, and predictive modeling, where both spatial and temporal dependencies play a critical role. For instance, in health data, CNN layers can identify abnormalities in a heart rate pattern at a given time, while LSTM layers assess how these patterns evolve over days or weeks to flag potential risks.

* **Enhanced Feature Extraction**:

CNN layers in CNN-LSTM focus on extracting spatial features from input data before passing it to LSTM layers. This improves the model’s ability to identify intricate patterns in high-dimensional data, which a normal LSTM might miss.

* **Reduced Input Dimensionality**:

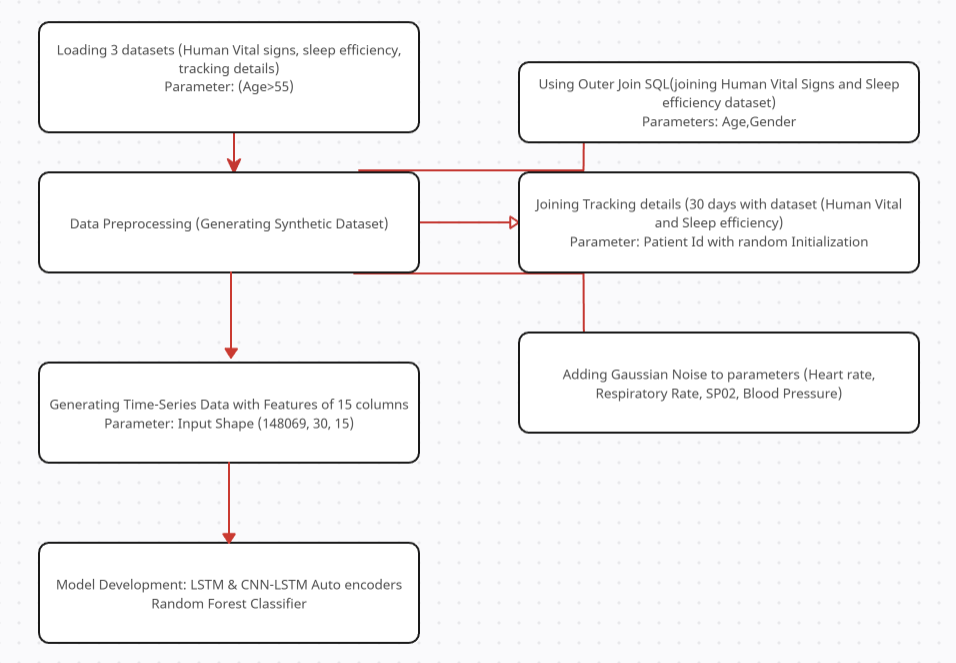
By using convolution and pooling operations, CNN layers reduce the dimensionality of the input, making it computationally efficient for LSTM layers to process sequential dependencies.

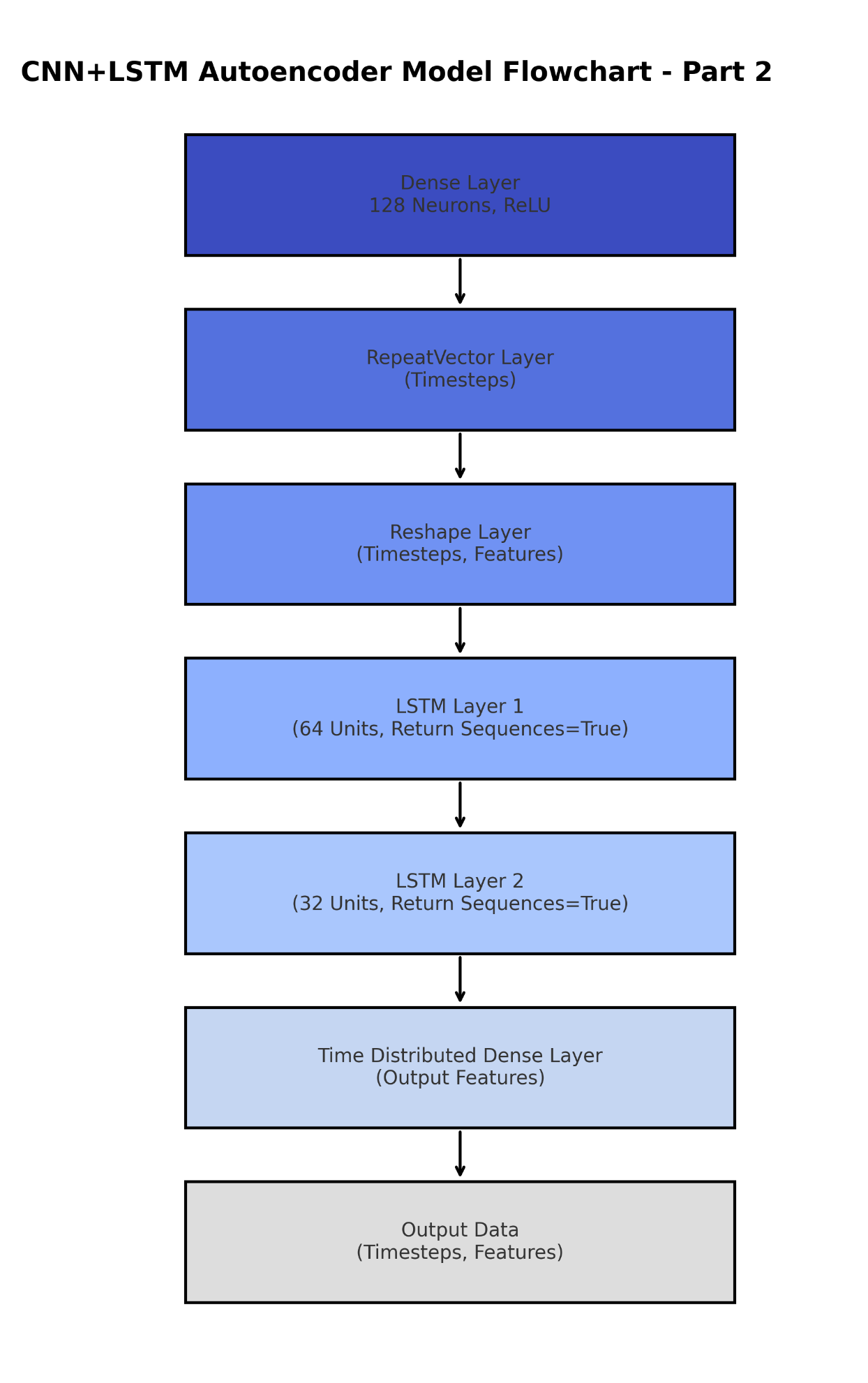
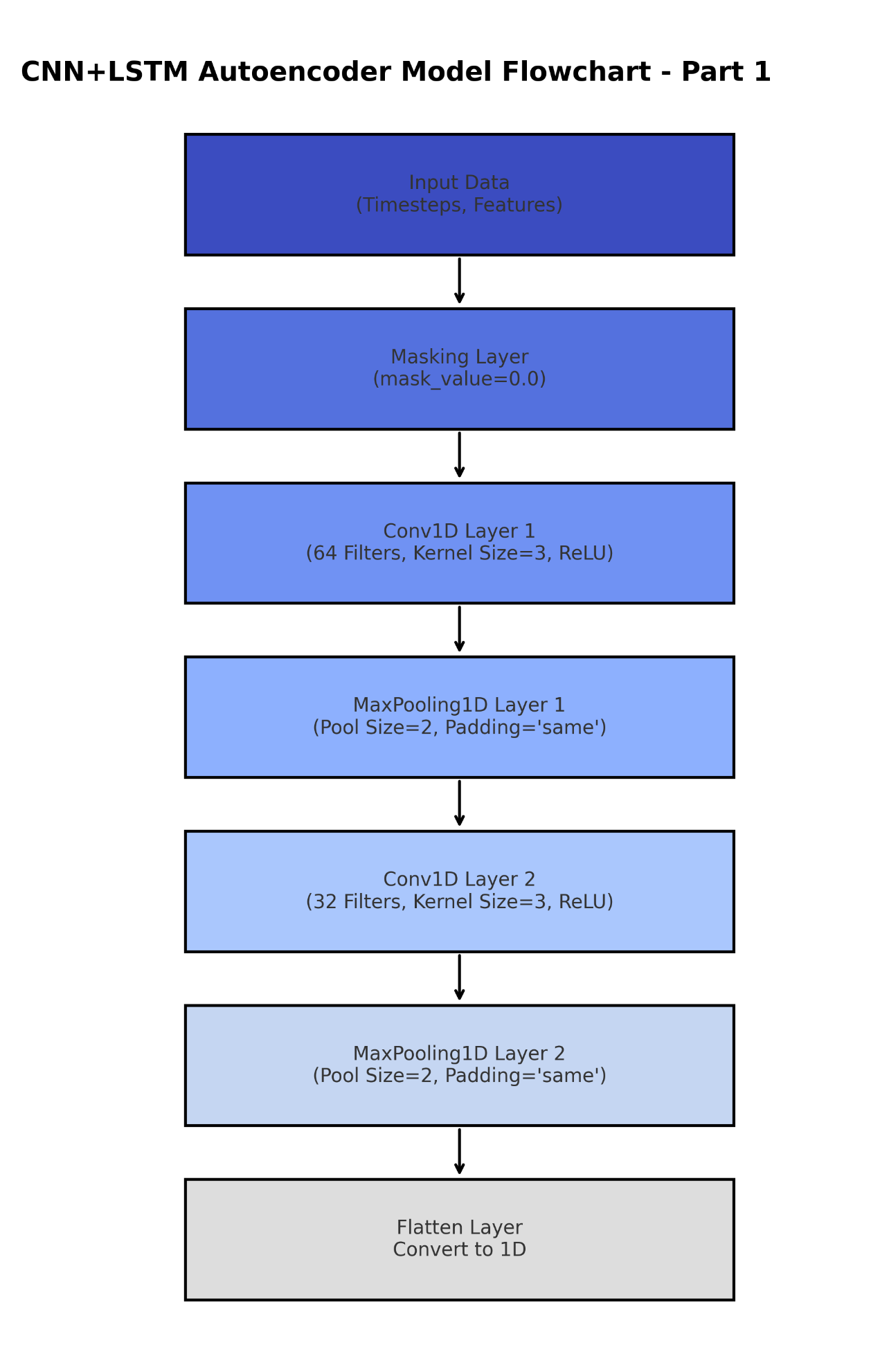
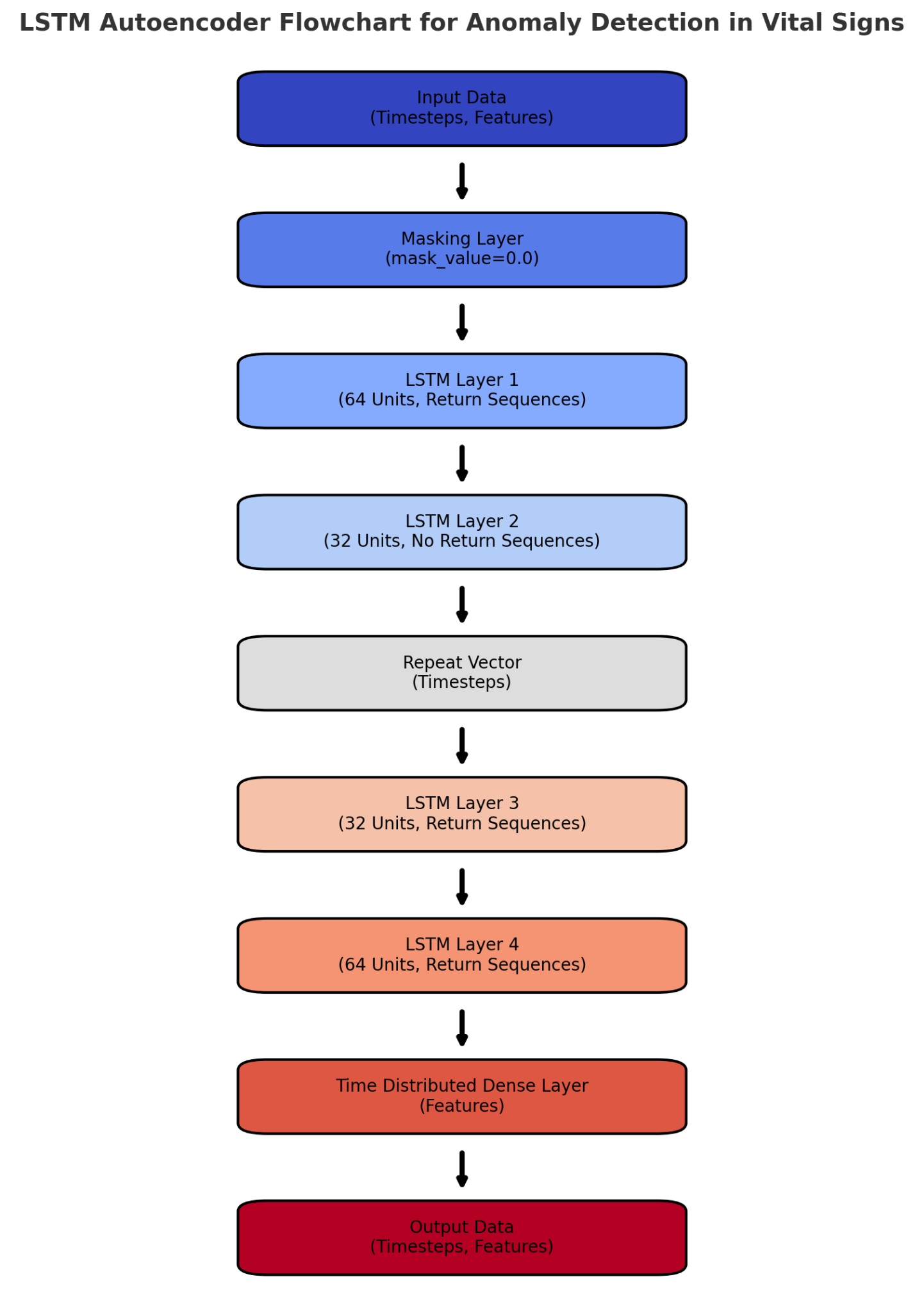
* **Improved Performance on Structured Data**:

CNN-LSTM excels in scenarios where input data has spatial correlations (e.g., time-series with multi-dimensional signals or images). Normal LSTM lacks the capability to leverage such spatial relationships effectively.

**CHAPTER 5**

**Block diagram:**

****

****

**CHAPTER 6**

**6.1 Data collection**

Data collection is the first and most important step in developing an effective machine learning model. In wellness monitoring, the data gathered plays a critical role in determining the model’s accuracy and performance. In this project, data is collected through various sensors, providing real-time measurements of vital signs and lifestyle metrics that are essential for accurate health predictions.

Sensors capture physiological data such as **heart rate**, **respiratory rate**, **body temperature**, **oxygen saturation**, and **blood pressure** (both **systolic** and **diastolic**). These measurements are crucial for assessing an individual’s cardiovascular and respiratory health. In addition, data on **age**, **gender**, **weight**, and **height** are collected to help contextualize the individual’s health profile.Behavioral data, including **sleep duration**, **sleep efficiency**, **REM sleep percentage**, and **exercise frequency**, are tracked through wearables and smart devices. **Lifestyle factors** such as **caffeine consumption**, **alcohol intake**, and **smoking status** are also captured, offering insights into habits that influence overall health.Derived metrics such as **Heart Rate Variability (HRV)**, **Pulse Pressure**, **Body Mass Index (BMI)**, and **Mean Arterial Pressure (MAP)** are calculated from raw sensor data, providing further insights into the individual’s health. These additional metrics help build a more comprehensive understanding of health risks and can aid in detecting abnormalities.

This collected data is essential for training a machine learning model, allowing it to predict health outcomes and identify potential health issues more accurately. The quality and relevance of the data directly affect the model’s performance, making precise and consistent data collection vital for developing a reliable wellness monitoring system.

**6.2 Data Pre Processing**

* **Load Datasets**

Collect the following datasets:

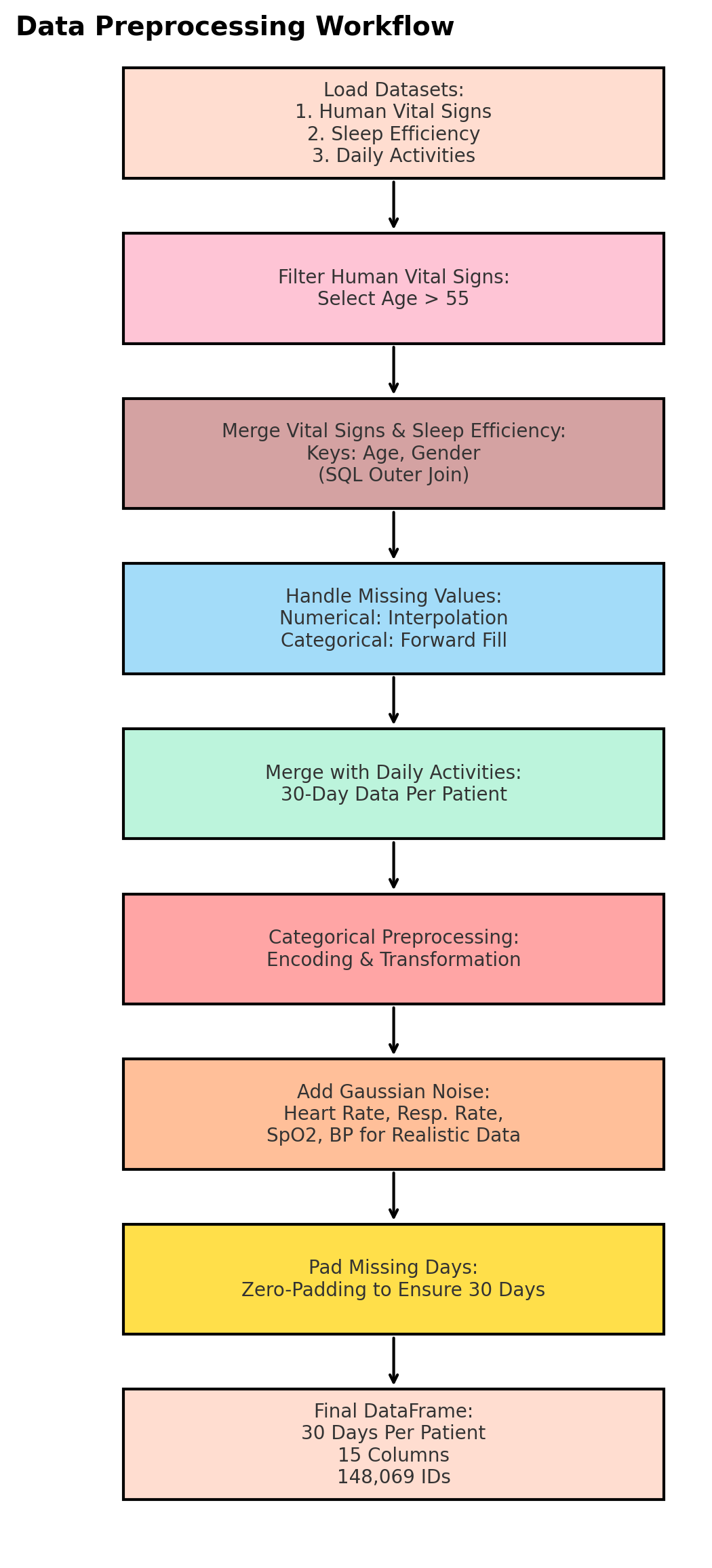
* + **Human Vital Signs:** Includes parameters like heart rate, respiratory rate, SpO2, blood pressure, and more.
  + **Sleep Efficiency:** Provides details about sleep quality, efficiency percentages, and related metrics.
  + **Daily Activities:** Contains 30-day activity patterns, such as exercise levels, step count, and calorie consumption.
* **Filter Data**
* From the **Human Vital Signs** dataset, select records for patients aged **above 55 years**.
* Extract relevant Patient IDs, narrowing down the dataset for focused processing.
* **Merge Vital Signs and Sleep Efficiency**
* Combine the filtered **Vital Signs** dataset with the **Sleep Efficiency** dataset.
* Use **age** and **gender** as keys for merging.
* Perform an **SQL Outer Join** to retain all records, even if data is missing in one of the datasets.
* **Handle Missing Values**
* **Numerical Columns:** Use **interpolation** to estimate missing values based on existing trends.
* **Categorical Columns:** Apply **forward fill (ffill)** to propagate the last valid observation for filling gaps.
* **Merge with Daily Activities**
* Merge the combined **Vital Signs + Sleep Efficiency** dataset with the **Daily Activities** dataset.
* Ensure every patient is matched with their corresponding activity records for the last **30 days**.
* **Categorical Preprocessing**

Convert categorical variables into machine-readable formats:

* + Apply **label encoding** for binary categories.
  + Use **one-hot encoding** for multi-class variables.
* **Add Gaussian Noise**

Introduce **Gaussian noise** to simulate realistic variations in vital signs:

* + Apply noise to **heart rate**, **respiratory rate**, **SpO2**, and **blood pressure**.
  + Use parameters:
* **Mean:** 0
* **Standard Deviation:** Based on expected variability (e.g., ±5 bpm for heart rate).
  + Clip the noisy values to realistic ranges (e.g., **SpO2:** 0-100, **heart rate:** 40-200 bpm).
* **Pad Missing Days**
* Use **zero-padding** to ensure all patients have a uniform 30 days of activity data.
* Fill gaps with zeros for patients missing some days of activity records.
* **Finalize the DataFrame**
* Structure the final dataframe to have:
  + **Rows:** 30 days of data for each patient.
  + **Columns:** 15 features, including vital signs, sleep efficiency, and daily activities.
* Save the processed dataset in a suitable format, such as .csv or .parquet.

****

**CHAPTER 7**

**7.1 Details of the dataset:**

### 1. ****Human Vital Sign Dataset****

* **Description**: This dataset comprises various human vital signs, including metrics such as heart rate, blood pressure, respiratory rate, and body temperature. It's designed to assist in analyzing and understanding human physiological conditions.
* **Source**: [Human Vital Sign Dataset](https://www.kaggle.com/datasets/nasirayub2/human-vital-sign-dataset)
* **Key Features**:
  + **Heart Rate**: Beats per minute.
  + **Blood Pressure**: Systolic and diastolic measurements.
  + **Respiratory Rate**: Breaths per minute.
  + **Body Temperature**: Measured in degrees Celsius or Fahrenheit.
* **Usage**: Ideal for studies related to health monitoring, medical research, and developing predictive models for health conditions.

### 2. ****Sleep Efficiency Dataset****

* **Description**: This dataset contains information on individuals' sleep patterns and efficiency. It includes data on sleep duration, quality, and factors that may influence sleep, such as caffeine intake and exercise.
* **Source**: [Sleep Efficiency Dataset](https://www.kaggle.com/datasets/equilibriumm/sleep-efficiency)
* **Key Features**:
  + **Sleep Duration**: Total hours slept per night.
  + **Sleep Efficiency**: Percentage of time spent asleep while in bed.
  + **Caffeine Intake**: Amount of caffeine consumed before sleep.
  + **Exercise Frequency**: Number of exercise sessions per week.
* **Usage**: Useful for analyzing factors affecting sleep quality and for developing models to predict sleep efficiency.

### 3. ****Bellabeat Tracking Details Dataset****

* **Description**: This dataset provides tracking details from Bellabeat devices, focusing on physical activity, sleep, and other health metrics over a period. It includes daily steps, calories burned, and active minutes.
* **Source**: [Bellabeat Dataset](https://www.kaggle.com/datasets/elijahtoumoua/bellabeat)
* **Key Features**:
  + **Daily Steps**: Number of steps taken each day.
  + **Calories Burned**: Total calories expended per day.
  + **Active Minutes**: Duration of physical activity per day.
  + **Sleep Duration**: Hours of sleep recorded per night.
* **Usage**: Beneficial for analyzing daily activity patterns, sleep behavior, and their correlations with overall health.

**CHAPTER 8**

**Results & Analysis:**

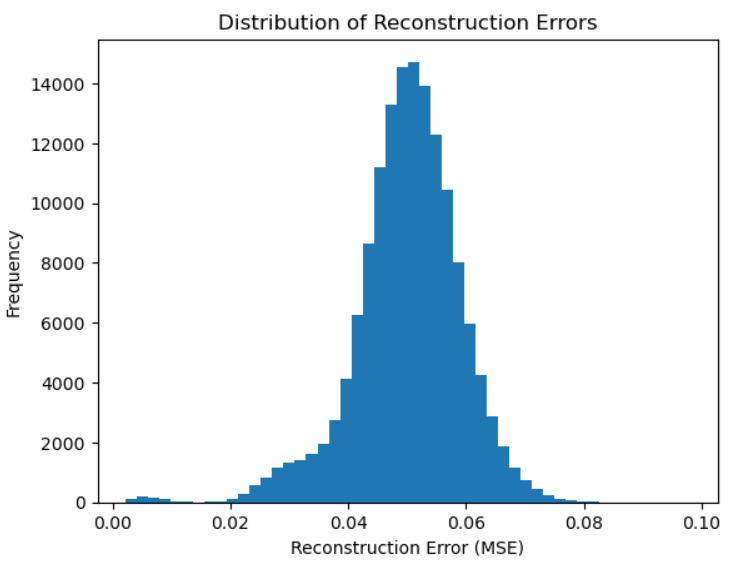
****

Fig:LSTM Model Reconstruction Errors

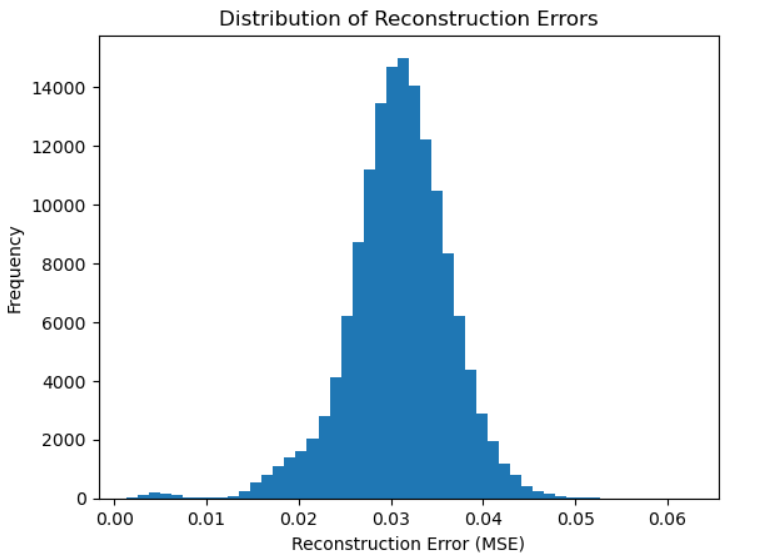
****

Fig: CNN-LSTM Model Reconstruction Errors

### ****Key Observations****:

1. **LSTM Model Reconstruction Errors (First Graph)**:
   * Distribution has a slightly wider spread.
   * Mean reconstruction error appears higher compared to the CNN-LSTM model.
   * Indicates that the LSTM model struggles more with reconstructing the data, possibly due to insufficient feature extraction or overfitting.
2. **CNN-LSTM Model Reconstruction Errors (Second Graph)**:
   * Distribution is narrower and more centered around a smaller mean reconstruction error.
   * Indicates better reconstruction capabilities, likely due to CNN layers extracting spatial features efficiently before the LSTM processes temporal dependencies.

### ****Analysis****:

* The **CNN-LSTM model** has a better performance in terms of reconstruction error.
* The **LSTM model** alone may struggle due to:
  + Lack of spatial feature extraction.
  + Potential overfitting to temporal patterns without learning spatial relationships.
* The CNN-LSTM model should be preferred for anomaly detection since it has lower reconstruction errors, indicating a better understanding of normal data patterns.
* Use these distributions to set an **anomaly detection threshold**:
* Use the tail of the error distribution (e.g., 95th percentile) as the cutoff for anomalies.

**Results & Analysis:**

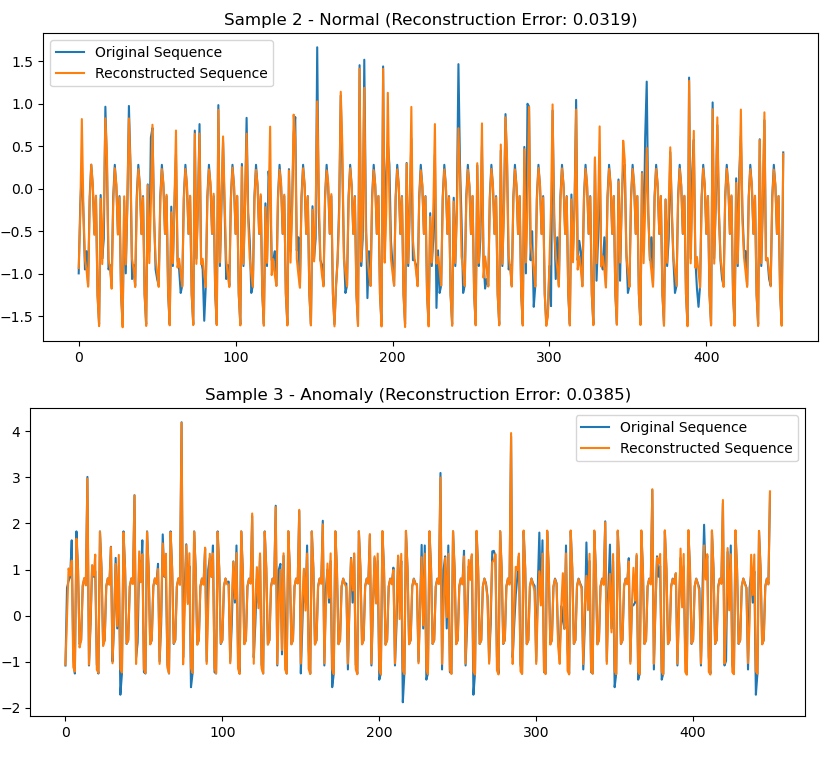


Fig: Anomaly & Normal reconstructed Error for CNN+LSTM model

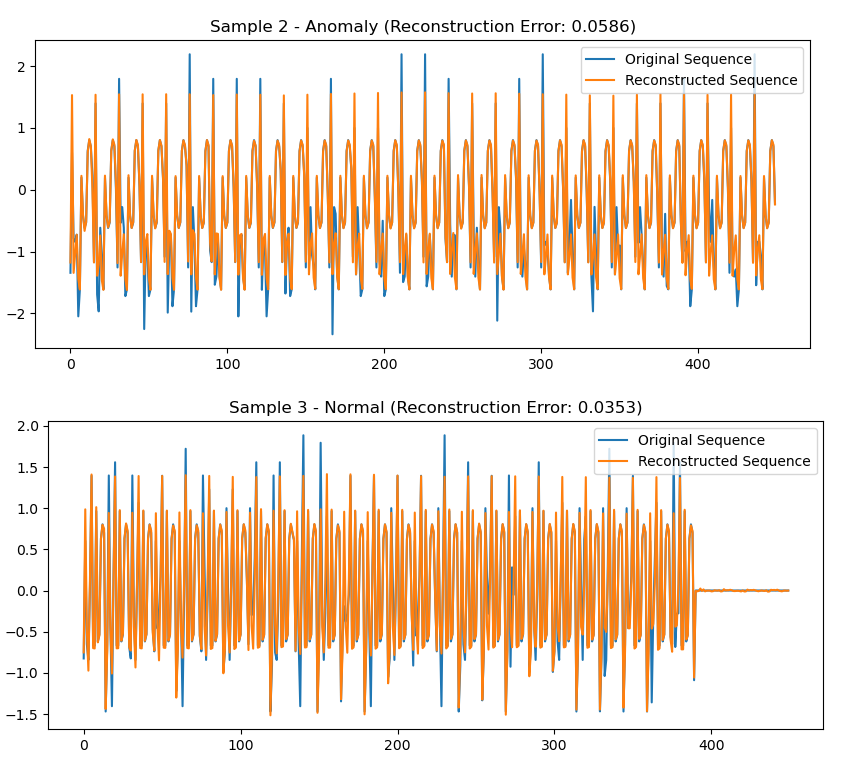
****

Fig: Anomaly & Normal Reconstruction Errors

**CNN-LSTM Model Observations (First Image):**

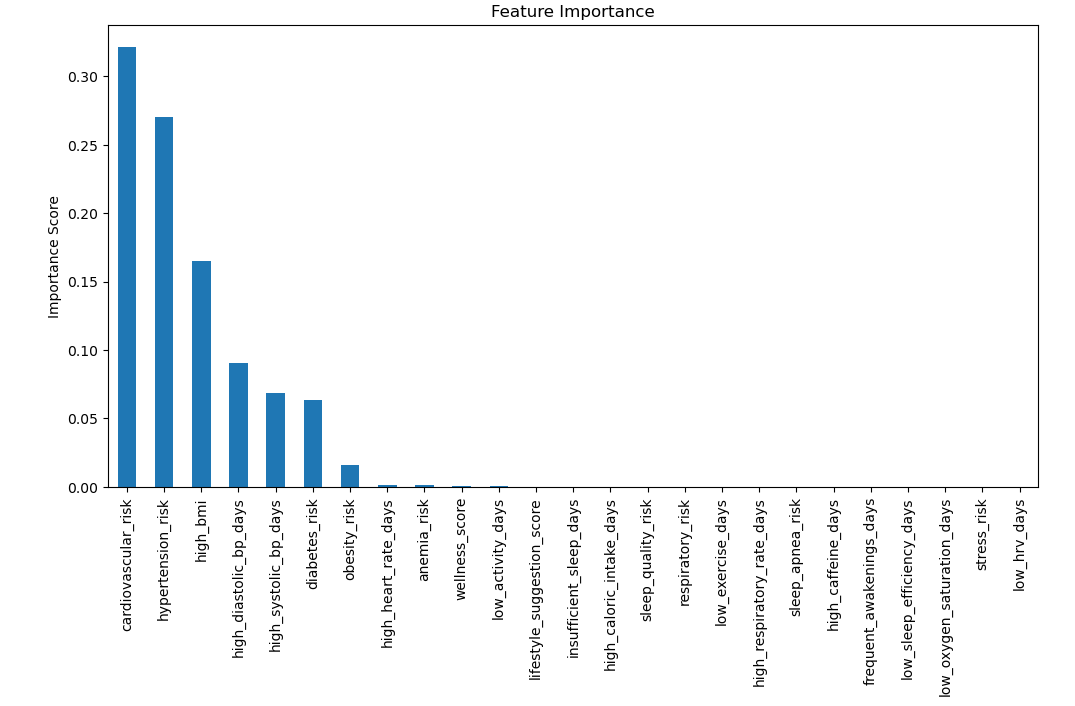
1. **Normal Data (Sample 2)**:
   * Reconstruction error: **0.0319** (low).
   * The reconstructed sequence aligns well with the original, indicating good generalization.
2. **Anomalous Data (Sample 3)**:
   * Reconstruction error: **0.0385** (slightly higher than normal).
   * The model struggles to reconstruct the anomalies, which results in higher reconstruction errors. This difference can be used for anomaly detection

**LSTM Model Observations (Second Image):**

1. **Anomalous Data (Sample 2)**:
   * Reconstruction error: **0.0586** (significantly higher).
   * The model struggles to reconstruct anomalies compared to the CNN-LSTM model, indicating less generalization capability.
2. **Normal Data (Sample 3)**:
   * Reconstruction error: **0.0353** (slightly higher than the CNN-LSTM model).
   * Reconstructed sequence aligns reasonably well but shows more deviation compared to the CNN-LSTM model.

**Key Takeaways:**

1. **Performance**:
   * The **CNN-LSTM model** consistently outperforms the LSTM model in reconstructing both normal and anomalous data.
   * The **LSTM model** has a higher reconstruction error, especially for anomalies.
2. **Anomaly Detection**:
   * The reconstruction error gap between normal and anomalous data is more distinct in the CNN-LSTM model.
   * This makes the CNN-LSTM model more reliable for detecting anomalies based on reconstruction errors.
3. **Generalization**:
   * The CNN-LSTM model benefits from spatial feature extraction (via CNN layers) before processing temporal dependencies, improving reconstruction accuracy.



The feature importance analysis highlights that **cardiovascular risk**, **hypertension risk**, and **high BMI** are the most critical predictors for the model, reflecting their strong influence on health outcomes. Blood pressure metrics, such as **high diastolic and systolic BP days**, and **diabetes risk**, also play significant roles, emphasizing the importance of managing these factors. Moderately important features include **high heart rate days** and **insufficient sleep days**, indicating the relevance of sleep and heart rate variability in predicting health risks. Less impactful features like **stress risk** and **low sleep efficiency** suggest a potential for further refinement or exclusion to streamline the model. This analysis underscores the need to prioritize cardiovascular and lifestyle-related factors for better health predictions.

**CHAPTER 9**

**Future scope**

The integration of sensor-based data collection with machine learning for wellness monitoring holds immense promise for future advancements in healthcare. The data collected, such as heart rate, oxygen saturation, blood pressure, sleep patterns, and lifestyle factors, can be leveraged to further enhance predictive health models and contribute to better health management and personalized care.

**Enhanced Accuracy in Health Predictions:** By continuously collecting diverse health metrics through sensors, the system can achieve more accurate and real-time health assessments. The use of advanced machine learning algorithms will help in reducing errors, such as false positives and false negatives, providing more reliable insights into an individual's health status. The ability to assess multiple variables simultaneously allows for better predictions and early detection of potential health issues.

**Early Detection and Prevention:** The integration of sensor data with machine learning will significantly improve early detection capabilities. By continuously monitoring an individual's vital signs and lifestyle factors, the system can identify early signs of health problems such as cardiovascular diseases, respiratory issues, or metabolic disorders. Early intervention will lead to improved health outcomes and prevention of serious conditions.

**Personalized Health Recommendations:** One of the major advantages of using this system is the potential for personalized health monitoring. By analyzing individual data, the system can offer customized health recommendations, such as changes in exercise routines, dietary suggestions, or sleep improvements, tailored to each person's specific needs. This personalization can help optimize health outcomes and encourage healthier lifestyle choices.

**Integration with Wearable Devices and Health Platforms:** The system can be integrated with a variety of wearable devices (such as smartwatches, fitness trackers, and health monitors) and health platforms to collect more comprehensive data. This integration could provide continuous monitoring and enable the system to offer real-time health feedback and alerts to users or healthcare providers. Enhanced sensor technology can further improve data accuracy, enabling even more precise health assessments.

**Big Data and AI Advancements:** With the growing availability of big data in health and wellness, the system can benefit from being trained on larger datasets, improving its predictive power. The increasing advancements in artificial intelligence, such as deep learning and reinforcement learning, will allow for the refinement of models, leading to better health insights and automated decision-making.

**Telemedicine and Remote Healthcare Monitoring:** Sensor-based health monitoring systems can enhance telemedicine by enabling healthcare providers to remotely monitor patients in real time. This is especially beneficial for patients in rural or underserved areas who may not have easy access to medical facilities. With the integration of AI models, remote diagnostics can become more accurate, making healthcare more accessible and efficient.

**Research and Clinical Trials:** The data generated by this system can be valuable for clinical trials and medical research, contributing to the development of new treatment strategies and healthcare models. By analyzing large datasets from diverse populations, the system can assist in identifying new patterns or trends in health that may have been previously overlooked, paving the way for advancements in healthcare.

In conclusion, the future scope of this wellness monitoring system powered by sensor data and machine learning is vast. With the potential to improve early detection, provide personalized healthcare recommendations, integrate with wearable devices, and contribute to medical research, this technology holds the promise to revolutionize healthcare, offering more efficient, accurate, and accessible solutions for individuals and healthcare providers alike.

**CHAPTER 10**

**Conclusions:**

The analysis of the distribution of reconstruction errors and reconstruction performance for the LSTM and CNN-LSTM models reveals significant differences in their effectiveness for anomaly detection. The distribution of reconstruction errors for the LSTM model is wider and has a slightly higher mean, indicating that it struggles to reconstruct input data accurately. This higher variability makes it difficult to distinguish normal and anomalous samples, reducing its reliability for anomaly detection. In contrast, the CNN-LSTM model shows a narrower distribution of reconstruction errors with a lower mean, suggesting that it effectively captures both spatial and temporal dependencies, resulting in better reconstruction accuracy.

In terms of reconstruction performance, the CNN-LSTM model outperforms the LSTM model in both normal and anomalous scenarios. For normal samples, the CNN-LSTM model reconstructs sequences with minimal error, demonstrating superior generalization. For anomalous samples, the reconstruction errors are more distinct in the CNN-LSTM model, making it easier to define a threshold for anomaly detection. On the other hand, the LSTM model exhibits higher reconstruction errors for both normal and anomalous samples, with greater error variability, further highlighting its limitations.

Overall, the CNN-LSTM model is the preferred choice for anomaly detection due to its superior reconstruction accuracy and clearer distinction between normal and anomalous samples. It benefits from the CNN layers' ability to extract spatial features, which enhances the LSTM layers' temporal pattern learning capabilities. The LSTM model, while simpler, is less effective in handling complex sequences, particularly those with spatial dependencies. Therefore, for applications like detecting heart issues or abnormal physiological signals, the CNN-LSTM model is recommended, with anomaly detection thresholds set based on its error distribution (e.g., 95th percentile for normal samples).

**References**

1. W. H. Organization et al., “Trends in maternal mortality 2000 to 2017: estimates by who, unicef, unfpa, world bank group and the united nations population division,” 2019.
2. L. Clifton, D. A. Clifton, P. J. Watkinson, and L. Tarassenko, “Identification of patient deterioration in vital-sign data using one-class support vector machines,” in 2011 federated conference on computer science and information systems (FedCSIS). IEEE, 2011, pp. 125–131.
3. A. R. M. Forkan and I. Khalil, “Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring,” Pervasive and Mobile Computing, vol. 38, pp. 296–311, 2017.
4. D. Rav`ı, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, “Deep learning for health informatics,” IEEE journal of biomedical and health informatics, vol. 21, no. 1, pp. 4–21, 2016.
5. T. Mitsunaga, I. Hasegawa, M. Uzura, K. Okuno, K. Otani, Y. Ohtaki, A. Sekine, and S. Takeda, “Comparison of the national early warning score (news) and the modified early warning score (mews) for predicting admission and in-hospital mortality in elderly patients in the pre-hospital setting and in the emergency department,” PeerJ, vol. 7, p. e6947, 2019.
6. F. E. Shamout, T. Zhu, P. Sharma, P. J. Watkinson, and D. A. Clifton, “Deep interpretable early warning system for the detection of clinical deterioration,” IEEE journal of biomedical and health informatics, vol. 24, no. 2, pp. 437–446, 2019.
7. Hung, Y.-X.; Chiang, C.-Y.; Hsu, S.J.; Chan, C.-T. Abnormality Detection for Improving Elder’s Daily Life Independent. In Aging Friendly Technology for Health and Independence; Springer Nature: Cham, Switzerland, 2010; Volume 6159, pp. 186–194.
8. Zerkouk, M.; Chikhaoui, B. Long Short Term Memory Based Model for Abnormal Behavior Prediction in Elderly Persons. In Proceedings of the International Conference on Smart Homes and Health Telematics, New York, NY, USA, 14–16 October 2019; pp. 36–45.