



## OPEN ACCESS

## EDITED BY

Hyo Kyung Lee,  
Korea University, Republic of Korea

## REVIEWED BY

Ricardo Valentim,  
Federal University of Rio Grande do  
Norte, Brazil  
Xu Huang,  
Nanjing University of Science and  
Technology, China

## \*CORRESPONDENCE

Teresa Wu,  
✉ Teresa.Wu@asu.edu

RECEIVED 16 February 2024

ACCEPTED 26 September 2024

PUBLISHED 15 October 2024

## CITATION

Patharkar A, Cai F, Al-Hindawi F and Wu T  
(2024) Predictive modeling of biomedical  
temporal data in healthcare applications:  
review and future directions.  
*Front. Physiol.* 15:1386760.  
doi: 10.3389/fphys.2024.1386760

## COPYRIGHT

© 2024 Patharkar, Cai, Al-Hindawi and Wu.  
This is an open-access article distributed  
under the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited,  
in accordance with accepted academic  
practice. No use, distribution or reproduction  
is permitted which does not comply with  
these terms.

# Predictive modeling of biomedical temporal data in healthcare applications: review and future directions

Abhidnya Patharkar<sup>1,2</sup>, Fulin Cai<sup>1,2</sup>, Firas Al-Hindawi<sup>1,2</sup> and  
Teresa Wu<sup>1,2\*</sup>

<sup>1</sup>School of Computing and Augmented Intelligence, Arizona State University, Tempe, AZ, United States, <sup>2</sup>ASU-Mayo Center for Innovative Imaging, Arizona State University, Tempe, AZ, United States

Predictive modeling of clinical time series data is challenging due to various factors. One such difficulty is the existence of missing values, which leads to irregular data. Another challenge is capturing correlations across multiple dimensions in order to achieve accurate predictions. Additionally, it is essential to take into account the temporal structure, which includes both short-term and long-term recurrent patterns, to gain a comprehensive understanding of disease progression and to make accurate predictions for personalized healthcare. In critical situations, models that can make multi-step ahead predictions are essential for early detection. This review emphasizes the need for forecasting models that can effectively address the aforementioned challenges. The selection of models must also take into account the data-related constraints during the modeling process. Time series models can be divided into statistical, machine learning, and deep learning models. This review concentrates on the main models within these categories, discussing their capability to tackle the mentioned challenges. Furthermore, this paper provides a brief overview of a technique aimed at mitigating the limitations of a specific model to enhance its suitability for clinical prediction. It also explores ensemble forecasting methods designed to merge the strengths of various models while reducing their respective weaknesses, and finally discusses hierarchical models. Apart from the technical details provided in this document, there are certain aspects in predictive modeling research that have arisen as possible obstacles in implementing models using biomedical data. These obstacles are discussed leading to the future prospects of model building with artificial intelligence in healthcare domain.

## KEYWORDS

biomedical temporal data, biomedical data challenges, forecasting, clinical predictive modeling, temporal data modeling problems, statistical time-series models, temporal machine learning models, deep temporal models

# 1 Introduction

## 1.1 Biomedical time series data

Clinical or biomedical data advances medical research by providing insights into patient health, disease progression, and treatment efficacy. It underpins new diagnostics, therapies, and personalized medicine, improving outcomes and understanding complex conditions. In predictive modeling, biomedical data is categorized as spatial, temporal, and spatio-temporal (Khalique et al., 2020; Veneri et al., 2012). Temporal data is key, capturing health evolution over time and offering insights into disease progression and treatment effectiveness. Time series data, collected at successive time points, shows complex patterns with short- and long-term dependencies, crucial for forecasting and analysis (Zou et al., 2019; Lai et al., 2018). Properly harnessed, this data advances personalized medicine and treatment optimization, making it essential in contemporary research.

## 1.2 Applications of predictive modeling in biomedical time series analysis

Predictive modeling with artificial intelligence (AI) has gained significant traction across various domains, including manufacturing (Altarazi et al., 2019), heat transfer (Al-Hindawi et al., 2023; 2024), energy systems (Huang et al., 2024), and notably, the biomedical field (Cai et al., 2023; Patharkar et al., 2024). Predictive modeling in biomedical time series data involves various approaches for specific predictions and data characteristics. Forecasting models predict future outcomes based on historical data, such as forecasting blood glucose levels for diabetic patients using past measurements, insulin doses, and dietary information (Plis et al., 2014). Classification models predict categorical outcomes, like detecting cardiac arrhythmias from ECG data by classifying segments into categories such as normal, atrial fibrillation, or other arrhythmias, aiding in early diagnosis and treatment (Daydulo et al., 2023; Zhou et al., 2019; Chuah and Fu, 2007). Anomaly detection in biomedical time series identifies outliers or abnormal patterns, signifying unusual events or conditions. For example, monitoring ICU patients' vital signs can detect early signs of sepsis (Mollura et al., 2021; Shashikumar et al., 2017; Mitra and Ashraf, 2018), enabling timely intervention.

Table 1 summarizes the example applications of these models within the context of biomedical time series.

## 1.3 Challenges in biomedical time series data

Regardless of the particular medical application or predictive model type used, models that manage biomedical time series data must tackle the intrinsic challenges posed by clinical and biomedical data. This includes various categories, such as electronic health records (EHRs), administrative data, claims data, patient/disease registries, health surveys, and clinical trials data. As illustrated in Table 2, each biomedical data category presents

distinct challenges regarding quality, privacy, and completeness. During predictive modeling, further challenges arise. Specifically, we will investigate problems associated with missing data and imputation methods, the intricate nature of high-dimensional temporal relationships, and factors concerning the size of the dataset. Addressing these issues is crucial for developing strong and accurate predictive models in medical research.

### 1.3.1 Challenges in handling missing values and imputation methods in biomedical time series

Clinical data is often confronted with the issue of missing values, which can be caused by irregular data collection schedules or unexpected events (Xu et al., 2020). Medical measurements, recorded variably and at different times, may be absent, not documented, or affected by recording errors (Mulyadi et al., 2022), which makes the data irregular. Dealing with missing values in data sets usually involves either directly modeling data sets with missing values or filling in the missing values (a.k.a. imputation) to complete datasets for traditional analysis methods using data imputation techniques.

Current imputation techniques can be divided into four categories: case deletion, basic statistical imputation, machine learning-based imputation (Luo et al., 2018), and aggregating irregularly sampled data into discrete time periods (Ghaderi et al., 2023). Each of these methods comes with specific challenges in the context of handling biomedical temporal data. The deletion or omission of cases may lead to the loss of important information, particularly when the rate of missingness is high, which is critical in sensitive applications such as biomedical predictive modeling, where data is scarce and human lives are at risk. However, in certain cases, it is possible to do data omission without any potential risk to the outcome of the study. For instance, (Pinto et al., 2022), employs interrupted time series analysis to assess the impact of the "Syphilis No!" initiative in reducing congenital syphilis rates in Brazil. The results indicate significant declines in priority municipalities after the intervention. The study showcases the efficacy of public health interventions in modifying disease trends using statistical analysis of temporal data. Data collection needed to be conducted consistently over time and at evenly spaced intervals for proper analysis. To prevent bias due to the COVID-19 pandemic, December 2019 was set as the final data collection point, encompassing 20 months before the intervention (September 2016 to April 2018) and 20 months after the intervention (May 2018 to December 2019). This approach illustrates how the author addressed potential issues of irregular data or missing values in this context.

Contrary to data omission, statistical imputation techniques, such as mean or median imputation offer an alternative that reduces the effect of missing data, however, such methods do not take into account the temporal information but rather offer a summarized statistical imputation that often does not provide accurate replacement of the missing data. This could be critical in biomedical applications with scarce datasets, where the weight of a single data point could heavily affect the predictive power of the model. The use of machine learning-based imputation methods, such as Maximum Likelihood Expectation-Maximization, k-Nearest Neighbors, and Matrix Factorization, might offer a more accurate imputation that takes into account the specificity of the data point contrary to statistical aggregation methods,

TABLE 1 Overview of predictive modeling techniques for biomedical time series and their example applications across healthcare scenarios.

Model type	Description	Bio-medical application example
Forecasting	Predicts a continuous value based on historical data	Predicting blood glucose levels for diabetic patients using past glucose measurements, insulin doses, and dietary information to forecast potential hypo- or hyperglycemic events (Plis et al., 2014)
Classification	Predicts categorical outcomes based on temporal data	Detecting cardiac arrhythmias (such as normal, atrial fibrillation, or other arrhythmias) (Daydulo et al., 2023; Zhou et al., 2019; Chuah and Fu, 2007)
Anomaly Detection	Identifies outliers or abnormal patterns within time series data	Sepsis detection. (Mollura et al., 2021; Shashikumar et al., 2017; Mitra and Ashraf, 2018)

TABLE 2 Overview of clinical data types and challenges. This table lists the main types of clinical and biomedical data, their definitions, and key challenges.

Data type	Definition	Challenges
Electronic Health Records (EHRs)	Digital records of patients medical history, treatments, and outcomes	<ul style="list-style-type: none"> <li>• Data Standardization: Different formats across providers</li> <li>• Data Quality: Missing, incomplete, or inaccurate data</li> <li>• Privacy and Security: Ensuring compliance with regulations like HIPAA</li> <li>• Interoperability: Difficulties in data exchange between systems</li> </ul>
Administrative Data	Data related to healthcare administration, such as hospital admissions and discharge records	<ul style="list-style-type: none"> <li>• Limited Clinical Detail: Lack of in-depth clinical information</li> <li>• Data Timeliness: Potential delays in data availability</li> <li>• Standardization Issues: Variability in recording and categorization</li> <li>• Privacy Concerns: Maintaining patient confidentiality</li> </ul>
Claims Data	Data from insurance claims used for billing and reimbursement	<ul style="list-style-type: none"> <li>• Purpose and Detail: Primarily for billing, may lack clinical details</li> <li>• Lag Time: Delays between care and data availability</li> <li>• Coding Errors: Inaccuracies in coding (e.g., ICD codes)</li> <li>• Complexity: Requires specialized knowledge for interpretation</li> </ul>
Patient/Disease Registries	Databases that track patients with specific conditions or diseases	<ul style="list-style-type: none"> <li>• Data Completeness: Ensuring all relevant data is captured</li> <li>• Data Standardization: Different definitions and methods across registries</li> <li>• Funding and Maintenance: Need for consistent resources</li> <li>• Privacy Issues: Protecting patient confidentiality</li> </ul>
Health Surveys	Data collected from health-related surveys and questionnaires	<ul style="list-style-type: none"> <li>• Response Bias: Non-response or inaccurate self-reporting</li> <li>• Sampling Issues: Ensuring representative samples</li> <li>• Data Quality: Depends on survey design and execution</li> <li>• Timeliness: Time-consuming design, conduct, and analysis</li> </ul>
Clinical Trials Data	Data from controlled trials testing the efficacy of treatments or interventions	<ul style="list-style-type: none"> <li>• Complexity and Cost: Expensive and logistically complex</li> <li>• Regulatory Hurdles: Compliance with regulatory requirements</li> <li>• Data Sharing: Balancing patient confidentiality and proprietary interests</li> <li>• Generalizability: Trial participants may not represent the broader population</li> </ul>

however, many of them still do not consider temporal relations between observations (Luo et al., 2018; Jun et al., 2019), and they usually are computationally expensive. Furthermore, without incorporating domain knowledge, these approaches can introduce bias and lead to invalid conclusions. Both machine learning and statistical techniques may not consider data distribution or variable relationships and may fail to capture complex patterns in multivariate time-series data due to the neglect of correlated variables, potentially resulting in underestimated or overestimated imputed values (Jun et al., 2019). Additionally, in real-time clinical decision support systems, timely and accurate data is crucial, as delays or errors in imputation can lead to incorrect decisions that directly affect patient outcomes. These systems demand high-speed processing, requiring imputation algorithms to be both computationally efficient and accurate. Moreover, the dynamic nature of clinical environments, where patient conditions can change rapidly, necessitates imputation methods that can adapt quickly to evolving data.

Aggregating measurements into discrete time periods can address irregular intervals, but it may lead to a loss of granular information (Ghaderi et al., 2023). Additionally, in time series prediction, missing values and their patterns are often correlated with target labels, referred to as informative missingness (Che et al., 2018). These limitations make it ill-advised to ignore, impute, or aggregate these values when handling biomedical time series data, but rather employ a model that is capable of handling the sparsity and the irregularity of clinical time series data.

### 1.3.2 Complexities of high-dimensional temporal dependencies in biomedical data

Besides missing data challenges, hospitalized patients have a wide range of clinical events that are recorded in their electronic health records (EHRs). EHRs contain two different kinds of data: structured information, like diagnoses, treatments, medication prescriptions, vital signs, and laboratory tests, and unstructured data, like clinical notes and physiological signals (Xie et al., 2022; Lee and Hauskrecht, 2021), making them multivariate or high-dimensional (Niu et al., 2022).

The complexity of the relationships existing in such high-dimensional multivariate time series data can be difficult to capture and analyze. Analysts often try to predict future outcomes based on past data, and the accuracy of these predictions depends on how well the interdependencies between the various series are modeled (Shih et al., 2019). It is often beneficial to consider all relevant variables together rather than focusing on individual variables to build a prediction model, as this provides a comprehensive understanding of correlations in multivariate time series (MTS) data (Du et al., 2020). It thus becomes a requirement for predictive models employed in biomedical applications to take into account correlations among multiple dimensions and make predictions accordingly. It is equally crucial to ensure that only the features with a direct impact on the outcome are considered in the analysis. For instance, the study by Barreto et al. (2023) investigates the deployment of machine learning and deep learning models to forecast patient outcomes and allocate beds efficiently during the COVID-19 crisis in Rio Grande do Norte, Brazil. Out of 20 available features, nine were chosen based on their clinical importance and their correlation with patient outcomes, selected through

discussions with clinical experts to guarantee the model's accuracy and interpretability.

In addition to the inherent high dimensionality of biomedical data sourced from diverse platforms such as EHRs, wearable devices monitoring neurophysiological functions, and intensive care units tracking disease progression through physiological measurements (Allam et al., 2021), also display a natural temporal ordering. This temporal structure demands a specialized analytical approach distinct from that applied to non-temporal datasets (Zou et al., 2019). The temporal dependency adds significant complexity to modeling due to the presence of two distinct recurring patterns: short-term and long-term. For instance, short-term patterns may repeat daily, whereas long-term patterns might span quarterly or yearly intervals within the time series (Lai et al., 2018). Biomedical data often exhibit long-term dependencies, such as those seen in biosignals like electroencephalograms (EEGs) and electrocardiograms (ECGs), which may span tens of thousands of time steps or involve specific medical conditions such as acute kidney injury (AKI) leading to subsequent dialysis (Sun et al., 2021; Lee and Hauskrecht, 2021). Concurrently, short-term dependencies can manifest in immediate physiological responses to medical interventions, such as the administration of norepinephrine and subsequent changes in blood pressure (Lee and Hauskrecht, 2021). Another instance is presented by Valentim et al. (2022), who have created a model to forecast congenital syphilis (CS) cases in Brazil based on maternal syphilis (MS) incidences. The model takes into account the probability of proper diagnosis and treatment during prenatal care. It integrates short-term dependencies by assessing the immediate effects of prenatal care on birth outcomes, and long-term dependencies by analyzing syphilis case trends over a 10-year period. This strategy aids in enhancing public health decision-making and syphilis prevention planning.

Analyzing these recurrent patterns and longitudinal structures in biomedical data is essential to facilitate the creation of time-based patient trajectory representations of health events that facilitate more precise disease progression modeling and personalized treatment predictions (Allam et al., 2021; Xie et al., 2022). By incorporating both short-term fluctuations and long-term trends, robust predictive models can uncover hidden patterns in patient health records, advancing our understanding and application of digital medicine. Failing to consider these recurrent patterns can undermine the accuracy of time series forecasting in biomedical contexts such as digital medicine, which involves continuous recording of health events over time.

Additionally, early detection of diseases is of paramount importance. This can be achieved by utilizing existing biomarkers along with advanced predictive modeling techniques, or by introducing new biomarkers or devices aimed at early disease detection. For instance, early diagnosis of osteoporosis is essential to mitigate the significant socioeconomic impacts of fractures and hospitalizations. The novel device, Osseus, as cited by Albuquerque et al. (2023), addresses this by offering a cost-effective, portable screening method that uses electromagnetic waves. Osseus measures signal attenuation through the patient's middle finger to predict changes in bone mineral density with the assistance of machine learning models. The advantages of using Osseus include enhanced accessibility to osteoporosis screening,

reduced healthcare costs, and improved patient quality of life through timely intervention.

### 1.3.3 Dataset size considerations

The quantity of data available in a given dataset must be carefully considered, as it significantly influences model selection and overall analytical approach. For instance, when patients are admitted for brief periods, the clinical sequences generated are often fewer than 50 data points (Liu, 2016). Similarly, the number of data points for specific tests, such as mean corpuscular hemoglobin concentration (MCHC) lab results, can be limited due to the high cost of these tests, often resulting in less than 50 data points (Liu and Hauskrecht, 2015). Such limited data points pose challenges for predictive modeling, as models must be robust enough to derive meaningful insights from small samples without overfitting.

Conversely, some datasets may have a moderate sample length, ranging from 55 to 100 data points, such as the Physionet sepsis dataset (Reyna et al., 2019; 2020; Goldberger et al., 2000). These moderate-sized datasets offer a balanced scenario where the data is sufficient to train more complex models, but still requires careful handling to avoid overfitting and ensure generalizability.

In other cases, datasets can be extensive, particularly when long-span time series data is collected via sensor devices. These devices continuously monitor physiological parameters, resulting in large datasets with thousands of time steps (Liu, 2016). For example, wearable devices tracking neurophysiological functions or intensive care unit monitors can generate vast amounts of data, providing a rich source of information for predictive modeling. However, handling such large datasets demands models that are computationally efficient and capable of capturing long-term dependencies and complex patterns within the data.

The amount of data available is a major factor in choosing the appropriate model. Sparse datasets require models that can effectively handle limited information, often necessitating advanced techniques for data augmentation and imputation to make the most out of available data. Moderate datasets allow for the application of more sophisticated models, including machine learning and deep learning techniques, provided they are carefully tuned to prevent overfitting. Large datasets, on the other hand, enable the use of highly complex models, such as deep neural networks, which can leverage the extensive data to uncover intricate patterns and relationships.

## 1.4 Strategies in forecasting for biomedical time series data

While our discussion has generally revolved around the challenges in predictive modeling of biomedical temporal data, this review specifically emphasizes forecasting. From the earlier discourse, it is clear that a forecasting model for clinical or biomedical temporal data needs to adeptly manage missing, irregular, sparse, and multivariate data, while also considering its temporal properties and the capacity to model both short-term and long-term dependencies. The model should be able to make multi-step predictions, and the selection of a suitable model is determined by the amount of data available and the temporal length of the time series under consideration.

In this review, we initially examine three main categories of forecasting models: statistical, machine learning, and deep learning models. We look closely at the leading models within each category, assessing their ability to tackle the complexities of biomedical temporal data, including issues like data irregularity, sparsity, and the need to capture detailed temporal dependencies, alongside multi-step predictions. Since each category has its unique advantages as well as limitations in addressing the specific challenges of biomedical temporal datasets, other sets of models mentioned in the literature, known as hierarchical time series forecasting and combination or ensemble forecasting that merge the benefits of various forecasting models to produce more accurate forecasts are also covered.

The rest of the paper is structured as follows: In Section 2, statistical models are introduced. Section 3 covers machine learning models, while Section 4 focuses on deep learning models. This is followed by Section 5, which is a discussion section that summarizes the findings, discusses ensemble as well as hierarchical models, and explores future directions for the application of AI in clinical datasets. Finally, Section 6 concludes the paper.

## 2 Statistical models

The most popular predictive statistical models for temporal data are Auto-Regressive Integrated Moving Average (ARIMA) models, Exponential Weighted Moving Average (EWMA) models, and Regression models which are reviewed in the following sections.

### 2.1 Auto-Regressive Integrated Moving Average models

(Yule, 1927) proposed an autoregressive (AR) model, and (Wold, 1948) introduced the Moving Averaging (MA) model, which were later combined by Box and Jenkins into the ARMA model (Janacek, 2010) for modeling stationary time series. The ARIMA model, an extension of ARMA, incorporates differencing to make the time series stationary before forecasting, represented by ARIMA (p,d,q), where p is the number of autoregressive terms, d is the degree of differencing, and q is the number of moving average terms. ARIMA models have been applied in real-world scenarios, such as predicting COVID-19 cases. Ding et al. (2020) used an ARIMA (1,1,2) model to forecast COVID-19 in Italy. In another study, (Bayyurt and Bayyurt, 2020), utilized ARIMA models for predictions in Italy, Turkey, and Spain, achieving a Mean Absolute Percentage Error (MAPE) value below 10%. Similarly, (Tandon et al., 2022), employed an ARIMA (2,2,2) model to forecast COVID-19 cases in India, reporting a MAPE of 5%, along with corresponding mean absolute deviation (MAD) and multiple seasonal decomposition (MSD) values.

When applying ARIMA models to biomedical data, we select the appropriate model using criteria like Akaike Information Criterion (AIC) or Bayesian information criterion (BIC), estimate parameters using tools like R or Python's statsmodels, and validate the model through residual analysis. ARIMA models are effective for univariate time series with clear patterns, supported by extensive documentation and software, but they require stationarity and may

be less effective for data with complex seasonality. Moreover, if a time series exhibits long-term memory, ARIMA models may produce unreliable forecasts (Al Zahrani et al., 2020), signifying that they are inadequate for capturing long-term dependencies. Additionally, ARIMA models necessitate a minimum of 50 data points in the time series to generate accurate forecasts (Montgomery et al., 2015). Therefore, ARIMA models should not be used for biomedical data that require the modeling of long-term relationships or have a small number of data points.

Several extensions such as Seasonal ARIMA (SARIMA) have been introduced for addressing seasonality. For instance, the research by Liu et al. (2023) examined 10 years of inpatient data on Acute Mountain Sickness (AMS), uncovering evident periodicity and seasonality, thereby establishing its suitability for SARIMA modeling. The SARIMA model exhibited high accuracy for short-term forecasts, assisting in comprehending AMS trends and optimizing the allocation of medical resources. An additional extension of ARIMA, proposed for long-term forecasts, is ARFIMA. In the study by Qi et al. (2020), the Seasonal Autoregressive Fractionally Integrated Moving Average (SARFIMA) model was utilized to forecast the incidence of hemorrhagic fever with renal syndrome (HFRS). The SARFIMA model showed a better fit and forecasting accuracy compared to the SARIMA model, indicating its superior capability for early warning and control of infectious diseases by capturing long-range dependencies. Additionally, it is apparent that ARIMA models cannot incorporate exogenous variables. Therefore, a variation incorporating exogenous variables, known as the ARIMAX model, has been proposed. The study by Mahmudimanesh et al. (2022) applied the ARIMAX model to forecast cardiac and respiratory mortality in Tehran by analyzing the effects of air pollution and environmental factors. The key variables encompass air pollutants (CO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>10</sub>) and environmental data (temperature, humidity). The ARIMAX model is selected for its capacity to include exogenous variables and manage non-static time series data.

For multi-step ahead forecasting in temporal prediction models, two methods exist. The first, known as the plug-in or iterated multi-step (IMS) prediction that involves successively using the single step predictor, treating each prediction as if it were an observed value to obtain the expected future value. The second approach is to create a direct multi-step (DMS) prediction as a function of the observations, and to select the coefficients in this predictor by minimizing the sum of squares of the multi-step forecast errors. Haywood and Wilson (2009) developed a test to decide which of two approaches is more dependable based on a given lead-time. In addition to this test, there are other ways to decide which technique is most suitable for forecasting multiple steps ahead. One of these methods can be used to decide the best choice for multi-step ahead prediction either for ARIMA or other types of models depending on the amount of historical data and the lead-time.

## 2.2 Exponential weighted moving average models

The EWMA method, based on Roberts (2000), uses first-order exponential smoothing as a linear combination of the current observation and the previous smoothed observation. The smoothed observation  $\tilde{y}_t$  at time  $t$  is given by the equation  $\tilde{y}_t =$

$\lambda y_t + (1 - \lambda)y_{t-1}$ , where  $\lambda$  is the weight assigned to the latest observation. This recursive equation requires an initial value  $\tilde{y}_0$ . Common choices for  $\tilde{y}_0$  include setting it equal to the first observation  $y_1$  or the average of available data, depending on the expected changes in the process. The smoothing parameter  $\lambda$  is typically chosen by minimizing metrics such as Mean Squared Error (MSE) or MAPE (Montgomery et al., 2015).

Several modifications of simple exponential smoothing exist to account for trends and seasonal variations, such as Holt's method (Holt, 2004) and Holt-Winter's method (Winters, 1960). These can be used in either additive or multiplicative forms. For modeling and forecasting biomedical temporal data, the choice of method depends on the data characteristics. Holt's method is more appropriate for data with trends. On the other hand, EWMA is suitable for stationary or relatively stable data, making it effective in scenarios without a clear trend, such as certain biomedical measurements. For instance, Rachmat and Suhartono (2020) performed a comparative analysis of the simple exponential smoothing model and Holt's method for forecasting the number of goods required in a hospital's inpatient service, assessing performance using error percentage and MAD. Their findings indicated that the EWMA model outperformed Holt's method, as it produced lower forecast errors. This outcome is logical since the historical data of hospitalized patients lack any discernible trend.

EWMA models are also intended for univariate, regularly-spaced temporal data, as demonstrated in the example above (Rachmat and Suhartono, 2020), which uses a single variable (number of goods) over a period of time as input for model construction. This model is not suitable for biomedical data that involves multiple variables influencing the forecast unless its extension for multivariate data is employed. As highlighted by De Gooijer and Hyndman (2006), there has been surprisingly little progress in developing multivariate versions of exponential smoothing methods for forecasting. Poloni and Sbrana (2015) attributes this to the challenges in parameter estimation for high-dimensional systems. Conventional multivariate maximum likelihood methods are prone to numerical convergence issues and high complexity, which escalate with model dimensionality. They propose a novel strategy that simplifies the high-dimensional maximum likelihood problem into several manageable univariate problems, rendering the algorithm largely unaffected by dimensionality.

EWMA models cannot directly handle data that is not evenly spaced, and thus cannot be used to directly model biomedical data with a large number of missing values without imputation. These models are capable of multi-step ahead prediction either through DMS or IMS approach. To emphasize long-range dependencies, the parameter  $\lambda$  can be set to a low value, while a higher value will give more importance to recent past value (Rabyk and Schmid, 2016). The range of  $\lambda$  values typically used for reasonable forecasting is 0.1–0.4, depending on the amount of historical data available for modeling (Montgomery et al., 2015).

## 2.3 Regression models

Several regression models are available, and in this discussion, we focus on two specific types: multiple linear regression (MLR) (Galton, 1886; Pearson, 1922; Pearson, 2023) and multiple

polynomial regression (MPR) (Legendre, 1806; Gauss, 1823). These models are particularly relevant for biomedical data analysis as they accommodate the use of two or more variables to forecast values. In MLR, there is one continuous dependent variable and two or more independent variables, which may be either continuous or categorical. This model operates under the assumption of a linear relationship between the variables. On the other hand, MPR shares the same structure as MLR but differs in that it assumes a polynomial or non-linear relationship between the independent and dependent variables. This review provides examination of these two regression models.

### 2.3.1 Multiple linear regression models

The estimated value of output  $y$  at time  $t$ , denoted as  $y_t$  with a MLR model for a certain set of predictors is given by the following Equation 1.

$$y_t = X_t\beta + \epsilon_t \quad (1)$$

where,  $X_t = (1, x_{1t}, x_{2t}, \dots, x_{kt})$  is a vector of  $k$  explanatory variables at time  $t$ ,  $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$  are regression coefficients, and  $\epsilon_t$  is a random error term at time  $t$ ,  $t = 1, \dots, N$  (Fang and Lahdelma, 2016). It can be solved with least squares method (Pearson, 1901) to obtain the regression coefficients.

$R^2$  value can be calculated to check the accuracy of model fitting. The value of  $R^2$  that is closer to 1 indicates better model performance. Metrics such as Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Theil's inequality coefficient (TIC) are commonly utilized to assess the forecasting model's performance. While RMSE is scale-sensitive, MAPE and TIC are scale-insensitive. Lower values for these three metrics signify a well-fitting forecasting model.

Zhang et al. (2021) developed an MLR model aimed at being computationally efficient and accurate for forecasting blood glucose levels in individuals with type 1 diabetes. These MLR models can predict specific future intervals (e.g., 30 or 60 min ahead). The dataset is divided into training, validation, and testing subsets; missing values are handled using interpolation and forward filling, and the data is normalized for uniformity. The MLR model showed strong performance, especially in 60-min forward predictions, and was noted for its computational efficiency in comparison to deep learning models. It excelled in short-term time series forecasts with significant data variability, making it optimal for real-time clinical applications.

### 2.3.2 Multiple polynomial regression models

The estimated value of  $y_t$  with say a second-order MPR model for a certain set of predictors is given by the following Equation 2.

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_n x_{nt} + \beta_{n+1} x_{1t}^2 + \beta_{n+2} x_{1t} x_{2t} + \dots + \beta_{2n} x_{1t} x_{nt} + \beta_{2n+1} x_{2t}^2 + \beta_{2n+2} x_{2t} x_{3t} + \dots + \epsilon \quad (2)$$

where,  $\beta_1, \beta_2$  are regression coefficients,  $x_1, x_2, \dots, x_{nt}$  are predictor variables, and  $\epsilon$  is a random error. The ordinary least squares method (Legendre, 1806; Gauss, 1823) is applicable for solving this, similar to how it is used with MLR models. Furthermore, the evaluation metrics utilized for MLR are also suitable for MPR models.

Wu et al. (2021) utilized US COVID-19 data from January 22 to July 20 (2020), categorizing it into nationwide and state-level data sets. Positive cases were identified as Temporal Features

(TF), whereas negative cases, total tests, and daily positive case increases were identified as Characteristic Features (CF). Various other features were employed in different manners, such as the daily increment of hospitalized COVID-19 patients. An MPR model was created for forecasting single-day outcomes. The model consisted of pre-processing and forecasting phases. The pre-processing phase included quantifying temporal dependency through time-window lag adjustment, selecting CFs, and performing bias correction. The forecasting phase involved developing MPR models on pre-processed data sets, tuning parameters, and employing cross-validation techniques to forecast daily positive cases based on state classification.

The various applications of multiple regression models stated above, linear or polynomial, reveal their inability to directly capture temporal patterns. Although these models can accommodate multiple input variables, their design limits them to forecasting a singular outcome with one model. One of the extensions proposed to tackle this problem is multivariate MLR (MVMLR). Suganya et al. (2020) employs MVMLR to forecast four continuous COVID-19 target variables (confirmed cases and death counts after one and 2 weeks) using cumulative confirmed cases and death counts as independent variables. The methodology includes data preprocessing, feature selection, and model evaluation using metrics like Accuracy,  $R^2$  score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

It is clear from the design of the regression models that they are unable to process missing input data. Unless all the predictor variables are present or substituted, the value of the output variable cannot be determined. Therefore, it becomes essential to apply imputation techniques prior to employing the regression models for forecasting.

The regression models do not usually require a large amount of data; it has been demonstrated to be effective with as few as 15 data points per case (Filipow et al., 2023). Multi-step ahead prediction can be accomplished with either IMS or DMS approaches when dealing with temporal data like previous cases. Nonetheless, as mentioned previously, since these methods do not inherently capture temporal dependencies, forecasts can be generated as long as the temporal order is maintained while training, and testing the model.

## 3 Machine learning models

Many machine learning models are employed to construct forecasting models for temporal data sets. The most popular models for temporal data sets include Support vector regression (SVR), k-nearest neighbors regression (KNNR), Regression trees (Random forest regression [RFR]), Markov process (MP) models, Gaussian process (GP) models. We will examine these techniques in the following sections.

### 3.1 Support vector regression

The origin of Support Vector Machines (SVMs) can be traced back to Vapnik (1999). Initially, SVMs were designed to address the issue of classification, but they have since been extended to the

realm of regression or forecasting problems (Vapnik et al., 1996). The SVR approach has the benefit of transforming a nonlinear problem into a linear one. This is done by mapping the data set  $x$  into a higher-dimensional, linear feature space. This allows linear regression to be performed on the new feature space. Various kernels are employed to convert non-linear data into linear data. The most commonly used are linear kernel, polynomial kernel, and radial basis or Gaussian kernel.

Upon transforming a nonlinear dataset  $x$  into a higher-dimensional, linear feature space, the prediction function  $f(x)$  is expressed by Equation 3.

$$f(x) = w^T \phi(x) + b \quad (3)$$

The SVR algorithm solves a nonlinear regression problem by transforming the training data  $x_i$  (where  $i$  ranges from one to  $N$ , with  $N$  being the size of the training data set) into a new feature space, denoted by  $\phi(x)$ . This transformation allows establishing a linear relationship between input and output, using the weight matrix  $w$  and bias matrix  $b$  to further refine the model.

In SVR, selecting optimal hyperparameters ( $C, \epsilon$ ) is crucial for accurate forecasting. The parameter  $C$  controls the balance between minimizing training error and generalization. A higher  $C$  reduces training errors but may overfit, while a lower  $C$  results in a smoother decision function, possibly sacrificing training accuracy. The parameter  $\epsilon$  sets a tolerance margin where errors are not penalized, forming an  $\epsilon$ -tube around predictions. A larger  $\epsilon$  simplifies the model but may underfit, whereas a smaller  $\epsilon$  provides more detail, potentially leading to overfitting. Optimal values for  $C$  and  $\epsilon$  may require additional methods (Liu et al., 2021).

SVR is often combined with other algorithms for parameter optimization. Evolutionary algorithms frequently determine SVR parameters. For example, Hamdi et al. (2018) used a combination of SVR and differential evolution (DE) to predict blood glucose levels with continuous glucose monitoring (CGM) data. The DE algorithm was used to determine the optimal parameters of the SVR model, which was then built based on these parameters. The model was tested using real CGM data from 12 patients, and RMSE was used to evaluate its performance for different prediction horizons. The RMSE values obtained were 9.44, 10.78, 11.82, and 12.95 mg/dL for prediction horizons (PH) of 15, 30, 45, and 60 min, respectively. It should be noted that when these evolutionary algorithms are employed for determining parameters, SVR encounters notable disadvantages, including a propensity to get stuck in local minima (premature convergence).

Moreover, SVR can occasionally lack robustness, resulting in inconsistent outcomes. To mitigate these challenges, hybrid algorithms and innovative approaches are applied. For instance, Empirical Mode Decomposition (EMD) is employed to extract non-linear or non-stationary elements from the initial dataset. EMD facilitates the decomposition of data, thereby improving the effectiveness of the kernel function Fan et al. (2017).

Essentially, SVR is an effective method for dealing with MTS data (Zhang et al., 2019). SVR, which operates on regression-based extrapolation, fits a curve to the training data and then uses this curve to predict future samples. It allows for continuous predictions rather than only at fixed intervals, making it applicable to irregularly spaced time series (Godfrey and Gashler, 2017). Nonetheless,

due to its structure, SVR struggles to capture complex temporal dependencies (Weerakody et al., 2021).

It is suitable for smaller data sets as the computational complexity of the problem increases with the size of the sample Liu et al. (2021). It excels at forecasting datasets with high dimensionality Gavrishchaka and Banerjee (2006) due to the advanced mapping capabilities of kernel functions Fan et al. (2017). Additionally, multi-step ahead prediction in the context of SVR's application to temporal data can be achieved either with the DMS or IMS approach (Bao et al., 2014).

## 3.2 K-nearest neighbors regression

In 1951, Evelyn Fix and Joseph Hodges developed the KNN algorithm for discriminant examination analysis (Fix and Hodges, 1989). This algorithm was then extended to be used for regression or forecasting. The KNN method assumes that the current time series segment will evolve in the future in a similar way to a past time series segment (not necessarily a recent one) that has already been observed (Kantz and Schreiber, 2004). The task is thus to identify past segments of the time series that are similar to the present one according to a certain norm. Given a time series  $y_N(N)$  with  $N$  samples, the segment made of the last  $m$  samples is denoted as  $y_M(N)$ , reflecting the current disturbance pattern. The KNN algorithm searches for  $k$  past time series intervals most comparable to  $y_M(N)$  within the memory  $y_N(N)$  using various distance metrics. For each nearest neighbor, a following time series of length  $h$  is generated, known as prediction contributions. Forecasting can then be done using unweighted or weighted approaches. In the unweighted approach, the prediction is the mean of the prediction contributions. In the weighted approach, the prediction is a weighted average based on the distance of each nearest neighbor from the current segment. Weights are assigned inversely proportional to the distances.

Gopakumar et al. (2016) employed the KNN algorithm to forecast the total number of discharges from an open ward in an Australian hospital, which lacked real-time clinical data. To estimate the next-day discharge, they used the median of similar discharges from the past. The quality of the forecast was evaluated using the mean forecast error (MFE), MAE, symmetric MAE (SMAPE), and RMSE. The results of these metrics were reported to be 1.09, 2.88, 34.92%, and 3.84, respectively, with an MAE error improvement of 16.3% over the naive forecast.

KNN regression is viable for multivariate temporal datasets, as illustrated by Al-Qahtani and Crone (2013). Nevertheless, its forecasting accuracy diminishes as the dimensionality of the data escalates. Consequently, it is critical to meticulously select pertinent features that impact the target variable to enhance model performance.

KNN proves effective for irregular temporal datasets (Godfrey and Gashler, 2017) due to its ability to identify previous matching patterns rather than solely depending on recent data. This distinctive characteristic renders KNN regression a favored choice for imputing missing data (Aljuaid and Sasi, 2016) prior to initiating any forecasting. Furthermore, it excels in capturing seasonal variations or local trends, such as aligning the administration of a medication that elevates blood pressure with a low blood



pressure condition. Conversely, its efficacy in identifying global trends is limited, particularly in scenarios like septic shock, where multiple health parameters progressively deteriorate over time (Weerakody et al., 2021).

The KNN algorithm necessitates distance computations for  $k$ -nearest neighbors. Selecting an appropriate distance metric aligned with the dataset's attributes is essential, with Euclidean distance being prevalent, though other metrics may be more suitable for specific datasets. Ehsani and Drabløs (2020) examines the impact of various distance measures on cancer data classification, using both common and novel measures, including Sobolev and Fisher distances. The findings reveal that novel measures, especially Sobolev, perform comparably to established measures.

As the size of the training dataset increases, the computational demands of the algorithm also rise. To mitigate this issue, approximate nearest neighbor search algorithms can be employed (Jones et al., 2011). Furthermore, the algorithm requires a large amount of data to accurately detect similar patterns. Several methods have been suggested to accelerate the process; for example, (Garcia et al., 2010), presented two GPU-based implementations of the brute-force kNN search algorithm using CUDA and CUBLAS, achieving speed-ups of up to 64X and 189X over the ANN C++ library on synthetic data.

Similarly to other forecasting models, KNN is applicable for multistep ahead predictions using strategies such as IMS or DMS (Martínez et al., 2019). It is imperative to thoroughly analyze the clinical application and characteristics of the clinical data prior to employing KNN regression for forecasting, given its unique attributes. Optimizing the number of neighbors ( $k$ ) and the segment length ( $m$ ) through cross-validation is crucial. Employing appropriate evaluation metrics (e.g., MFE, MAE, SMAPE, RMSE) is necessary to assess the model's performance.

### 3.3 Random forest regression

Random Forests (RFs), introduced by Breiman (2001), are a widely-used forecasting data mining technique. According to Bou-Hamad and Jamali (2020), they are tree-based ensemble methods used for predicting either categorical (classification) or numerical (regression) responses. In the context of regression, known as Random Forest Regression (RFR), RF models strive to derive a prediction function  $f(x)$  that reduces the expected value of a loss function  $L(Y, f(X))$ , with the output  $Y$  typically evaluated using the squared error loss. RFR builds on base learners, where each learner is a tree trained on bootstrap samples of the data. The final prediction is the average of all tree predictions as shown by Equation 4.

$$f(x) = \frac{1}{K} \sum_{k=1}^K I_k(x) \quad (4)$$

where  $K$  is the number of trees, and  $I_k(x)$  is the  $k$ -th tree. Trees are constructed using binary recursive partitioning based on criteria such as MSE.

Zhao et al. (2019) developed a RFR model to forecast the future estimated glomerular filtration rate (eGFR) values of patients to predict the progression of Chronic Kidney Disease (CKD). The data set used was from a regional health system and included 120,495 patients from 2009 to 2017. The data was divided into three

tables: eGFR, demographic, and disease information. The model was optimized through grid-search and showed good fit and accuracy in forecasting eGFR for 2015–2017 using the historical data from the past years. The forecasting accuracy decreased over time, indicating the importance of previous eGFR records. The model was successful in predicting CKD stages, with an average  $R^2$  of 0.95, 88% Macro Recall, and 96% Macro Precision over 3 years.

The study presented in Zhao et al. (2019) indicates that RFR is effective for forecasting multivariate data. Another research by Hosseinzadeh et al. (2023) found that RFR performs better with multivariate data than with univariate data, especially when the features hold substantial information about the target. Research by Tyralis and Papacharalampous (2017) indicated that RF incorporating many predictor variables without selecting key features exhibited inferior performance relative to other methods. Conversely, optimized RF utilizing a more refined set of variables showed consistent reliability, highlighting the importance of thoughtful variable selection.

Similar to SVR, RFR is able to process non-linear information, although it does not have a specific design for capturing temporal patterns (Helmini et al., 2019). RFR is capable of handling irregular or missing data. El Mrabet et al. (2022) compared RFR for fault detection with Deep Neural Networks (DNNs), and found that RFR was more resilient to missing data than DNNs, showing its superior ability to manage missing values. To apply RFR to temporal data, it must be suitably modeled. As an example, Hosseinzadeh et al. (2023) has demonstrated one of the techniques, which involves forecasting stream flow by modeling the RFR as a supervised learning task with 24 months of input data and corresponding 24 months of output sequence. The construction of sequences involves going through the entire data set, shifting 1 month at a time. The study showed that extending the look-back window beyond a certain time frame decreases accuracy, indicating RFR's difficulty in capturing long-term dependencies when used in temporal modeling context. For a forecasting window of 24 months, the look-back window must be at least 24 months to avoid an increase in MAPE. This implies that although RFR can be used for temporal modeling, its effectiveness is more in capturing short-term dependencies rather than long-term ones. The experiments conducted by Tyralis and Papacharalampous (2017) also support this, showing that utilizing a small number of recent variables as predictors during the fitting process significantly improves the RFR's forecasting accuracy.

RFR can be used to forecast multiple steps ahead, similar to other regression models used for temporal forecasting (Alhnaity et al., 2021). Regarding data management, RFR necessitates a considerable volume of data to adjust its hyperparameters. It can swiftly handle such extensive datasets, leading to a more accurate model (Moon et al., 2018).

### 3.4 Markov process models

Two types of Markov Process (MP) models exist: Linear Dynamic System (LDS) and Hidden Markov Model (HMM). Both of these models are based on the same concept: a hidden state variable that changes according to Markovian dynamics can be measured. The learning and inference algorithms for both models are similar in structure. The only difference is that the HMM uses a discrete state

variable with any type of dynamics and measurements, while the LDS uses a continuous state variable with linear-Gaussian dynamics and measurements. These models are discussed in more detail in the following sections.

### 3.4.1 Linear dynamic system

LDS, introduced by Kalman (1963), models the dynamics of sequences using hidden states and discrete time. It assumes evenly spaced time intervals within sequences, where the state transition and state-observation probabilities are given by  $q_i$  and  $o_i$ , respectively. These probabilities are determined by the Equations 5, 6.

$$q_i = Aq_{i-1} + \epsilon_i \quad (5)$$

$$o_i = Bq_i + \zeta_i \quad (6)$$

The terms  $A$  and  $B$  represent the transition and emission matrices, respectively, whereas  $\epsilon_i$  and  $\zeta_i$  denote Gaussian noise components. Specifically, the stochastic element  $\epsilon_i$  adheres to a zero-mean Gaussian distribution  $\epsilon_i \sim \mathcal{N}(0, P)$ , characterized by a zero-mean vector and covariance matrix  $P$ . On the other hand, the stochastic component  $\zeta_i$  follows a zero-mean Gaussian distribution  $\zeta_i \sim \mathcal{N}(0, R)$ , which is also characterized by a zero-mean vector and covariance matrix  $R$ . The initial state distribution ( $q_1$ ) is defined, with mean  $\xi$  and covariance matrix  $\psi$ , i.e.,  $q_1 \sim \mathcal{N}(\xi, \psi)$ . The set of LDS parameters is denoted as  $\lambda = (A, B, P, R, \xi, \psi)$ . In applied scenarios, these parameters necessitate estimation from empirical data. Two standard approaches for learning LDS are the Expectation-Maximization (EM) (Ghahramani and Hinton, 1996) and spectral learning algorithms (Katayama, 2005; Overchee and Moor, 1996; Doretto et al., 2003). EM iteratively maximizes the likelihood of observations by cycling between expectation (E-step) and maximization (M-step). It is precise but can be slow and prone to local optima, especially with limited training data. Spectral learning algorithms provide a non-iterative, closed-form solution using singular value decomposition (SVD) to estimate LDS parameters. They are faster but may be less precise than EM.

A new data-driven state-space dynamic model was developed by Wang et al. (2014) using an extended Kalman filter to estimate time-varying coefficients based on three variate time series data corresponding to glucose, insulin, and meal intake from type 1 diabetic subjects. This model was used to forecast blood glucose levels and was evaluated against a standard model (forgetting-factor-based recursive ARX). The results showed that the proposed model was superior in terms of fit, temporal gain, and J index, making it better for early detection of glucose trends. Furthermore, the model parameters could be estimated in real time, making it suitable for adaptive control. This model was tested for various prediction horizons, demonstrating the model's suitability for multi-step ahead prediction.

The LDS is apt for modelling multivariate temporal data, yet it is confined to data sampled at regular time intervals. As a result, its application to irregularly spaced data (Shamout et al., 2021) or time series with missing values may be problematic. In such instances, modifications and extension are needed. For example, Liu et al. (2013) presented a novel probabilistic method for modeling clinical time series data that accommodates irregularly sampled observations using LDS combined with GP models. They defined

the model by a series of GPs, each confined to a finite window, with dependencies between consecutive GPs represented via an LDS. Their experiments on real-world clinical time series data demonstrate that their model excels in modeling clinical time series and either outperforms or matches alternative time series prediction models.

Typically, implementing the LDS model starts with thorough data preparation, requiring uniform sampling. In cases of irregular sampling or datasets with missing values, proper management through interpolation or imputation is essential for using the model without alterations, as mentioned above. The model architecture is constructed using hidden state variables ( $q_i$ ) to encapsulate the latent processes, alongside measurable observation variables ( $o_i$ ) representing directly observable quantities. Parameters such as the state transition matrix ( $A$ ), the emission matrix ( $B$ ), and the covariance matrices for process noise ( $P$ ) and observation noise ( $R$ ) should be initialized based on prior knowledge or through randomization techniques. Parameter learning is facilitated through the EM algorithm or spectral learning methods, with practical considerations dictating the choice: EM being preferred for its precision with limited datasets and spectral methods for their computational expediency.

The LDS or Kalman filter remains a cornerstone for tracking and estimation due to its attributes of simplicity, optimality, tractability, and robustness. However, nonlinear system applications present complex challenges, often mitigated by the Extended Kalman Filter (EKF) (Lewis, 1986) which linearizes nonlinear models to leverage the linear Kalman filter. Also, various advancements have been proposed for LDS, particularly when addressing nonlinear or non-Gaussian dynamics. For example, approximate filtering methodologies such as the unscented Kalman filter (Julier and Uhlmann, 1997), alongside Monte Carlo-based techniques including the particle filter (Gordon et al., 1993) and the ensemble Kalman filter (Evensen, 1994), are also utilized similar to EKF. Model evaluation is conducted through cross-validation employing metrics such as MSE or RMSE. For forecasting applications, the model can be employed for one-step ahead forecasts or extended to iterative multi-step predictions.

### 3.4.2 Hidden markov model

Hidden Markov Models (HMMs), introduced by Baum and colleagues in the late 1960s and early 1970s (Baum and Petrie, 1966; Baum and Eagon, 1967; Baum and Sell, 1968; Baum et al., 1970; Baum, 1972), are powerful tools for linking hidden states with observed events, assuming an underlying stochastic process. An HMM consists of a set of hidden states, a transition probability matrix, a sequence of observations, observation likelihoods, and an initial state distribution. A critical assumption in HMMs is output independence, where the probability of an observation depends solely on the state that produced it.

HMMs address three fundamental problems: (1) Likelihood estimation: Using the forward or backward algorithm to compute the probability of an observed sequence given the model parameters; (2) Decoding: Employing the Viterbi algorithm to determine the optimal sequence of hidden states corresponding to a sequence of observations; and (3) Learning: Applying the Baum-Welch algorithm, a special case of the EM algorithm, to estimate HMM parameters from observation sequences.

Sotoodeh and Ho (2019) proposed a novel feature representation based on the HMM to predict the length of stay of patients admitted to the ICU. This representation was composed of a specified time resolution and a summary statistic calculated for a specific time window for each feature (e.g., average, most recent, maximum, etc.). An HMM was then trained on these features, and used to generate a series of states for each patient, with the first and last states being used as it was thought that these could better explain the variance in the length of stay. This feature matrix was then used as the input to a regression model to estimate the length of stay. Experiments were conducted to determine the optimal number of states, overlapping or non-overlapping time windows, aggregation of ICU types, summary measure for each time window, and selection of time window probabilities. The model was compared to other baseline models, and was found to have a lower RMSE than all of them.

It is evident from the application here that HMM is capable of dealing with multivariate data. Additionally, it is designed to process temporal data that is spaced at regular intervals of time (Shamout et al., 2021). Unfortunately, it is not able to process temporal data that is irregular or has missing values. Cao et al. (2015) employed both DMS and IMS strategies to forecast multiple future system states and anticipate the evolution of a fault in the Tennessee Eastman (TE) chemical process using HMM. They reported the accuracy of 1,2,3,,20 step-ahead predictions, which were similar for both approaches, with the DMS approach being slightly more accurate than the IMS approach. This is understandable, as the IMS approach has to contend with additional complexities, such as cumulative errors, decreased precision, and increased uncertainty. This demonstrates the capability of HMM to make predictions for multiple steps in the future.

HMM can be constructed using either raw time series data or extracted features. Samaee and Kobravi (2020) introduced a forecasting model aimed at forecasting the timing of tremor bursts with a nonlinear hidden Markov model. This model was trained using the Baum-Welch algorithm, employing both raw Electromyogram (EMG) data and extracted features such as integrated EMG, mean frequency, and peak frequency. The study found that an HMM trained on raw EMG data performed better at forecasting tremor occurrences, suggesting that raw data more accurately captures tremor dynamics compared to extracted features. This is likely due to the short time window being insufficient for feature-based methods. Therefore, it is crucial to determine whether raw time series data or extracted features yield better performance in HMM construction.

In general, MP models are well-recognized for their efficacy in capturing short-term relationships (Manaris et al., 2011) between adjacent symbols or sequences with strong inter-symbol ties. However, they prove inadequate for representing long-distance dependencies between symbols that are spatially or temporally distant (Yoon and Vaidyanathan, 2006; Manaris et al., 2011). To enhance the representational scope of these models, certain methodologies must be employed. For instance, Yoon and Vaidyanathan (2006) proposed context-sensitive HMMs capable of capturing long-distance dependencies, thereby enabling robust pairwise correlations between distant symbols.

Additionally, a limitation of Markov models is that the intrinsic dimensionality of its hidden states is not known beforehand. If the dimensionality is too large, there is a risk of the model becoming

overfitted. Therefore, it is often necessary to try out different training sizes and intrinsic dimensionality of the hidden states to create a model that fits (Liu, 2016).

### 3.5 Gaussian process models

The Gaussian process (GP), introduced by Williams and Rasmussen (Williams and Rasmussen, 2006), is a non-parametric, non-linear Bayesian model in statistical machine learning. A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution. This model extends the multivariate Gaussian to infinite-sized collections of real-valued variables, defining the distribution over random functions. A GP is represented by the mean function:  $m(x) = \mathbb{E}[f(x)]$ , and the covariance function:  $K^G(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$ , where  $f(x)$  is a real-valued process and  $x$  and  $x'$  are two input vectors.

In the context of biomedical temporal data, GP shows promises for modeling and forecasting due to their flexibility and ability to incorporate uncertainty. For example, GP can be used to model patient vital signs over time or predict disease progression (Siami-Namini et al., 2019). The key advantage of GP is their ability to provide uncertainty estimates along with predictions, which is crucial in biomedical applications where uncertainty quantification can inform clinical decisions. The GP can compute the distribution of function values for any set of inputs. This initial distribution, known as the prior, is a multivariate Gaussian represented by Equation 7.

$$f(X^*) \sim \mathcal{N}(m(X^*), K^G(X^*, X^*)) \quad (7)$$

When given observed data, the GP updates this to the posterior distribution, which also follows a multivariate Gaussian. This updated distribution incorporates the observed data, providing more accurate predictions. The posterior distribution is influenced by the observed values and accounts for noise in the data.

GPs extend the multivariate Gaussian distribution into an infinite function space, making them suitable for time series modeling. They can handle observations taken at any time, whether regularly or irregularly spaced, and can make future predictions by calculating the posterior mean for any given time index. Additionally, GPs can act as non-linear transformation operators by replacing the linear transformations used in traditional temporal models with GP, offering a flexible approach to modeling complex data.

GP parameters consist of mean and covariance function parameters. The mean function, dependent on time, represents the expectation before observations. In cases of uncertain trend directions, constant-offset mean functions are common. If prior knowledge about the long-term trend exists, it can be incorporated into GP models, optimizing mean function parameters using gradient-based methods. In clinical scenarios with diverse patient ages and circumstances, aligning time origins is challenging. A practical approach is setting mean functions to a constant ( $m(t) = M$ ), making the GP time-invariant. The constant  $M$  is determined by averaging all patient observations. To optimize the covariance function parameters  $\Theta$ , one can maximize the marginal likelihood  $p(Y|X)$ . The log marginal likelihood for GP is calculated

where  $Y$  includes all training observations. The covariance matrix for noisy observations is represented by  $K^Y$ . It is calculated as  $K^Y = K^G + \sigma^2 I$ , where  $K^G$  is the covariance matrix for noise-free function values, and  $\sigma$  is a standard deviation of the noise, represented as,  $\epsilon \sim \mathcal{N}(0, \sigma)$ . The partial derivatives of the marginal likelihood with respect to each parameter in  $\Theta$  are then derived. These derivatives are used in gradient-based optimization methods to maximize  $p(Y|X)$ , thereby optimizing the covariance function parameters.

A prevalent limitation of GP models pertains to their high computational demands. Sparse GP methodologies have been devised to mitigate this challenge (Williams and Rasmussen, 2006; Quinero-Candela and Rasmussen, 2005), primarily by identifying a subset of pseudo inputs to alleviate computational load. Further optimization of computational efficiency can be achieved through the application of the Kronecker product (Stegle et al., 2011), synchronization of training data across identical time intervals for each dimension (Evgeniou et al., 2005), or the implementation of recursive algorithms tailored for online settings (Pillonetto et al., 2008). Applications necessitating near real-time retraining are more apt to benefit from these approaches, whereas methods that extend over more prolonged temporal frameworks exhibit reduced sensitivity to such computational constraints. Another shortcoming of GP is that it models each time series separately, disregarding the interactions between multiple variables. To tackle this problem and capture the multivariate behavior of MTS, the multi-task Gaussian process (MTGP) was proposed (Bonilla et al., 2007).

### 3.5.1 Multi-task Gaussian process

MTGP is an extension of GP that models multiple tasks (e.g., MTS) simultaneously by utilizing the learned covariance between related tasks. It uses  $K^C$  to model the similarities between tasks and  $K^G$  to capture the temporal dependence with respect to time stamps. The covariance function of MTGP is given by Equation 8.

$$K^M = K^C \otimes K^G + D \otimes I_T \quad (8)$$

where  $K^C$  is a positive semi-definite matrix and  $K_{j,k}^C$  measures the similarity between time series  $j$  and time series  $k$ .  $D$  is an  $n \times n$  diagonal matrix in which  $D_{j,j}$  is the noise variance  $\delta_j^2$  for the  $j^{\text{th}}$  time series.  $\otimes$  is the Kronecker product.

The parameters of GP-based models are composed of parameters that define the mean and covariance functions. Generally, the covariance function ensures that values of the function for two close times tend to have a high covariance, while values from inputs that are distant in time usually have a low covariance. These parameters can be acquired from data that includes one or multiple examples of time series. The predictions of values at future times are equivalent to the calculation of the posterior distribution for those times.

Proper data preprocessing is essential when building MTGP models for forecasting time series. This involves transformations such as detrending and applying logarithmic adjustments. Methods like spectral mixture kernels or Bayesian Nonparametric Spectral Estimation can be employed for initialization. Post-training, it is vital to visualize and interpret cross-channel correlations to better understand the inherent patterns, thereby supporting practical and accurate forecasting applications (de Wolff et al., 2021).

Shukla (2017) proposed to use MTGP to forecast blood pressure from Photoplethysmogram (PPG) signals and compared

its performance to Artificial Neural Networks (ANNs). Ten features were extracted from the PPG signal, and five of them were chosen as the tasks (or targets) to construct the MTGP model. These features were systolic blood pressure, diastolic blood pressure, systolic upstroke time, diastolic time and cardiac period. Four different ANN models were built based on one or more of the above tasks. The models were evaluated on clinical data from the MIMIC Database, with the absolute error  $e$  calculated for each heart beat as the performance measure. The results showed that the performance of MTGP was either comparable to or better than the ANNs and existing methods of computing BP from non-invasive data. MTGP is thus applicable for modeling multivariate temporal data with multiple prediction targets. In a study by Dürichen et al. (2014), MTGP was employed on three diverse biomedical data sets. The experiments aimed to illustrate that forecasting all correlated variables simultaneously enhanced prediction performance, contrasting with individual variable predictions. MTGP has been demonstrated to be successful in multi-step ahead forecasting for a variety of biomedical domain applications mentioned here, as well as in other domains (Cai et al., 2020).

GP models, with an appropriate choice of covariance function, can capture rapid changes in a time series and can be applied to time series modeling problems by representing observations as a function of time. This means that there is no restriction on when the observations are made or if they are regularly or irregularly spaced in time. Liu (2016) and Cheng et al. (2020) demonstrated that, with the appropriate selection of a covariance function, it is possible to model both the short-term dependencies or long-term correlations of temporal data. GP models also work well with small amounts of data (Liu, 2016). It is possible to predict with a certain degree of certainty (confidence interval) using GP (Roberts et al., 2013), which is usually essential for temporal modeling of medical data that necessitates a certain degree of assurance to be employed by medical professionals to make their decisions. However, this approach has some limitations, the most serious being that the mean function of the GP is a function of time and must be set to a constant value in order to make the GP independent of the time origin. This significantly restricts its ability to represent changes or different modes in time series dynamics.

## 4 Deep learning models

The use of Deep Learning techniques for predicting time series data has gained significant attention. While there are various models available for handling time-series data, in this review, we will focus on some commonly used models for forecasting clinical data sets over time. Specifically, we will explore Recurrent Neural Networks (RNN), Long Short Term Memory Networks (LSTM), and Transformer models.

### 4.1 Recurrent Neural Networks

The concept of RNN was introduced by Elman (1990) for identifying patterns in sequential data. RNNs accept sequential data as input and process it recursively. In an RNN, nodes are linked

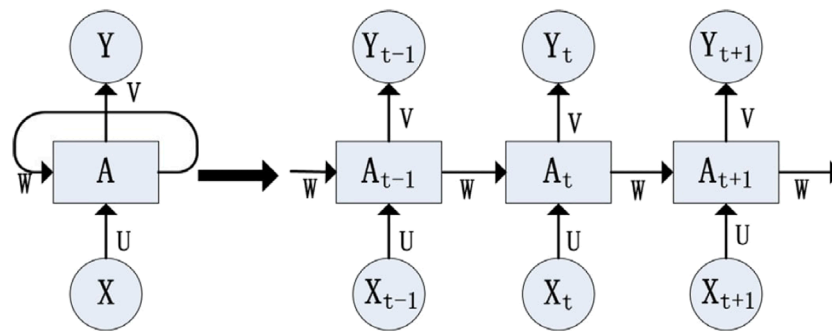


FIGURE 1  
RNN structure (reproduced from Liu et al., 2021, licensed under CC BY 4.0).

sequentially, where the input at time  $t$  depends on the output at time  $t - 1$ . The structure and functions of RNNs are depicted in Figure 1.

In this structure, the input layer ( $X$ ) is weighted by  $U$ , the hidden layer ( $A$ ) by  $W$ , and the output layer ( $Y$ ) by  $V$ . The equations employed for calculations are Equations 9, 10.

$$Y_{t+1} = g(VA_{t+1}) \quad (9)$$

$$A_{t+1} = f(UX_{t+1} + WA_t) \quad (10)$$

The above formula is iterative in nature and can be expanded using the Equation 11 as:

$$Y_{t+1} = Vf(UX_{t+1} + Wf(UX_t + Wf(UX_{t-1} + \dots))) \quad (11)$$

The equation above demonstrates that the RNN network's output  $Y_{t+1}$  is influenced by the current input  $A_{t+1}$ , as well as the previous inputs  $A_t, A_{t-1}, \dots$ . RNNs effectively handle sequential and correlated data by considering historical inputs. The work of Chandra et al. (2021) demonstrates its applicability in multi-step ahead prediction. Although their demonstration focuses on univariate cases, RNN has also been successfully applied to multivariate cases. In their study, Zhu et al. (2020) utilized four data fields for each instance: sampling time, CGM values, meal intake, and insulin dose. They employed a deep learning approach using an extension of RNN, dilated RNN (DRNN), to forecast glucose levels for the next 30 min. The DRNN model exhibits superior performance compared to current models like autoregressive (ARX), SVR, and neural networks for glucose prediction (NNPG), when evaluated on the OhioT1DM dataset. The RMSE values reported are ARX: 20.1 mg/dL, SVR: 21.7 mg/dL, NNPG: 22.9 mg/dL, and DRNN: 18.9 mg/dL. RNNs are frequently used to handle missing values or irregularities in multivariate temporal datasets. There are two main approaches to achieve this: imputation and data generation, or a forecasting approach. When using the first approach, RNNs leverage temporal correlations within each series and correlations among multiple features to fill in missing values or create a time series that captures the original characteristics. On the other hand, the latter approach involves the development of more advanced RNN-based solutions that provide a deeper understanding of the missing data, as well as the patterns and relationships within the data (Weerakody et al., 2021).

Implementing RNNs for modeling and forecasting biomedical temporal data necessitates meticulous attention to data

preprocessing, model structure, tuning of hyperparameters, and evaluation techniques. The recommendations for each aspect are outlined as discussed in Hewamalage et al. (2021). Deseasonalization is advised for datasets exhibiting seasonal trends unless consistent seasonal patterns exist, which RNNs can inherently manage. Data normalization enhances training convergence, while the sliding window approach divides the time series into overlapping sequences for model input. Hyperparameter tuning is crucial for achieving optimal RNN performance. Principal hyperparameters include the learning rate, batch size, and the number of layers. The learning rate must be selected judiciously; for ideal convergence, the Adagrad optimizer typically needs a higher learning rate ranging between 0.01 and 0.9, whereas the Adam optimizer performs effectively within a narrower range of 0.001–0.1. The batch size should be commensurate with the dataset size, and usually, one or two layers are sufficient, as additional layers may result in overfitting. Setting high values for the standard deviation of regularization parameters for Gaussian noise and L2 weight regularization can cause significant underfitting, reducing the neural network's efficacy in generating forecasts. One category of RNN models, stacked RNNs, which involve multiple RNN layers, are employed for forecasting and often utilize skip connections to alleviate vanishing gradient issues. Another category of RNN models, known as sequence-to-sequence (S2S) models, is typically applied in sequential data transformations and is useful for tasks like multi-step forecasting. Assessing RNN performance against traditional methods like ARIMA using standard metrics and cross-validation confirms their competitiveness. Enhancements to RNN methods, such as attention mechanisms and ensemble methods, further boost their performance. Attention mechanisms enable the model to concentrate on relevant parts of the input sequence, while ensemble methods combine several RNN models to produce robust forecasts, reducing biases and variances.

RNNs excel at capturing short-term dependencies (Helmini et al., 2019). They are more sensitive to time series data than traditional convolutional neural networks (CNNs) and can retain memory during data transmission. However, as previously mentioned, when the input sequence lengthens, the network demands more temporal references, leading to a deeper network. In longer sequences, it becomes challenging for the gradient to propagate back from later sequences to earlier ones, resulting in

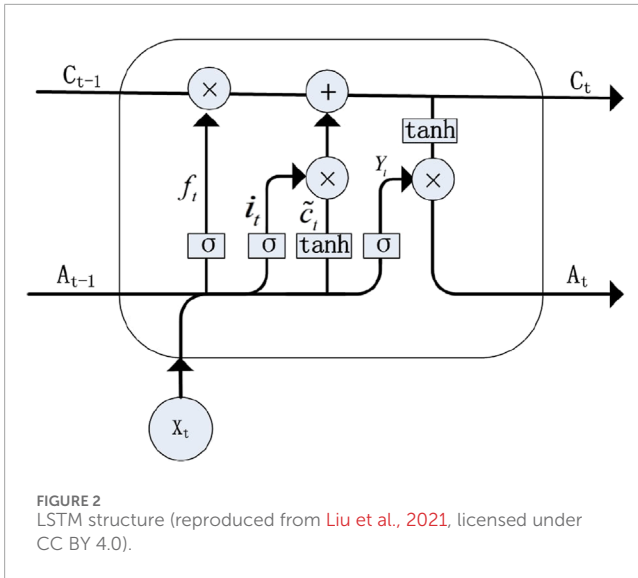


FIGURE 2  
LSTM structure (reproduced from Liu et al., 2021, licensed under CC BY 4.0).

the vanishing gradient problem. Consequently, RNNs struggle with long-term dependencies. To mitigate this vanishing (or exploding) gradient issue, a modification of the RNN known as the long short-term memory (LSTM) model was introduced by Hochreiter and Schmidhuber (1997).

#### 4.1.1 Long Short Term Memory Networks

To overcome the challenges of vanishing and exploding gradients in RNNs, the LSTM model was introduced. This architecture employs a cell state to maintain long-term dependencies, as discussed by Helmini et al. (2019). The model effectively manages gradient dispersion by establishing a retention mechanism between input and feedback. Figure 2 illustrates the LSTM structure (Weerakody et al., 2021). Additionally, LSTM models are proficient in capturing short-term dependencies, primarily through the use of a hidden state. LSTM units are controlled by three gates: the input gate, the output gate, and the forget gate. These gates regulate the flow of information and maintain the cell state, enabling LSTMs to retain important information over long periods. The key equations (Equations 12–17) governing LSTM operations are mentioned as follows:

$$f_t = \sigma(W_{fA}A_{t-1} + W_{fX}X_t + b_f) \quad (12)$$

$$i_t = \sigma(W_{iA}A_{t-1} + W_{iX}X_t + b_i) \quad (13)$$

$$\tilde{c}_t = \tanh(W_{cA}A_{t-1} + W_{cX}X_t + b_c) \quad (14)$$

$$c_t = (f_t \circ c_{t-1}) + (i_t \circ \tilde{c}_t) \quad (15)$$

$$Y_t = \sigma(W_{YA}A_{t-1} + W_{YX}X_t + b_Y) \quad (16)$$

$$A_t = (Y_t \circ \tanh(c_t)) \quad (17)$$

In these equations,  $\sigma$  represents the sigmoid function, and  $\circ$  denotes element-wise multiplication. The forget gate ( $f_t$ ) controls the retention of the previous cell state ( $c_{t-1}$ ), the input gate ( $i_t$ )

manages the incorporation of new information, and the output gate ( $Y_t$ ) determines the output based on the cell state ( $c_t$ ).  $W_{fA}$ ,  $W_{fX}$ ,  $W_{iA}$ ,  $W_{iX}$ , and  $W_{cA}$  are different weights associated with the forget gate, input gate, and the current input unit state.

A deep learning neural network (NN) model based on LSTM with the addition of two fully connected layers was proposed by Idriss et al. (2019), for forecasting blood glucose levels. To determine the optimal parameters for the model, several experiments were conducted using data from 10 diabetic patients. The performance of the proposed LSTM NN, as measured by RMSE, was compared to that of a simple LSTM model and an autoregressive (AR) model. The results indicated that the LSTM NN achieved higher accuracy (mean RMSE = 12.38 mg/dL) compared to both the existing LSTM model (mean RMSE = 28.84 mg/dL) for all patients and the AR model (mean RMSE = 50.69 mg/dL) for 9 out of 10 patients. LSTM is therefore valuable in the representation of time-based information.

One popular extension of the LSTM network is a Bidirectional LSTM (BiLSTM) model which is obtained by modifying the architecture of the LSTM network to include two LSTM layers: one processing the input sequence from left to right (forward direction) and the other from right to left (backward direction). This bidirectional traversal allows the model to have information from both past and future contexts, enhancing its ability to capture complex patterns and dependencies. The outputs from both layers are concatenated at each time step, providing a richer representation of the input sequence. This approach results in improved performance for tasks like time series forecasting, as BiLSTM models can leverage additional training from both directions to better understand sequential data (Abbasimehr and Paki, 2022). For instance, in a study by Said et al. (2021), a bidirectional LSTM (BiLSTM) was employed to analyze multivariate data from countries grouped based on demographic, socioeconomic, and health sector indicators along with the information on lockdown measures, to predict the cumulative number of COVID-19 cases in Qatar from December 1st to 31 December 2020.

LSTM is also combined with multi-head attention mechanisms. This approach aims to address the non-linear patterns and complexities often found in real-world time series data, which traditional forecasting techniques struggle to predict accurately (Siami-Namini et al., 2019). When dealing with irregular temporal data that contain missing values, traditional LSTM models face challenges and may produce suboptimal analyses and predictions. This is because applying the LSTM model to irregular temporal data, either by filling in missing values or using temporal smoothing, does not enable the model to differentiate between actual observations and imputed values. Therefore, caution is advised when using an LSTM model on a dataset where multiple missing values have been imputed.

## 4.2 Transformer models

The Transformer model for natural language processing (NLP) was introduced by Vaswani et al. (2017). This model is composed of an encoder-decoder network, which differs from the traditional sequential structure of RNN. Transformer model utilizes the Self-Attention mechanism to enable parallel training and capture global

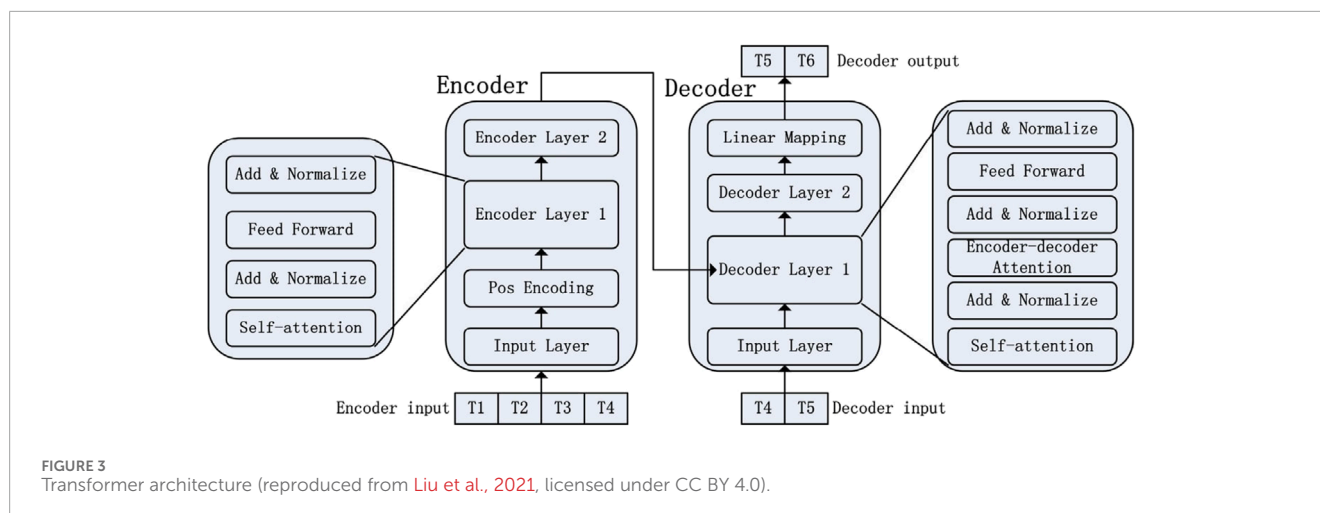


FIGURE 3  
Transformer architecture (reproduced from Liu et al., 2021, licensed under CC BY 4.0).

information. The encoder takes historical time series data as input, while the decoder predicts future values using an autoregressive approach. This means that the decoder's generated output at each step is based on previously generated outputs. To establish a connection between the encoder and decoder, an attention mechanism is employed. This allows the decoder to learn how to effectively focus ("pay attention") on relevant parts of the historical time series before making predictions. The decoder utilizes masked self-attention to prevent the network from accessing future values during training, thereby avoiding information leakage. The typical architecture of the Transformer model is depicted in the Figure 3.

Originally designed for NLP tasks, the Transformer architecture has found application in temporal forecasting as well. To model irregular temporal data, various methods have been proposed. For instance, Tipirneni and Reddy (2022) introduced the Self-supervised Transformer for Time-Series (STraTS) model, which treats each time-series as observation triplets (time, variable, value) instead of matrices as done by conventional methods. This approach eliminates the need for aggregation or imputation. STraTS utilizes a Continuous Value Embedding (CVE) scheme to retain detailed time information without discretization.

The study by Harerimana et al. (2022) utilized a Multi-Headed Transformer (MHT) model to forecast clinical time-series variables from charted vital signs, leveraging the transformer architecture's attention mechanism to capture complex temporal dependencies. The dataset is split into training and testing sets per patient, using past 24-h data for recursive future predictions. Training involves a fixed dimension of 512 for all layers, and the model is evaluated using metrics like Area under the Receiver Operating Characteristic Curve (AUC-ROC), MSE, and MAPE. The MHT model outperforms traditional models (LSTM, Temporal Convolutional Network, TimeNet) in forecasting vital signs, length of stay, and in-hospital mortality, demonstrating superior accuracy and robustness by focusing on influential past time steps, validating its efficacy in handling clinical time-series data.

The Transformer architecture is a relatively new concept, and ongoing research is being conducted to explore its capabilities. For instance, Li et al. (2019) suggest that unlike RNN-based methods, the Transformer enables the model to access any part of the

time series history, disregarding the distance. This characteristic potentially makes it more adept at capturing recurring patterns with long-term dependencies. However, Zeng et al. (2023) presented an opposing viewpoint, questioning the effectiveness of Transformer-based solutions in long-term time series forecasting (LTSF). They argue that while Transformers are adept at capturing semantic correlations in sequences, their self-attention mechanism, which is invariant to permutations, may result in the loss of crucial temporal information necessary for accurate time series modeling. In support of their claim, the researchers introduced LTSF-Linear, a simple one-layer linear model, and discovered that it outperformed more complex Transformer-based LTSF models on nine real-life data sets. In addition, a temporal fusion transformer (TFT) was suggested by Zhang et al. (2022) as a method that effectively captures both short-term and long-term dependencies. Hence, when employing Transformer-based approaches for temporal forecasting, it is crucial to take into account these distinct viewpoints and conduct experiments to determine the most effective modeling technique for the specific forecasting task, considering the presence of short-term and long-term dependencies.

While DL models are capable of generating precise predictions, they are frequently perceived as black-box models that lack interpretability and transparency in their internal processes (Vellido, 2019). This presents a significant issue as medical professionals are often hesitant to trust machine recommendations without a clear understanding of the underlying rationale. In addition, significant quantities of clinical data are utilized to generate standardized inputs for training DL models. The challenge of acquiring extensive clinical data sets poses a challenge in the integration of DL clinical models into real-world clinical systems (Xiao et al., 2018).

## 5 Discussion

This section is comprised of two subsections. The first subsection summarizes the overview of the models and their capacities in addressing the difficulties encountered in forecasting of clinical datasets. The second subsection explores the future prospects

concerning the practical obstacles in implementing AI models for biomedical data modeling.

## 5.1 Summary of models for biomedical temporal data forecasting

### 5.1.1 Summary of statistical, ML, and DL models

This review focuses on predictive models for biomedical temporal data, which face several challenges such as missing values due to irregular data collection or errors. Traditional methods use imputation or deletion, but models that handle missing values without these steps are preferable, as patterns of missing data might hold valuable information termed as “informative missingness”. EHRs often feature MTS data, so models must capture these correlations. Temporal data complexity requires models to consider short-term and long-term patterns. Short-term patterns might involve events like norepinephrine administration linked to recent hypotension, while long-term patterns could involve past acute kidney injury necessitating dialysis. Models should account for these dependencies and support multi-step ahead forecasting for early disease detection. Data availability varies with clinical events, thus impacting model selection. These challenges are crucial for accurate, effective predictions in clinical settings. [Table 3](#) summarizes the advantages and disadvantages of the discussed models, supplemented by literature insights.

Forecasting is categorized into statistical, ML, and DL methods. We focused on models frequently used in biomedical temporal modeling, evaluating their effectiveness. For statistical methods, we analyzed ARIMA, EWMA, and regression models. In ML, we assessed SVR, RFR, KNNR, MP, and GP models. For DL methods, we evaluated RNN, LSTM, and Transformer models. Our analysis found the MTGP model effective for irregularly spaced data, capturing both short-term and long-term dependencies with an appropriate covariance function. It predicts multiple steps ahead and accounts for autocorrelation within and correlation between time series, making it suitable for multivariate temporal analysis with small to moderate data. However, MTGP’s computational cost can be high with large data, and a constant mean function may limit its ability to represent time series dynamics. While MTGP is suitable for biomedical temporal modeling, alternative approaches include improving current models, adopting ensemble methods, or using hierarchical approaches discussed later in this paper.

Improving existing models by incorporating new techniques can address limitations in temporal analysis of biomedical data. For instance, while RNNs struggle with long-range dependencies, they handle other temporal challenges well. To overcome this, [Zhu et al. \(2020\)](#) introduced a dilated RNN, enhancing neuron receptive fields to capture long-term dependencies, enabling 30-min glucose level forecasts. Similarly, HMMs lack long-range correlation modeling. [Yoon and Vaidyanathan \(2006\)](#) introduced context-sensitive HMM (csHMM), capturing long-range correlations by adding context-sensitivity to model states. Additionally, the interpretability in DL models is essential. [Tipirneni and Reddy \(2022\)](#) proposed an interpretable model with outputs as linear combinations of individual feature components. Slight

modifications to the original models can address specific limitations.

Even though various modifications have been suggested to address the shortcomings of individual models, certain limitations remain insurmountable. A recently emerging solution involves combining multiple models to create a fusion model, which allows for the integration of their strengths and mitigation of their weaknesses. These fusion models, also known as combination or ensemble forecasting models, is examined in the next subsection.

### 5.1.2 Fusion models

A different approach to enhance forecasting precision involves merging multiple models, also known as combination or ensemble forecasting models. The paper by [Wang et al. \(2023\)](#) provides a comprehensive overview of the evolution and effectiveness of combining multiple forecasts to enhance prediction accuracy. Combining forecasts, known as “ensemble forecasts,” integrates information from various sources, avoiding the need to identify a single “best” forecast amidst model uncertainty and complex data patterns. The review covers simple combination methods, such as equally weighted averages, which surprisingly often outperform more sophisticated techniques due to their robustness and lower risk of overfitting. Linear combinations, which determine optimal weights based on historical performance, and nonlinear combinations, which account for nonlinear relationships using methods like neural networks, are also discussed. [Wang et al. \(2023\)](#) emphasize the potential of learning-based combination methods, such as stacking and cross-learning, which improve accuracy by training meta-models on multiple time series. In stacking, several forecasting models are trained on the original dataset, and their predictions are combined by a meta-model to provide an optimal forecast. Cross-learning builds on this by utilizing data from various time series to train the meta-model. The review also highlights the crucial role of diversity and precision in forecast combinations, pointing out that successful combinations are enhanced by diverse individual forecasts.

These techniques have been successfully applied to biomedical data forecasting. For example, [Naemi et al. \(2020\)](#) introduced a customizable real-time hybrid model, leveraging the Nonlinear Autoregressive Exogenous (NARX) model along with Ensemble Learning (EL) (RFR and AdaBoost), to forecast patient severity during their stay at Emergency Departments (ED). This model makes use of patient vital signs such as Pulse Rate (PR), Respiratory Rate (RR), Arterial Blood Oxygen Saturation (SpO<sub>2</sub>), and Systolic Blood Pressure (SBP), which are recorded during treatment. The model forecasts the severity of illness in hospitalized patients at ED for the upcoming hour based on their vital signs from the previous 2 hours. The effectiveness of the NARX-EL models is evaluated against other baseline models including ARIMA, a fusion of NARX and LR, SVR, and KNNR. The findings revealed that the proposed hybrid models could predict patient severity with significantly higher accuracy. Furthermore, it was noted that the NARX-RF model excels at predicting abrupt changes and unexpected adverse events in patients’ vital signs, exhibiting an  $R^2$  score of 0.978 and NRMSE of 6.16%. [Kandula et al. \(2018\)](#) used a super-ensemble technique to combine information from



TABLE 3 Advantages and disadvantages of models for handling biomedical temporal data.

Model type	Advantages	Disadvantages
ARIMA	<ul style="list-style-type: none"> <li>- Captures linear dependencies and trends</li> <li>- Interpretable parameters</li> <li>- Works well with stationary data</li> </ul>	<ul style="list-style-type: none"> <li>- Needs data to be stationarized</li> <li>- Lacks ability to handle missing values</li> <li>- Inability to manage multivariate data</li> <li>- Can not capture long-range dependencies</li> </ul>
EWMA	<ul style="list-style-type: none"> <li>- Proficient in temporal modeling</li> <li>- Simple and computationally efficient</li> <li>- Adapts quickly to recent changes in data</li> <li>- Useful for smoothing noisy data</li> <li>- Effective in short-range modeling</li> </ul>	<ul style="list-style-type: none"> <li>- Not suitable for complex patterns</li> <li>- Unsuitable for handling multivariate data</li> <li>- Critical initialization and parameter selection</li> <li>- Capable of long-range modeling with parameter adjustment</li> </ul>
MLR	<ul style="list-style-type: none"> <li>- Interpretable coefficients</li> <li>- Insights into variables' relationships</li> <li>- Performs well with small-mid datasets</li> <li>- Efficiently manages multivariate data</li> <li>- Can be adapted for temporal modeling</li> </ul>	<ul style="list-style-type: none"> <li>- Assumes linear relationships</li> <li>- Sensitive to multicollinearity</li> <li>- Requires features to be linearly related to the target.</li> </ul>
MPR	<ul style="list-style-type: none"> <li>- Can capture higher-order relationships</li> <li>- More flexible than MLR.</li> <li>- Suitable for polynomial relationships</li> <li>- Manages multivariate data efficiently</li> </ul>	<ul style="list-style-type: none"> <li>- Prone to overfitting with high polynomial-degrees</li> <li>- Interpretation of coefficients can be complex</li> <li>- Unable to handle missing values</li> </ul>
SVR	<ul style="list-style-type: none"> <li>- Effective in high-dimensional spaces</li> <li>- Can capture nonlinear relationships</li> <li>- Robust to overfitting with regularization</li> <li>- Manages multivariate data efficiently</li> <li>- Although not designed for temporal modeling, but can be adapted to capture them</li> </ul>	<ul style="list-style-type: none"> <li>- Computationally complex</li> <li>- Needs support from other algorithms for hyperparameter tuning</li> <li>- Lacks robustness resulting in inconsistent outcomes</li> <li>- Struggles to capture complex temporal dependencies</li> <li>- Memory intensive for large datasets</li> </ul>
KNNR	<ul style="list-style-type: none"> <li>- Non-parametric and flexible</li> <li>- Can be adapted for temporal modeling</li> <li>- Proficient in handling missing values</li> <li>- Efficiently manages multivariate data</li> <li>- Effective in short-range modeling due to its unique structure</li> </ul>	<ul style="list-style-type: none"> <li>- Expensive for large datasets</li> <li>- Memory intensive</li> <li>- Falls short in capturing global dependencies</li> </ul>
RFR	<ul style="list-style-type: none"> <li>- Handles nonlinear relationships</li> <li>- Robust to overfitting</li> <li>- Can handle high-dimensional data</li> <li>- Manages multivariate data efficiently</li> <li>- Capable of handling irregular or missing data</li> </ul>	<ul style="list-style-type: none"> <li>- Time consuming for large datasets</li> <li>- Requires careful tuning of hyperparameters</li> <li>- Difficulty in handling long-range dependencies</li> </ul>
LDS	<ul style="list-style-type: none"> <li>- Captures temporal dependencies</li> <li>- Efficiently handles multivariate data</li> <li>- Captures short-term relationships</li> </ul>	<ul style="list-style-type: none"> <li>- Complex parameter tuning</li> <li>- Cannot deal with irregular data</li> <li>- Difficulty with nonlinear relationships</li> </ul>
HMM	<ul style="list-style-type: none"> <li>- Captures hidden influencing states</li> <li>- Useful for sequential data modeling</li> <li>- Efficiently handles multivariate data</li> <li>- Can capture short-term dependencies efficiently</li> </ul>	<ul style="list-style-type: none"> <li>- Training complexity</li> <li>- Lacks interpretability of hidden states</li> <li>- Prone to overfitting when intrinsic dimensionality exceeds data</li> <li>- Struggles with capturing long-term dependencies</li> </ul>
MTGP	<ul style="list-style-type: none"> <li>- Models multiple tasks simultaneously</li> <li>- Captures correlations between tasks</li> <li>- Provides uncertainty estimates</li> <li>- Can forecast efficiently with irregular data</li> <li>- Flexible covariance function that can capture both short-range and long-range dependencies</li> </ul>	<ul style="list-style-type: none"> <li>- Complex to implement and tune</li> <li>- If the GP is made time independent, it restricts the representation of changes in time series dynamics</li> <li>- Computationally intensive on large-scale</li> </ul>
RNN	<ul style="list-style-type: none"> <li>- Proficient in handling missing values</li> <li>- Can handle variable-length sequences</li> <li>- Effective for multivariate sequential data modeling</li> </ul>	<ul style="list-style-type: none"> <li>- Vanishing/exploding gradient problem</li> <li>- Training can be slow</li> <li>- Difficulty with very long-term dependencies</li> </ul>

(Continued on the following page)

TABLE 3 (Continued) Advantages and disadvantages of models for handling biomedical temporal data.

Model type	Advantages	Disadvantages
LSTM	<ul style="list-style-type: none"> <li>- Handles vanishing gradient problem</li> <li>- Captures long-term dependencies effectively</li> <li>- Robust to sequence length variations</li> </ul>	<ul style="list-style-type: none"> <li>- Training complexity</li> <li>- Lacks interpretability</li> <li>- Requires careful hyperparameters tuning</li> <li>- May produce suboptimal analyses and predictions when modeling imputed data</li> </ul>
Transformer	<ul style="list-style-type: none"> <li>- Highly suitable for multivariate temporal modeling</li> <li>- Parallel processing of sequences</li> <li>- Scalable to large datasets</li> <li>- Effective in short-range modeling</li> </ul>	<ul style="list-style-type: none"> <li>- Computationally intensive</li> <li>- Requires large amounts of data</li> <li>- Lacks interpretability</li> <li>- Fine-tuning can be complex</li> <li>- Uncertain effectiveness in managing long-term dependencies</li> </ul>

different forecasting methods robustly. This method yielded a more accurate comprehensive forecast on average than a single model. They compared three forecasting approaches for predicting seven characteristics of seasonal influenza during the 2016–2017 USA season: a mechanistic method, a weighted average of two statistical methods, and a super-ensemble of eight statistical and mechanistic models. The study found the meta-ensemble approach to be the most accurate overall. [Katari et al. \(2023\)](#) employed a combination of Decision Tree (DT) and Ada Boosting algorithms for heart disease prediction. The study highlights the importance of early diagnosis due to high mortality rates. The hybrid model outperformed traditional methods in accuracy, true positive rate (TPR), and precision. Results indicate this combination approach enhances heart disease prediction and aids clinical decision-making.

It is evident that combining forecasts is a crucial component in contemporary forecasting methods for temporal biomedical datasets, providing notable benefits over using single models. Nevertheless, it is crucial to thoroughly understand the data and the aim of forecasting to create an effective ensemble model. Furthermore, it is essential to employ appropriate evaluation metrics for assessing biomedical temporal forecasts. Advancements in research on efficient combination techniques may arise from the capability to manage large and varied datasets, alongside the development of automatic selection methods that balance expertise and diversity when selecting and combining models for forecasting ([Wang et al., 2023](#)).

### 5.1.3 Coherent forecasting

This type of forecasting a.k.a. hierarchical time series (HTS) represents a set of data sequences organized by aggregation constraints, reflecting many real-world applications in research and industry. Forecasting in such hierarchical structures is challenging and time-consuming due to the need to ensure forecasting consistency among hierarchy levels based on their dimensional attributes, such as geography or product categories. Coherent forecasts are essential, meaning that higher-level forecasts must equal the sum of lower-level forecasts. This coherency requirement adds complexity to the original time series forecasting problem ([Sagheer et al., 2021](#)).

For biomedical data scenarios, HTS forecasting is applied in predicting instances similar to emergency medical services (EMS)

requirements ([Rostami-Tabar and Hyndman, 2024](#)) and mortality rates across various U.S. states ([Li and Hyndman, 2021](#); [Li et al., 2024](#)). Forecasting is crucial for EMS as it promotes consistency and synchronized resource allocation, enhancing decision-making processes and leading to better patient outcomes by avoiding the imbalance between demand and resources. In mortality rate predictions, forecasting addresses differences in mortality patterns across different geographic regions. Maintaining adherence between state-level and national-level mortality forecasts is vital for precise policy planning and resource management, aiding in reducing life expectancy disparities and enhancing public health results.

Different reconciliation procedures like top-down, bottom-up, and middle-out have been developed to maintain consistency across levels by generating base forecasts and then adjusting them. These procedures vary in approach: bottom-up starts from the lowest level and aggregates upwards, top-down begins at the highest level and disaggregates downwards, and middle-out combines both methods starting from an intermediate level. Each has its strengths and weaknesses, and none has proven universally superior. [Hyndman et al. \(2011\)](#) proposed an optimal combination approach, which independently forecasts all levels and then combines them using regression to ensure coherence. The Minimum Trace (MinT) method ([Wickramasuriya et al., 2018](#)) is another widely adopted approach for reconciliation. This technique uses the complete covariance matrix of forecast errors to generate a set of coherent forecasts. It aims to minimize the MSE of these coherent forecasts across the whole series, under the assumption of unbiasedness.

The approach detailed by [Rostami-Tabar and Hyndman \(2024\)](#) involves implementing forecast reconciliation for the hierarchical data of ambulance demand. It utilizes an ensemble of models: Exponential Smoothing State Space model (ETS), Poisson regression with Generalized Linear Model (GLM), and time series GLM (TSGLM). It generates base forecasts independently for each hierarchy level and reconciles them using the MinT method, minimizing forecast variances for coherence. Validation is done via time series cross-validation, with accuracy measured by mean absolute scaled error (MASE) and continuous ranked probability scores (CRPS). The methodology by [Li and Hyndman \(2021\)](#) ensures coherent mortality forecasts using a forecast reconciliation approach. Independent state-level forecasts are generated with

the Lee-Carter model and then reconciled using the Minimum Trace (MinT) method together with the sampling approach by [Jeon et al., \(2019\)](#) to ensure consistency with national-level forecasts. Validation is performed using out-of-sample forecasting, with accuracy measured by MAPE and the Winkler score. The study uses U.S. mortality data from 1969 to 2017 and projects rates up to 2027. Another paper by [Li et al. \(2024\)](#) uses boosting with stochastic mortality models as weak learners. The authors extend gradient boosting with age-based and spatial shrinkage, iteratively fitting the Lee-Carter model to residuals and adding graph Laplacian-based penalties to align forecasts of adjacent age groups and states. Validation uses US male mortality data (1969–2019), with forecasting performance assessed using MASE.

Traditionally, methods like ARIMA and exponential smoothing generate base forecasts but fail to capture individual and grouped time series dynamics, especially with time variation or sudden changes. They also struggle with exploiting complete hierarchical information, affecting forecasting efficiency. Recently, ML algorithms like artificial neural networks, extreme gradient boosting, and SVR have been employed to improve accuracy by considering nonlinear relationships and dynamic changes. However, they often still rely on traditional methods and may overlook useful hierarchical information. Overall, HTS forecasting remains a complex problem with ongoing research aimed at finding more efficient and accurate methods to ensure coherent and reliable forecasts across all levels of the hierarchy ([Sagheer et al., 2021](#)). Note: A list and description of open source tools for forecasting is provided in the [Supplementary Material](#) of this article.

## 5.2 Future directions

Extensive research has been conducted to interrogate biomedical temporal data in medical and health applications. Challenges remain, and are summarized into six key areas: (1) standardizing diverse data formats; (2) managing data quality; (3) ensuring model interpretability; (4) protecting patient privacy; (5) enabling real-time monitoring; and (6) addressing bias to create fair models. To grasp the potential future developments, we present a use case to illustrate six future directions within the clinical context. Specifically, taking Mr. Smith (45 years old) as a persona who is concerned about his risk of developing Alzheimer's Disease (AD).

### 5.2.1 Data harmonization to standardize data format

Time series analysis plays a critical role in the early detection of AD by enabling the continuous monitoring of specific biomarkers over time. This approach is crucial for understanding the progression of the disease through its various stages, from preclinical AD to mild cognitive impairment (MCI), and ultimately to dementia. The primary biomarkers used in detecting and monitoring AD include beta-amyloid and tau proteins, which are typically measured in cerebrospinal fluid (CSF), along with imaging biomarkers such as PET scans for assessing beta-amyloid burden and MRI scans for detecting changes in brain volume. These biomarkers are indispensable for identifying the onset and progression of the disease, often before clinical symptoms become evident ([Hernandez-Lorenzo et al., 2022](#)).

During his visit to the physician, Mr. Smith is advised to undergo a series of tests, including genetic screening, neuroimaging, and cognitive assessments. These tests generate a diverse array of data types, ranging from genetic biomarkers to neuroimaging data (e.g., MRI scans) and time-series data derived from cognitive assessments. However, the data collected from Mr. Smith originate from multiple sources: a local hospital, a specialized lab for genetic testing, and a cognitive assessment app. To create a unified dataset, data harmonization is necessary, ensuring consistency across different formats, terminologies, and units. Implementing interoperable technologies can greatly facilitate seamless data exchange across disparate healthcare systems. Future research should focus on developing advanced harmonization techniques for time series data to ensure accurate and consistent integration from various sources. Additionally, integrating multi-modal data, such as clinical, genetic, and imaging information, will be crucial for creating personalized prediction models.

### 5.2.2 Data quality

As Mr. Smith assesses his risk of developing AD, data from various tests play a critical role in forecasting his condition. However, his data may contain missing values due to irregular monitoring, different data collection protocols, or the progression of his condition. Addressing these gaps is crucial for building a reliable predictive model. A promising approach involves filling these gaps and using the missing data as a valuable signal. Missing biomarker readings can be estimated using methods like forward-filling or zero imputation. The model can also incorporate indicators to highlight absent data points, learning from the pattern of missing data. For example, if Mr. Smith's cognitive scores are missing for several months, the model can predict these values and use the absence of scores as a feature. This allows the model to detect patterns that may reveal insights such as health changes or inconsistencies in monitoring.

Ensuring data quality is essential for reliable predictive models in clinical research. Future directions should integrate advanced ML techniques that handle missing data and leverage the temporal patterns surrounding these gaps. By combining models that analyze available data and sequences of missing data, we can improve predictive accuracy, uncover hidden trends, and identify critical periods signaling disease progression. This approach enhances timely, personalized predictions for patients like Mr. Smith.

### 5.2.3 Interpretability

As Mr. Smith assesses his risk of developing AD, advanced ML models analyzing the biomarkers to identify the intervention strategies become crucial. Current models offer predictive power but often function as “black boxes” making it challenging to understand risk factors and the associated impacts. To address this, interpretability methods are essential to know the factors behind risk predictions. One important future direction on interpretability is to use attention mechanisms that prioritize key biomarkers and time points, focusing on early disease prediction characteristics. For example, attention-based models can highlight critical data points, such as changes in biomarkers that signal the onset of AD.

A significant biomarker decline flagged by the model would make the risk assessment more transparent, aiding the physician's

understanding and decisions (e.g., intervention). Alternatively, time-based SHAP (SHapley Additive exPlanations) techniques enhance model prediction transparency by assessing feature importance at specific times. Future work could focus on developing interpretability frameworks in personalized, real-time risk assessments for AD and other conditions, ensuring predictions are accurate and understandable for patients and clinicians.

#### 5.2.4 Data privacy

As Mr. Smith evaluates his risk of AD, the sensitive data gathered requires rigorous privacy safeguards. Data privacy is crucial for legal compliance and maintaining trust in the healthcare system. Sharing sensitive information in research while retaining data utility is challenging. Anonymization is a technique to safeguard against reidentification while maintaining the usefulness of research data. Blockchain technology is another method, providing secure means for sharing data. Federated learning (FL) is also beneficial for collaborative studies, enabling ML models to be trained on Mr. Smith's data locally without the need for centralization, thus decreasing privacy risks. Informed consent is another essential aspect for research purposes. If consent is dynamic, it allows for real-time management, permitting alterations as new research develops. Future directions include implementing these techniques independently or as hybrid frameworks that improve privacy protection without sacrificing research utility. Establishing international standards for these methods is imperative for harmonizing global privacy practices and enhancing security and trust in collaborative research.

#### 5.2.5 Real-time detection

Let's assume, the physician seeing Mr. Smith recommends the use of a wearable device that monitors essential physiological indicators such as sleep patterns and heart rate variability (HRV) to assess his AD risks. Note these devices have already demonstrated potential in identifying early signs of cognitive decline (Saif et al., 2020). With continuous, real-time monitoring, Mr. Smith would be empowered to take proactive actions—such as making lifestyle changes or seeking further medical evaluations—that could potentially delay the progression of the disease. We have observed an emerging trend in health domain to embed wearable devices into regular health surveillance, facilitating the early identification and treatment of AD or other disease conditions. A future direction in predictive modeling is high-fidelity model enabling real-time, or near real-time (e.g., 15 min) detection. Some related research questions include data storage (where data to be stored, cloud or locally), model calibration and fine tuning strategies (e.g., transfer learning).

#### 5.2.6 Bias and fairness

A typical problem in AI models is the possibility of bias if they are trained on unrepresentative datasets. For example, if a model is trained mainly on data from old Asian females, it might inaccurately evaluate Mr. Smith, who is a middle-aged American male. Future directions for utilizing AI-driven models should emphasize making these models unbiased and dependable for various populations. A critical measure is the creation and validation of AI models with datasets that include a broad spectrum of demographics, such as

different ages, ethnicities, and genders. Another approach to ensure fairness in AI algorithms is through regular audits and validation by independent experts. These audits can uncover and fix biases that could distort predictions. Independent audits help guarantee that AI models are equitable and effective for diverse groups thereby offering reliable health assessments. Additionally, it is essential for both healthcare providers and patients to recognize the potential biases in AI tools. By carefully reviewing AI-generated advice alongside clinical expertise and other diagnostic tools, healthcare providers can ensure that the AI model's predictions are accurate and contextual.

## 6 Conclusion

In summary, the review paper outlines the challenges faced in predictive modeling for biomedical temporal data, such as managing missing values, addressing correlations between variables, capturing both short-term and long-term dependencies, performing multi-step ahead predictions, and considering data availability. It assesses models in three categories—statistical, machine learning, and deep learning—to evaluate their effectiveness in forecasting data amidst these challenges. Recognizing limitations in each approach, it discusses alternative methods like model enhancements or ensemble/combo forecasting techniques to potentially improve forecasting accuracy. The review also covers hierarchical forecasting for biomedical datasets with relevant structures. Moreover, it explores issues like data quality, privacy concerns, data harmonization, interpretability, real-time detection, and bias/fairness considerations in integrating AI or ML into clinical practices. These challenges underline the necessity for thorough data evaluation, strong privacy laws, and a deep understanding of the goals of predictive modeling. Moreover, successfully implementing these models necessitates a joint effort from the different fields, along with an inclusive approach that tackles not just the technical aspects of the model but also the broader ethical and fairness issues in healthcare environments.

## Author contributions

AP: Conceptualization, Formal Analysis, Investigation, Writing—original draft, Writing—review and editing. FC: Writing—original draft, Writing—review and editing. FA-H: Writing—review and editing. TW: Conceptualization, Supervision, Writing—original draft, Writing—review and editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. We extend our heartfelt appreciation to the National Science Foundation (NSF)—Partnership for Innovation—“Avoiding Kidney Injuries” (Award #2122901) and CPS: TTP Option: Medium: i-HEAR: immersive Human-On-the-Loop Environmental Adaptation for

Stress Reduction (Award #2038905) for their generous support of this work.

## Acknowledgments

“Writefull” Version 2.2.0 with Overleaf is utilized for improving the quality of the writing for this manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Abbasimehr, H., and Paki, R. (2022). Improving time series forecasting using lstm and attention models. *J. Ambient Intell. Humaniz. Comput.* 13, 673–691. doi:10.1007/s12652-020-02761-x
- Albuquerque, G. A., Carvalho, D. D., Cruz, A. S., Santos, J. P., Machado, G. M., Gendriz, I. S., et al. (2023). Osteoporosis screening using machine learning and electromagnetic waves. *Sci. Rep.* 13, 12865. doi:10.1038/s41598-023-40104-w
- Al-Hindawi, F., Siddiquee, M. M. R., Wu, T., Hu, H., and Sun, Y. (2024). Domain-knowledge inspired pseudo supervision (dips) for unsupervised image-to-image translation models to support cross-domain classification. *Eng. Appl. Artif. Intell.* 127, 107255. doi:10.1016/j.engappai.2023.107255
- Al-Hindawi, F., Soori, T., Hu, H., Siddiquee, M. M. R., Yoon, H., Wu, T., et al. (2023). A framework for generalizing critical heat flux detection models using unsupervised image-to-image translation. *Expert Syst. Appl.* 227, 120265. doi:10.1016/j.eswa.2023.120265
- Alhnaity, B., Kollias, S., Leontidis, G., Jiang, S., Schamp, B., and Pearson, S. (2021). An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth. *Inf. Sci.* 560, 35–50. doi:10.1016/j.ins.2021.01.037
- Aljuaid, T., and Sasi, S. (2016). “Proper imputation techniques for missing values in data sets,” in *2016 international conference on data science and engineering (ICDSE)*, 1–5. doi:10.1109/ICDSE.2016.7823957
- Allam, A., Feuerriegel, S., Rebhan, M., and Krauthammer, M. (2021). Analyzing patient trajectories with artificial intelligence. *J. Med. internet Res.* 23, e29812. doi:10.2196/29812
- Al-Qahtani, F. H., and Crone, S. F. (2013). “Multivariate k-nearest neighbour regression for time series data—a novel algorithm for forecasting UK electricity demand,” in *The 2013 international joint conference on neural networks (IJCNN)* (IEEE), 1–8. doi:10.1109/IJCNN.2013.6706742
- Altarazi, S., Allaf, R., and Alhindawi, F. (2019). Machine learning models for predicting and classifying the tensile strength of polymeric films fabricated via different production processes. *Materials* 12, 1475. doi:10.3390/ma12091475
- Al Zahrani, S., Al Sameeh, F. A. R., Musa, A. C., and Shokeralla, A. A. (2020). Forecasting diabetes patients attendance at al-baha hospitals using autoregressive fractional integrated moving average (arfima) models. *J. Data Analysis Inf. Process.* 8, 183–194. doi:10.4236/jdaip.2020.83011
- Bao, Y., Xiong, T., and Hu, Z. (2014). Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing* 129, 482–493. doi:10.1016/j.neucom.2013.09.010
- Barreto, T. d. O., Veras, N. V. R., Cardoso, P. H., Fernandes, F. R. d. S., Medeiros, L. P. d. S., Bezerra, M. V., et al. (2023). Artificial intelligence applied to analyzes during the pandemic: covid-19 beds occupancy in the state of rio grande do norte, Brazil. *Front. Artif. Intell.* 6, 1290022. doi:10.3389/frai.2023.1290022
- Baum, L. E., and Eagon, J. A. (1967). *An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology.*
- Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of markov processes. *Inequalities* 3, 1–8.
- Baum, L. E., and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. statistics* 37, 1554–1563. doi:10.1214/aoms/1177699147
- Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. statistics* 41, 164–171. doi:10.1214/aoms/1177697196
- Baum, L. E., and Sell, G. (1968). Growth transformations for functions on manifolds. *Pac. J. Math.* 27, 211–227. doi:10.2140/pjm.1968.27.211
- Bayyurt, L., and Bayyurt, B. (2020). *Forecasting of covid-19 cases and deaths using arima models*, 2020. medrxiv. doi:10.1101/2020.04.17.20069237
- Bonilla, E. V., Chai, K., and Williams, C. (2007). Multi-task Gaussian process prediction. *Adv. neural Inf. Process. Syst.* 20.
- Bou-Hamad, I., and Jamali, I. (2020). Forecasting financial time-series using data mining models: a simulation study. *Res. Int. Bus. Finance* 51, 101072. doi:10.1016/j.ribaf.2019.101072
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/A:1010933404324
- Cai, F., Patharkar, A., Wu, T., Lure, F. Y., Chen, H., and Chen, V. C. (2023). Stride: systematic radar intelligence analysis for adrd risk evaluation with gait signature simulation and deep learning. *IEEE sensors J.* 23, 10998–11006. doi:10.1109/jksen.2023.3263071
- Cai, H., Jia, X., Feng, J., Li, W., Hsu, Y.-M., and Lee, J. (2020). Gaussian process regression for numerical wind speed prediction enhancement. *Renew. energy* 146, 2112–2123. doi:10.1016/j.renene.2019.08.018
- Cao, L., Fang, H., and Liu, X. (2015). “Multi-step ahead forecasting for fault prognosis using hidden markov model,” in *The 27th Chinese control and decision conference (2015 CCDC)* (IEEE), 1688–1692. doi:10.1109/CCDC.2015.7162191
- Chandra, R., Goyal, S., and Gupta, R. (2021). Evaluation of deep learning models for multi-step ahead time series prediction. *IEEE Access* 9, 83105–83123. doi:10.1109/ACCESS.2021.3085085
- Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Sci. Rep.* 8, 6085. doi:10.1038/s41598-018-24271-9
- Cheng, L.-F., Dumitrascu, B., Darnell, G., Chivers, C., Draugelis, M., Li, K., et al. (2020). Sparse multi-output Gaussian processes for online medical time series prediction. *BMC Med. Inf. Decis. Mak.* 20, 1–23. doi:10.1186/s12911-020-1069-4
- Chuah, M. C., and Fu, F. (2007). “Ecg anomaly detection via time series analysis,” in *Frontiers of high performance computing and networking ISPA 2007 workshops: ISPA 2007 international workshops SSDSN, UPWN, WISH, SGC, ParDMCom, HiPCoMB, and IST-AWSN niagara falls, Canada, august 28-september 1, 2007 proceedings 5* (Springer), 123–135.
- Daydulo, Y. D., Thamineni, B. L., and Dawud, A. A. (2023). Cardiac arrhythmia detection using deep learning approach and time frequency representation of ecg signals. *BMC Med. Inf. Decis. Mak.* 23, 232. doi:10.1186/s12911-023-02326-w
- De Gooijer, J. G., and Hyndman, R. J. (2006). 25 years of time series forecasting. *Int. J. Forecast.* 22, 443–473. doi:10.1016/j.ijforecast.2006.01.001
- de Wolff, T., Cuevas, A., and Tobar, F. (2021). Mogptk: the multi-output Gaussian process toolkit. *Neurocomputing* 424, 49–53. doi:10.1016/j.neucom.2020.09.085
- Ding, G., Li, X., Jiao, F., and Shen, Y. (2020). *Brief analysis of the arima model on the covid-19 in Italy*, 2020. medRxiv. doi:10.1101/2020.04.08.20058636
- Doretto, G., Chiuso, A., Wu, Y. N., and Soatto, S. (2003). Dynamic textures. *Int. J. Comput. Vis.* 51, 91–109. doi:10.1023/A:1021669406132

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2024.1386760/full#supplementary-material>

- Du, S., Li, T., Yang, Y., and Horng, S.-J. (2020). Multivariate time series forecasting via attention-based encoder-decoder framework. *Neurocomputing* 388, 269–279. doi:10.1016/j.neucom.2019.12.118
- Dürichen, R., Pimentel, M. A., Clifton, L., Schweikard, A., and Clifton, D. A. (2014). “Multi-task Gaussian process models for biomedical applications,” in *IEEE-EMBS international Conference on Biomedical and health informatics (BHI) (IEEE)*, 492–495. doi:10.1109/BHI.2014.6864410
- Ehsani, R., and Drablos, F. (2020). Robust distance measures for k nn classification of cancer data. *Cancer Inf.* 19, 1176935120965542. doi:10.1177/1176935120965542
- El Mrabet, Z., Sugunara, N., Ranganathan, P., and Abhyankar, S. (2022). Random forest regressor-based approach for detecting fault location and duration in power systems. *Sensors* 22, 458. doi:10.3390/s22020458
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res. Oceans* 99, 10143–10162. doi:10.1029/94jc00572
- Evgeniou, T., Micchelli, C. A., Pontil, M., and Shawe-Taylor, J. (2005). Learning multiple tasks with kernel methods. *J. Mach. Learn. Res.* 6.
- Fan, G.-F., Peng, L.-L., Zhao, X., and Hong, W.-C. (2017). Applications of hybrid emd with pso and ga for an svr-based load forecasting model. *Energies* 10, 1713. doi:10.3390/en10111713
- Fang, T., and Lahdelma, R. (2016). Evaluation of a multiple linear regression model and sarima model in forecasting heat demand for district heating system. *Appl. energy* 179, 544–552. doi:10.1016/j.apenergy.2016.06.133
- Filipow, N., Main, E., Tanriver, G., Raywood, E., Davies, G., Douglas, H., et al. (2023). Exploring flexible polynomial regression as a method to align routine clinical outcomes with daily data capture through remote technologies. *BMC Med. Res. Methodol.* 23, 114. doi:10.1186/s12874-023-01942-4
- Fix, E., and Hodges, J. L. (1989). Discriminatory analysis. nonparametric discrimination: consistency properties. *Int. Stat. Review/Revue Int. Stat.* 57, 238–247. doi:10.2307/1403797
- Galton, F. (1886). Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. G. B. Irel.* 15, 246–263. doi:10.2307/2841583
- Garcia, V., Debreuve, E., Nielsen, F., and Barlaud, M. (2010). “K-nearest neighbor search: fast gpu-based implementations and application to high-dimensional feature matching,” in *2010 IEEE international Conference on image processing (IEEE)*, 3757–3760.
- Gauss, C.-F. (1823). *Theoria combinationum observationum erroribus minimis obnoxiae (Henricus Dieterich)*.
- Gavrishchaka, V. V., and Banerjee, S. (2006). Support vector machine as an efficient framework for stock market volatility forecasting. *Comput. Manag. Sci.* 3, 147–160. doi:10.1007/s10287-005-0005-5
- Ghaderi, H., Foreman, B., Nayebi, A., Tipirneni, S., Reddy, C. K., and Subbian, V. (2023). A self-supervised learning-based approach to clustering multivariate time-series data with missing values (slac-time): an application to tbi phenotyping. *J. Biomed. Inf.* 143, 104401. doi:10.1016/j.jbi.2023.104401
- Ghahramani, Z., and Hinton, G. E. (1996). *Parameter estimation for linear dynamical systems*.
- Godfrey, L. B., and Gashler, M. S. (2017). Neural decomposition of time-series data for effective generalization. *IEEE Trans. neural Netw. Learn. Syst.* 29, 2973–2985. doi:10.1109/TNNLS.2017.2709324
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R., et al. (2000). *Physiobank, physiotoolkit, and physiome: components of a new research resource for complex physiologic signals*.
- Gopakumar, S., Tran, T., Luo, W., Phung, D., and Venkatesh, S. (2016). Forecasting daily patient outflow from a ward having no real-time clinical data. *JMIR Med. Inf.* 4, e25. doi:10.2196/medinform.5650
- Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc. F (radar signal processing) (IET)* 140, 107–113. doi:10.1049/ip-f-2.1993.0015
- Hamdi, T., Ali, J. B., Di Costanzo, V., Fnaiech, F., Moreau, E., and Ginoux, J.-M. (2018). Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. *Biocybern. Biomed. Eng.* 38, 362–372. doi:10.1016/j.bbe.2018.02.005
- Harerimana, G., Kim, J. W., and Jang, B. (2022). A multi-headed transformer approach for predicting the patient’s clinical time-series variables from charted vital signs. *IEEE Access* 10, 105993–106004. doi:10.1109/access.2022.3211334
- Haywood, J., and Wilson, G. T. (2009). A test for improved multi-step forecasting. *J. Time Ser. Analysis* 30, 682–707. doi:10.1111/j.1467-9892.2009.00634.x
- Helmini, S., Jihan, N., Jayasinghe, M., and Perera, S. (2019). Sales forecasting using multivariate long short term memory network models. *PeerJ Prepr.* 7, e27712v1. doi:10.7287/peerj.preprints.27712v1
- Hernandez-Lorenzo, L., Ilundain, I. S., and Rodrigo, J. L. A. (2022). “Timeseries biomarkers clustering for alzheimer’s disease progression,” in *2022 IEEE international conference on omni-layer intelligent systems (COINS) (IEEE)*, 1–7.
- Hewamalage, H., Bergmeir, C., and Bandara, K. (2021). Recurrent neural networks for time series forecasting: current status and future directions. *Int. J. Forecast.* 37, 388–427. doi:10.1016/j.ijforecast.2020.06.008
- Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi:10.1162/neco.1997.9.8.1735
- Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* 20, 5–10. doi:10.1016/j.ijforecast.2003.09.015
- Hosseinzadeh, P., Nassar, A., Boubrahimi, S. F., and Hamdi, S. M. (2023). ML-based streamflow prediction in the upper Colorado river basin using climate variables time series data. *Hydrology* 10, 29. doi:10.3390/hydrology10020029
- Huang, J., Ghalamsiah, N., Patharkar, A., Pradhan, O., Chu, M., Wu, T., et al. (2024). An entropy-based causality framework for cross-level faults diagnosis and isolation in building hvac systems. *Energy Build.* 317, 114378. doi:10.1016/j.enbuild.2024.114378
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., and Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Comput. statistics and data analysis* 55, 2579–2589. doi:10.1016/j.csda.2011.03.006
- Idriss, T. E., Idri, A., Abnane, I., and Bakkoury, Z. (2019). “Predicting blood glucose using an lsm neural network,” in *2019 federated conference on computer science and information systems (FedCSIS)*, 35–41. doi:10.15439/2019F159
- Janacek, G. (2010). Time series analysis forecasting and control. *J. Time Ser. Analysis* 31, 303. doi:10.1111/j.1467-9892.2009.00643.x
- Jeon, J., Panagiotelis, A., and Petropoulos, F. (2019). Probabilistic forecast reconciliation with applications to wind power and electric load. *Eur. J. Operational Res.* 279, 364–379. doi:10.1016/j.ejor.2019.05.020
- Jones, P. W., Osipov, A., and Rokhlin, V. (2011). “Randomized approximate nearest neighbors algorithm,” in *Proceedings of the national academy of sciences* 108, 15679–15686.
- Julier, S. J., and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. *Signal Process. Sens. fusion, target Recognit. VI (Spie)* 3068, 182–193. doi:10.1117/12.280797
- Jun, E., Mulyadi, A. W., and Suk, H.-I. (2019). “Stochastic imputation and uncertainty-aware attention to ehr for mortality prediction,” in *2019 international joint conference on neural networks (IJCNN)*, 1–7. doi:10.1109/IJCNN.2019.8852132
- Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *J. Soc. Industrial Appl. Math. Ser. A Control* 1, 152–192. doi:10.1137/0301010
- Kandula, S., Yamana, T., Pei, S., Yang, W., Morita, H., and Shaman, J. (2018). Evaluation of mechanistic and statistical methods in forecasting influenza-like illness. *J. R. Soc. Interface* 15, 20180174. doi:10.1098/rsif.2018.0174
- Kantz, H., and Schreiber, T. (2004). *Nonlinear time series analysis*, 7. Cambridge University Press. doi:10.1017/cbo9780511755798
- Katari, S., Likith, T., Sree, M. P. S., and Rachapudi, V. (2023). “Heart disease prediction using hybrid ml algorithms,” in *2023 international conference on sustainable computing and data communication systems (ICSCDS) (IEEE)*, 121–125.
- Katayama, T. (2005). *Subspace methods for system identification*. Springer. doi:10.1007/1-84628-158-X
- Khalique, F., Khan, S. A., Butt, W. H., and Matloob, I. (2020). An integrated approach for spatio-temporal cholera disease hotspot relation mining for public health management in Punjab, Pakistan. *Int. J. Environ. Res. Public Health* 17, 3763. doi:10.3390/ijerph17113763
- Lai, G., Chang, W.-C., Yang, Y., and Liu, H. (2018). “Modeling long-and short-term temporal patterns with deep neural networks,” in *The 41st international ACM SIGIR conference on research and development in information retrieval*, 95–104. doi:10.1145/3209978.3210006
- Lee, J. M., and Hauskrecht, M. (2021). Modeling multivariate clinical event time-series with recurrent temporal mechanisms. *Artif. Intell. Med.* 112, 102021. doi:10.1016/j.artmed.2021.102021
- Legendre, A. M. (1806). *Nouvelles méthodes pour la détermination des orbites des comètes: avec un supplément contenant divers perfectionnements de ces méthodes et leur application aux deux comètes de 1805 (Courcier)*.
- Lewis, F. L. (1986). *Optimal estimation: with an introduction to stochastic control theory. (No Title)*.
- Li, H., and Hyndman, R. J. (2021). Assessing mortality inequality in the us: what can be said about the future? *Insur. Math. Econ.* 99, 152–162. doi:10.1016/j.insmath.2021.03.014
- Li, L., Li, H., and Panagiotelis, A. (2024). Boosting domain-specific models with shrinkage: an application in mortality forecasting. *Int. J. Forecast.* doi:10.1016/j.ijforecast.2024.05.001
- Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y.-X., et al. (2019). Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Adv. neural Inf. Process. Syst.* 32.
- Liu, J., Yu, F., and Song, H. (2023). Application of sarima model in forecasting and analyzing inpatient cases of acute mountain sickness. *BMC Public Health* 23, 56. doi:10.1186/s12889-023-14994-4

- Liu, Z. (2016). *Time series modeling of irregularly sampled multivariate clinical data*. University of Pittsburgh. Ph.D. thesis.
- Liu, Z., and Hauskrecht, M. (2015). Clinical time series prediction: toward a hierarchical dynamical system framework. *Artif. Intell. Med.* 65, 5–18. doi:10.1016/j.artmed.2014.10.005
- Liu, Z., Wu, L., and Hauskrecht, M. (2013). “Modeling clinical time series using Gaussian process sequences,” in *Proceedings of the 2013 SIAM international conference on data mining (SIAM)*, 623–631.
- Liu, Z., Zhu, Z., Gao, J., and Xu, C. (2021). Forecast methods for time series data: a survey. *Ieee Access* 9, 91896–91912. doi:10.1109/ACCESS.2021.3091162
- Luo, Y., Cai, X., Zhang, Y., Xu, J., and Yuan, X. (2018). Multivariate time series imputation with generative adversarial networks. *Adv. neural Inf. Process. Syst.* 31.
- Mahmudimanesh, M., Mirzaee, M., Dehghan, A., and Bahrapour, A. (2022). Forecasts of cardiac and respiratory mortality in tehran, Iran, using arimax and cnn-lstm models. *Environ. Sci. Pollut. Res.* 29, 28469–28479. doi:10.1007/s11356-021-18205-8
- Manaris, B., Hughes, D., and Vassilandonakis, Y. (2011). “Monterey mirror: combining markov models, genetic algorithms, and power laws,” in *Computer science department, college of charleston, SC, USA, appeared in proceedings of 1st workshop in evolutionary music, 2011 IEEE congress on evolutionary computation (CEC 2011)* (New Orleans, LA, USA (Citeseer)), 33–40.
- Martínez, F., Frías, M. P., Pérez, M. D., and Rivera, A. J. (2019). A methodology for applying k-nearest neighbor to time series forecasting. *Artif. Intell. Rev.* 52, 2019–2037. doi:10.1007/s10462-017-9593-z
- Mitra, A., and Ashraf, K. (2018). Sepsis prediction and vital signs ranking in intensive care unit patients. *arXiv Prepr. arXiv:1812.06686*. doi:10.48550/arXiv.1812.06686
- Mollura, M., Lehman, L.-W. H., Mark, R. G., and Barbieri, R. (2021). A novel artificial intelligence based intensive care unit monitoring system: using physiological waveforms to identify sepsis. *Philosophical Trans. R. Soc. A* 379, 20200252. doi:10.1098/rsta.2020.0252
- Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley and Sons.
- Moon, J., Kim, Y., Son, M., and Hwang, E. (2018). Hybrid short-term load forecasting scheme using random forest and multilayer perceptron. *Energies* 11, 3283. doi:10.3390/en11123283
- Mulyadi, A. W., Jun, E., and Suk, H.-I. (2022). Uncertainty-aware variational-recurrent imputation network for clinical time series. *IEEE Trans. Cybern.* 52, 9684–9694. doi:10.1109/TCYB.2021.3053599
- Naemi, A., Mansourvar, M., Schmidt, T., and Wiil, U. K. (2020). “Prediction of patients severity at emergency department using narx and ensemble learning,” in *2020 IEEE international Conference on Bioinformatics and biomedicine (BIBM) (IEEE)*, 2793–2799. doi:10.1109/BIBM49941.2020.9313462
- Niu, K., Lu, Y., Peng, X., and Zeng, J. (2022). Fusion of sequential visits and medical ontology for mortality prediction. *J. Biomed. Inf.* 127, 104012. doi:10.1016/j.jbi.2022.104012
- [Dataset] Overchee, P., and Moor, B. (1996). *Subspace identification for linear system*.
- Patharkar, A., Huang, J., Wu, T., Forzani, E., Thomas, L., Lind, M., et al. (2024). Eigenentropy based time series signatures to support multivariate time series classification. *Sci. Rep.* 14, 16076. doi:10.1038/s41598-024-66953-7
- Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin philosophical Mag. J. Sci.* 2, 559–572. doi:10.1080/14786440109462720
- Pearson, K. (1922). Francis galton 1822-1922: a centenary appreciation, 1–28. doi:10.1037/11161-001
- Pearson, K. (2023). “The life, letters and labours of francis galton,” in *Scientific and medical knowledge production, 1796-1918* (London, United Kingdom: Routledge), 311–318.
- Pillonetto, G., Dinuzzo, F., and De Nicolao, G. (2008). Bayesian online multitask learning of Gaussian processes. *IEEE Trans. Pattern Analysis Mach. Intell.* 32, 193–205. doi:10.1109/TPAMI.2008.297
- Pinto, R., Valentim, R., da Silva, L. F., de Souza, G. F., de Moura Santos, T. G. F., de Oliveira, C. A. P., et al. (2022). Use of interrupted time series analysis in understanding the course of the congenital syphilis epidemic in Brazil. *Lancet Regional Health—Americas* 7, 100163. doi:10.1016/j.lana.2021.100163
- Plis, K., Bunescu, R., Marling, C., Shubrook, J., and Schwartz, F. (2014). “A machine learning approach to predicting blood glucose levels for diabetes management,” in *Workshops at the Twenty-Eighth AAAI conference on artificial intelligence*.
- Poloni, F., and Sbrana, G. (2015). A note on forecasting demand using the multivariate exponential smoothing framework. *Int. J. Prod. Econ.* 162, 143–150. doi:10.1016/j.ijpe.2015.01.017
- Qi, C., Zhang, D., Zhu, Y., Liu, L., Li, C., Wang, Z., et al. (2020). Sarfima model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with sarima. *BMC Med. Res. Methodol.* 20, 243–247. doi:10.1186/s12874-020-01130-8
- Quinero-Candela, J., and Rasmussen, C. E. (2005). A unifying view of sparse approximate Gaussian process regression. *J. Mach. Learn. Res.* 6, 1939–1959.
- Rabyk, L., and Schmid, W. (2016). Ewma control charts for detecting changes in the mean of a long-memory process. *Metrika* 79, 267–301. doi:10.1007/s00184-015-0555-7
- Rachmat, R., and Suhartono, S. (2020). Comparative analysis of single exponential smoothing and holt’s method for quality of hospital services forecasting in general hospital. *Bull. Comput. Sci. Electr. Eng.* 1, 80–86. doi:10.25008/bcsee.v1i2.8
- Reyna, M. A., Josef, C., Seyedi, S., Jeter, R., Shashikumar, S. P., Westover, M. B., et al. (2019). “Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019,” in *2019 computing in cardiology (cinc)* (IEEE).
- Reyna, M. A., Josef, C. S., Jeter, R., Shashikumar, S. P., Westover, M. B., Nemati, S., et al. (2020). Early prediction of sepsis from clinical data: the physionet/computing in cardiology challenge 2019. *Crit. Care Med.* 48, 210–217. doi:10.1097/CCM.0000000000004145
- Roberts, S. (2000). Control chart tests based on geometric moving averages. *Technometrics* 42, 97–101. doi:10.1080/00401706.2000.10485986
- Roberts, S., Osborne, M., Ebdon, M., Reece, S., Gibson, N., and Aigrain, S. (2013). Gaussian processes for time-series modelling. *Philosophical Trans. R. Soc. A Math. Phys. Eng. Sci.* 371, 20110550. doi:10.1098/rsta.2011.0550
- Rostami-Tabar, B., and Hyndman, R. J. (2024). Hierarchical time series forecasting in emergency medical services. *J. Serv. Res.*, 10946705241232169. doi:10.1177/10946705241232169
- Sagheer, A., Hamdoun, H., and Youness, H. (2021). Deep lstm-based transfer learning approach for coherent forecasts in hierarchical time series. *Sensors* 21, 4379. doi:10.3390/s21134379
- Said, A. B., Erradi, A., Aly, H. A., and Mohamed, A. (2021). Predicting covid-19 cases using bidirectional lstm on multivariate time series. *Environ. Sci. Pollut. Res.* 28, 56043–56052. doi:10.1007/s11356-021-14286-7
- Saif, N., Yan, P., Niotis, K., Scheyer, O., Rahman, A., Berkowitz, M., et al. (2020). Feasibility of using a wearable biosensor device in patients at risk for alzheimer’s disease dementia. *J. Prev. Alzheimer’s Dis.* 7, 104–111. doi:10.14283/jpad.2019.39
- Samae, S., and Kobravi, H. R. (2020). Predicting the occurrence of wrist tremor based on electromyography using a hidden markov model and entropy based learning algorithm. *Biomed. Signal Process. Control* 57, 101739. doi:10.1016/j.bspc.2019.101739
- Shamout, F., Zhu, T., and Clifton, D. A. (2021). Machine learning for clinical outcome prediction. *IEEE Rev. Biomed. Eng.* 14, 116–126. doi:10.1109/RBME.2020.3007816
- Shashikumar, S. P., Stanley, M. D., Sadiq, I., Li, Q., Holder, A., Clifford, G. D., et al. (2017). Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *J. Electrocardiol.* 50, 739–743. doi:10.1016/j.jelectrocard.2017.08.013
- Shih, S.-Y., Sun, F.-K., and Lee, H.-y. (2019). Temporal pattern attention for multivariate time series forecasting. *Mach. Learn.* 108, 1421–1441. doi:10.1007/s10994-019-05815-0
- Shukla, S. N. (2017). “Estimation of blood pressure from non-invasive data,” in *2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (IEEE)*, 1772–1775. doi:10.1109/EMBC.2017.8037187
- Siami-Namini, S., Tavakoli, N., and Namin, A. S. (2019). “The performance of lstm and bilstm in forecasting time series,” in *2019 IEEE International conference on big data (Big Data) (IEEE)*, 3285–3292.
- Sotoodeh, M., and Ho, J. C. (2019). Improving length of stay prediction using a hidden markov model. *AMIA Summits Transl. Sci. Proc.* 2019, 425–434.
- Stegle, O., Lippert, C., Mooij, J. M., Lawrence, N., and Borgwardt, K. (2011). Efficient inference in matrix-variate Gaussian models with iid observation noise. *Adv. neural Inf. Process. Syst.* 24.
- Suganya, R., Arunadevi, R., and Buhari, S. M. (2020). *Covid-19 forecasting using multivariate linear regression*.
- Sun, J., Xie, J., and Zhou, H. (2021). “Eeg classification with transformer-based models,” in *2021 IEEE 3rd global conference on life sciences and technologies (lifetech) (IEEE)*, 92–93. doi:10.1109/LifeTech52111.2021.9391844
- Tandon, H., Ranjan, P., Chakraborty, T., and Suhag, V. (2022). Coronavirus (covid-19): arima-based time-series analysis to forecast near future and the effect of school reopening in India. *J. Health Manag.* 24, 373–388. doi:10.1177/09720634221109087
- Tipirneni, S., and Reddy, C. K. (2022). Self-supervised transformer for sparse and irregularly sampled multivariate clinical time-series. *ACM Trans. Knowl. Discov. Data (TKDD)* 16, 1–17. doi:10.1145/3516367
- Tyralis, H., and Papacharalampous, G. (2017). Variable selection in time series forecasting using random forests. *Algorithms* 10, 114. doi:10.3390/a10040114
- Valentim, R. A., Caldeira-Silva, G. J., Da Silva, R. D., Albuquerque, G. A., De Andrade, I. G., Sales-Moioli, A. I. L., et al. (2022). Stochastic petri net model describing the relationship between reported maternal and congenital syphilis cases in Brazil. *BMC Med. Inf. Decis. Mak.* 22, 40. doi:10.1186/s12911-022-01773-1
- Vapnik, V. (1999). *The nature of statistical learning theory*. Springer science and business media.

- Vapnik, V., Golowich, S., and Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Adv. neural Inf. Process. Syst.* 9.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Adv. neural Inf. Process. Syst.* 30.
- Vellido, A. (2019). Societal issues concerning the application of artificial intelligence in medicine. *Kidney Dis.* 5, 11–17. doi:10.1159/000492428
- Veneri, G., Pretegianni, E., Rosini, F., Federighi, P., Federico, A., and Rufa, A. (2012). Evaluating the human ongoing visual search performance by eye tracking application and sequencing tests. *Comput. methods programs Biomed.* 107, 468–477. doi:10.1016/j.cmpb.2011.02.006
- Wang, Q., Molenaar, P., Harsh, S., Freeman, K., Xie, J., Gold, C., et al. (2014). Personalized state-space modeling of glucose dynamics for type 1 diabetes using continuously monitored glucose, insulin dose, and meal intake: an extended kalman filter approach. *J. diabetes Sci. Technol.* 8, 331–345. doi:10.1177/1932296814524080
- Wang, X., Hyndman, R. J., Li, F., and Kang, Y. (2023). Forecast combinations: an over 50-year review. *Int. J. Forecast.* 39, 1518–1547. doi:10.1016/j.ijforecast.2022.11.005
- Weerakody, P. B., Wong, K. W., Wang, G., and Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing* 441, 161–178. doi:10.1016/j.neucom.2021.02.046
- Wickramasuriya, S. L., Athanasopoulos, G., and Hyndman, R. J. (2018). Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *J. Am. Stat. Assoc.* 114, 804–819. doi:10.1080/01621459.2018.1448825
- Williams, C. K., and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT press.
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Manag. Sci.* 6, 324–342. doi:10.1287/mnsc.6.3.324
- Wold, H. O. (1948). On prediction in stationary time series. *Ann. Math. Statistics* 19, 558–567. doi:10.1214/aoms/1177730151
- Wu, J., Li, Z., and Yang, S. (2021). “Covid-19 dynamics prediction by improved multi-polynomial regression model,” in *The 2nd international conference on computing and data science*, 1–7. doi:10.1145/3448734.3450847
- Xiao, C., Choi, E., and Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *J. Am. Med. Inf. Assoc.* 25, 1419–1428. doi:10.1093/jamia/ocy068
- Xie, F., Yuan, H., Ning, Y., Ong, M. E. H., Feng, M., Hsu, W., et al. (2022). Deep learning for temporal data representation in electronic health records: a systematic review of challenges and methodologies. *J. Biomed. Inf.* 126, 103980. doi:10.1016/j.jbi.2021.103980
- Xu, X., Liu, X., Kang, Y., Xu, X., Wang, J., Sun, Y., et al. (2020). A multi-directional approach for missing value estimation in multivariate time series clinical data. *J. Healthc. Inf. Res.* 4, 365–382. doi:10.1007/s41666-020-00076-2
- Yoon, B.-J., and Vaidyanathan, P. (2006). Context-sensitive hidden markov models for modeling long-range dependencies in symbol sequences. *IEEE Trans. Signal Process.* 54, 4169–4184. doi:10.1109/TSP.2006.880252
- Yule, G. U. (1927). Vii. on a method of investigating periodicities disturbed series, with special reference to wolfer’s sunspot numbers. *Philosophical Trans. R. Soc. Lond. Ser. A, Contain. Pap. a Math. or Phys. Character* 226, 267–298. doi:10.1098/rsta.1927.0007
- Zeng, A., Chen, M., Zhang, L., and Xu, Q. (2023). Are transformers effective for time series forecasting? *Proc. AAAI Conf. Artif. Intell.* 37, 11121–11128. doi:10.1609/aaai.v37i9.26317
- Zhang, B., Ren, H., Huang, G., Cheng, Y., and Hu, C. (2019). Predicting blood pressure from physiological index data using the svr algorithm. *BMC Bioinforma.* 20, 109–115. doi:10.1186/s12859-019-2667-y
- Zhang, H., Zou, Y., Yang, X., and Yang, H. (2022). A temporal fusion transformer for short-term freeway traffic speed multistep prediction. *Neurocomputing* 500, 329–340. doi:10.1016/j.neucom.2022.05.083
- Zhang, M., Flores, K. B., and Tran, H. T. (2021). Deep learning and regression approaches to forecasting blood glucose levels for type 1 diabetes. *Biomed. Signal Process. Control* 69, 102923. doi:10.1016/j.bspc.2021.102923
- Zhao, J., Gu, S., and McDermaid, A. (2019). Predicting outcomes of chronic kidney disease from emr data based on random forest regression. *Math. Biosci.* 310, 24–30. doi:10.1016/j.mbs.2019.02.001
- Zhou, B., Liu, S., Hooi, B., Cheng, X., and Ye, J. (2019). Beatgan: anomalous rhythm detection using adversarially generated time series. *IJCAI* 2019, 4433–4439. doi:10.24963/ijcai.2019/616
- Zhu, T., Li, K., Chen, J., Herrero, P., and Georgiou, P. (2020). Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *J. Healthc. Inf. Res.* 4, 308–324. doi:10.1007/s41666-020-00068-2
- Zou, Y., Donner, R. V., Marwan, N., Donges, J. F., and Kurths, J. (2019). Complex network approaches to nonlinear time series analysis. *Phys. Rep.* 787, 1–97. doi:10.1016/j.physrep.2018.10.005