Healthcare Time Series Forecasting Using Gaussian Processes and Dynamic Time Warping-Based Data Selection

[1] Sahadeva Reddy Bollu
Assistant Professor
Dept. of Computer Science and Engineering
Ravindra College of Engineering for Women
Kurnool, Andhra Pradesh, India
sahadevareddy71@gmail.com

[2] M. Balakrishna
Assistant Professor
Dept. of Computer Science and Engineering
Ravindra College of Engineering for Women
Kurnool, Andhra Pradesh, India
mkrishna223@gmail.com

***Abstract*—The structure of healthcare time series data (vital signs, disease progression, and physiological sensor measurements) has high levels of complexity, leading to difficult forecasting tasks. There are many challenges facing traditional forecasting methodologies due to their inability to capture variability and non-linear relationships present in the healthcare datasets. This paper proposes a healthcare time series forecasting framework using a hybrid method of Gaussian Processes (GP) and Dynamic Time Warping (DTW) for data retrieval to enhance accuracy as well as efficiency in forecasting. In doing so, DTW is used to identify and select historical sequences that show similar temporal patterns as the target time series, allowing irrelevant records to be eliminated and improve the training of the models. Once a subset of historical sequences has been identified and selected via DTW, Gaussian Process regression is then applied to the remaining sequences to model temporal dependence and provide probabilistic forecasts, including estimates of forecast uncertainty. Performance evaluations of the proposed methodology in comparison to traditional machine learning based methods. and statistical methods using healthcare-related temporal datasets indicate the proposed method demonstrates significantly improved forecasting performance. Improvements include: improved forecasting accuracy, improved robustness to noisy observations, and reduced computational overhead. The combination of DTW-based subset selection and Gaussian Process modeling offers a viable solution to healthcare forecasting applications to enable more dependable clinical decision support and proactive monitoring of patients.**

***Index Terms*—Healthcare Analytics, Time Series Forecasting, Gaussian Processes, Dynamic Time Warping, Machine Learning, Predictive Healthcare, Clinical Decision Support.**

I. Introduction

The enormous increase in digital health care systems and their subsequent increase in the amount of time-dependent medical records created via electronic health records, wearables [1], and patient monitoring systems, and sensors has produced a tremendous amount of temporal, time-based medical data. Health care time series data contains valuable insight into patient condition, disease progression and response to treatment, and physiological trends. Timely and accurate forecasting of the temporal data provides critical support for clinical decision making; early disease detection; responsible allocation of resources; and personalized health care management.There are a variety of challenges associated with healthcare time series forecasting due to the complexity of medical data [2]. There are nonlinear patterns, irregular sampling intervals, missing observations, and high levels of variability between patients in both physiological signals and patient health records. Traditional statistical forecasting methods like ARIMA models can have difficulty capturing lower accuracy of predictions. As such, there is a growing need for more complex machine learning techniques capable of modeling complex temporal dependencies and providing accurate estimates of uncertainty [3].

Gaussian Processes (GPs) are a powerful and popular probabilistic machine learning model for predicting time series data. Unlike traditional regression models, GPs are able to model nonlinear relationships between variables and estimate the uncertainty associated with predicted values, both of which are important in healthcare where assessing risk is a common part of making decisions [3]. However, because the computation of a GP model requires , GPs are often not practical when working with large amounts of healthcare data.Similarity-based data selection methods can help to identify the most helpful historical data points to be used in the construction of a GP model [4]. One common technique for determining how similar two sequences are is known as Dynamic Time Warping (DTW); DTW allows for the alignment (re-sequencing) of two time series, even if they are of different lengths and have been distorted in time [9]. By identifying previous patterns that are most similar to a current (target) pattern, DTW is able to help find subsets of previously-collected data that will provide useful information while filtering out irrelevant or noisy data. This results in a computationally less intensive method of creating a GP model, as well as creating a more accurate model [5].

This research proposes a framework for the time series forecasting of healthcare data using a hybridized approach of Dynamic Time Warping (DTW)-based subsetting and Gaussian Process (GP) regression [9], [10]. The DTW algorithm will be used to find subsets of historical data with similar temporal patterns, and these subsets will then be used to fit a GP model that can predict future values [6]. The goal of this work is to increase the accuracy of forecasting while still being able to do so efficiently in terms of time and resources and provide estimates of confidence in the predictions made [6].

This paper makes the following major contributions:

1) A new framework for forecast construction in healthcare time series data that integrates Dynamic Time Warping-based sampling and selection of past events using Gaussian Process regression models [7].
2) A more effective means of identifying the most relevant historical data to be used as input into Gaussian Process regression models, thereby reducing the amount of computation required by these models [7].
3) Improved accuracy in forecasts by employing pattern matching based on similarity and incorporating probabilistic techniques in the learning process [6], [7].
4) An extensive experimental analysis of the proposed methodology is performed on multiple healthcare time series datasets, employing standard performance metrics.
5) The applicability of this framework for making clinical decisions, monitoring patients, and performing predictive analytics on healthcare data is demonstrated [8].

Sections II to VI of this paper detail several issues associated with advanced forecasting methods for time series data. The last section, Section VI, contains the summary of the entire study and outlines its proposed future directions and applications [8].

II. RELATED WORK

There has been much interest in time series forecasting in health care because it can support clinical decisions, track disease progression, and predict patient risk [10]. There are many statistical and machine learning methods available that can be applied to model temporal health data [10]. Classic methods for forecasting include Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing Techniques, both of which have been used to analyze the time series of medical records. While both of these classic time series forecasting methods have proven useful in modeling time series that are linear and stationary, these methods can struggle to find the nonlinear and complex relationships typically found in health data [11]. Several new machine learning methods have been developed to forecast health care [11]. For example, have received increased interest. Generally speaking, the aforementioned deep learning approaches require a large amount of training data and do not typically have high interpretability or provide good estimates of uncertainty [12].

Because a Gaussian Process (GP) can effectively model non-linear relationships while providing a prediction confidence interval, GP regression has become one of the more appealing choices for forecasting time series data in healthcare (due to the non-parametric nature of GP models, as well as the probabilistic framework of GP regression) [13]. A wide range of studies have demonstrated the ability of Gaussian Processes to predict physiological signals, monitor patients, and model disease progression [13]. However, the computational burden associated with GP modelling increases with the amount of training data available, thus making it more difficult to apply GP modelling to large-scale datasets in healthcare.To overcome this challenge of sample selection, researchers have used different approaches to develop similarity-based methodologies [14]. A simulation of one such similarity-based method is Dynamic Time Warping (DTW), which is commonly used in the measurement of similarity between two temporal sequences that differ both in length and tambal of time [15], [16]. There are many applications of DTW to health-related areas, such as ECG analysis, trajectory comparisons, activity recognition, and modeling of disease progression [15]. By identifying areas of alignment between two sequences, the DTW algorithm can provide relevant information from historical data, thereby improving prediction capabilities [16].

A graphical representation of the proposed method for forecasting is shown in Figure 1. The data used in the forecasting is composed of many sets of historical healthcare time series (and also the target series, which has some observations missing) [17]. To help obtain a temporal similarity between the historical sequences and the target sequence, the technique Dynamic Time Warping (DTW) is used to align them [17]. Once this has been completed, only the most similar training sequences remain which allows for the exclusion of irrelevant patterns in order to yield better computational efficiency. Using only the selected training sequences then, the target values will be predicted using a Gaussian Process (GP) regression model [18]. The GP does not only predict future observations of the target series, it also estimates the uncertainty associated with the predicted values [17], [19]. This is very helpful as decision-support systems can operate with multiple sources of uncertainty. Therefore, utilizing DTW and a probabilistic GP for forecasting will help improve not only the accuracy but also the reliability of the forecasts for heterogeneous healthcare time series [19].

III. Notation

Let's take into account a healthcare dataset that has multiple subjects and records temporal observations relating to those subjects. We can say N is the overall number of subjects in the dataset. For each of these subjects there are healthcare measurements that are taken on a continuous basis over time [20].

The input feature matrix for the i^{th} subject is represented as

$$X_i = \{x_i(t_1), x_i(t_2), \ldots, x_i(t_{m_i})\} \tag{1}$$

where $x_i(t_j)$ denotes the observed feature vector at time t_j; and m_i is the full potential of observations that can be recorded for subject index i.

Similarly, the response time-series corresponding would be expressed as

$$Y_i = \{y_i(t_1), y_i(t_2), \ldots, y_i(t_{m_i})\} \tag{2}$$

The target healthcare outcome $y_i(t_j)$ has been measured at the observed time t_j.

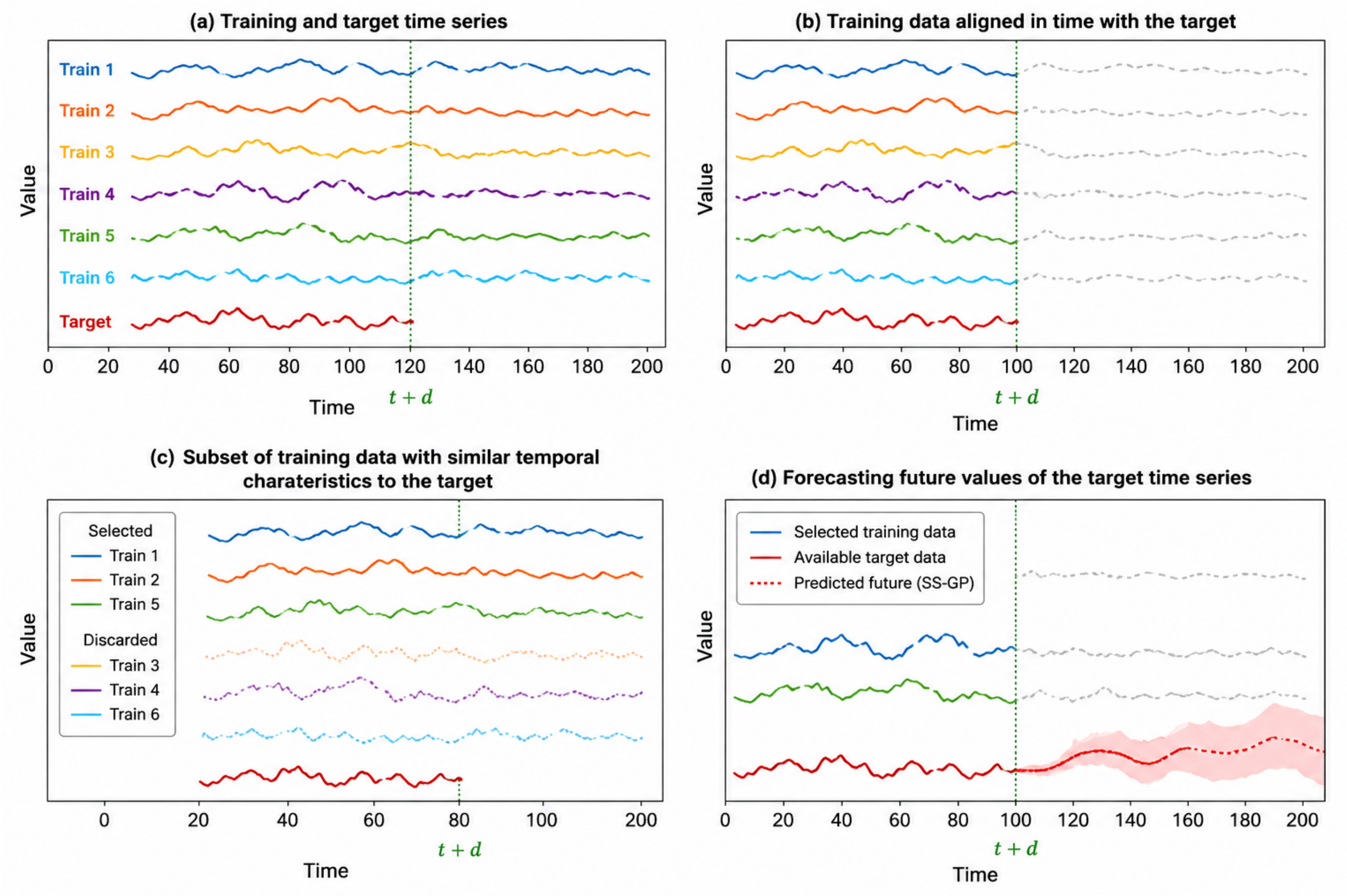

Fig. 1. Illustration of the proposed SS-GP forecasting framework. (a) Training and target time series, where target observations are available only up to time $t+d$. (b) Temporal alignment of the training sequences with the target sequence. (c) Selection of training sequences exhibiting temporal characteristics similar to those of the target sequence, while dissimilar sequences are discarded. (d) Forecasting future target values using the selected training sequences together with the available historical target observations. The predicted future trajectory is represented by the red dashed line.

The times at which the measurement of healthcare is taken are often taken on irregular or unevenly spaced intervals.

$$t_1 < t_2 < \cdots < t_{m_i} \tag{3}$$

Between multiple subjects, there are different time intervals from one observation to the next.

Let (X_T, Y_T) represent the target subject of interest for which future estimations are going to be computed. Please note that estimates can only be computed for the target subject based on observable information collected through time t_d.

$$Y_T = \{y_T(t_1), y_T(t_2), \ldots, y_T(t_d)\} \tag{4}$$

The goal here is to predict future values past the observation horizon at time t_d.

Dynamic time warping (DTW) is used in this approach to identify relevant historical patterns by measuring the distance (similarity) between the target and training/reference sequences within the context of DTW. The DTW distance between target and subject i is given by [21]:

$$D_i = DTW(Y_i, Y_T) \tag{5}$$

Based on these similarity scores, a subset of the most relevant subjects is selected as

$$S = \{Y_i \mid D_i < \delta\} \tag{6}$$

where δ denotes the DTW similarity threshold.

The selected subset S is then used to train a Gaussian Process (GP) forecasting model. The GP prior is defined as

$$f(t) \sim GP(\mu(t), k(t, t')) \tag{7}$$

where $\mu(t)$ represents the mean function and $k(t, t')$ denotes the covariance kernel.

The future healthcare outcome at prediction horizon h is estimated as

$$\hat{y}(t_d + h) = f(X_T, S) + \epsilon \tag{8}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise.

IV. Time Series Forecasting

A. Time Series Forecasting Methods

Many forecasting methods have been suggested for modelling temporal healthcare data, and this section provides a summary of most of the methods used in this study [22].

Maximum Likelihood Estimation (MLE): The maximum likelihood method is used to find the model parameters that will produce the highest possible likelihood given the observed data. observed target variable (Y). A pth degree polynomial forecasting model can be represented as follows [20]:

$$y(t) = \sum_{i=0}^{p} \beta_i t^i \tag{9}$$

where β_i The optimal polynomial coefficient vector can be determined through the process of estimating and predictions based on the observed measurements of the target individual through maximization of the probability associated with those measurements. Since estimates only rely upon the observation of the specific measurements of an individual, the final model is highly individualised. However, when limited or few observations have been utilised to build the model, the overall prediction accuracy of the model will typically decrease.

Maximum A Posteriori (MAP): Maximum a Posteriori (MAP) estimation builds on Maximum Likelihood Estimation (MLE) by including additional information from previously measured subjects [**?**], [22]. While MLE relies completely on the target data, MAP combines the likelihood of the data that has been collected with a prior distribution on the parameters. By using Bayes' theorem, the benefit of making use of both the sample data and the prior knowledge accumulated at the population level can result in better accuracy when there is little sample data to use for estimating parameter values [23].

Autoregressive Integrated Moving Average (ARIMA): Traditional statistical forecasting method that uses autoregressive and moving average components to identify temporal dependence. The simple representation for this model is given as ARIMA (p,d,q) where p = autoregressive order, d = differencing order and q = moving average order. ARIMA works well for time series data collected at regular intervals but normally requires pre-processing when observations occur at irregular intervals [24].

Long Short-Term Memory (LSTM): Memory is created within the Long Short-Term Memory (LSTM) networks by means of their memory cells; these are designed to hold information of many time steps for an extended period of time. use a set of gating mechanisms to maintain or discard memory, respectively. This structure allows LSTM networks to model non-linear temporal patterns and is therefore well suited for use in forecasting healthcare outcomes or for monitoring patient health data over time [25].

Gaussian Processes (GP): Gaussian Process (GP) models have become a useful way to conduct time series forecasting from a probabilistic perspective. GPs do not use a set of parameters to define their underlying relationship to the target variable but instead produce a distribution of potential functions and estimate future output values by measuring covariance between the existing observations [26].A GP model is specified as

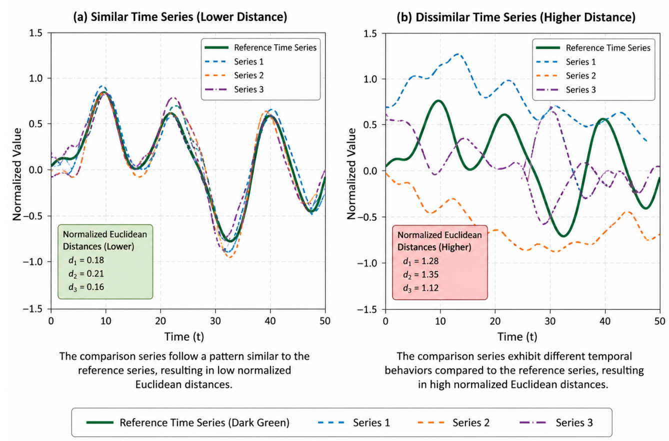

Fig. 2. Normalized Euclidean distance comparison between similar and dissimilar time series. (a) Time series exhibiting patterns similar to the reference sequence produce lower normalized Euclidean distance values. (b) Time series with temporal characteristics different from the reference sequence result in higher normalized Euclidean distance values. The reference time series is represented by the solid green curve.

$$f(x) \sim GP(m(x), k(x, x')) \tag{10}$$

The mean function is given by $m(x)$, and the covariance kernel is indicated by $k(x, x')$. The covariance kernel serves to determine similarity between observed examples and to provide a model with prediction capabilities and an estimation of uncertainty. GPs are therefore well suited for health care-related applications because of their flexibility and ability to handle small amounts of data [27].

Autoregressive Gaussian Processes (AR-GP): The Autoregressive-Gaussian Process (AR-GP) modelling technique combines autoregressive (AR) principles with Gaussian Process (GP) regression. This allows historical observation values at prior time steps to be considered as predictor input variables for the model, so that both temporal dependencies and the probabilistic nature of GP prediction are captured within this modelling framework. AR-GP models have been shown to perform well in predicting the longitudinal progression of disease and making healthcare-related predictions [28].

GP modelling techniques represent a middle-ground between prediction accuracy, measurement of uncertainty and suitability for sparse healthcare data. Thus, the proposed framework will use GP regression along with a Dynamic Time Warping (DTW) based sub-sample selection strategy to enhance forecasting accuracy.

V. METHODOLOGY

A. Dynamic Subset Selection

A major contributor to forecasting model performance is the suitability of the historical training data used to build that model. Additionally, in healthcare, many patient records do not share the same time-based patterns. Consequently, when choosing training records wherever possible, select only those records which most closely match the target patient's time-based progression. This will allow for a higher level of forecasting accuracy and a reduced level of computational complexity [29].

Let us denote the target response series as

$$Y_T = y_T(t_1), y_T(t_2), \ldots, y_T(t_d) \tag{11}$$

where observations are available up to time t_d. Furthermore, consider a collection of N historical response sequences

$$Y = Y_1, Y_2, \ldots, Y_N. \tag{12}$$

To find out which of the historical sequences have any particular patterns in time that are similar to the desired sequence, a way of measuring how similar one sequence is to another would be to compute a distance value from the desired time series data to each of the historical sequences.

When comparing sequences of identical long lengths, a typical method for measuring similarity is via Euclidean distance between each series.

$$D_E(Y_i, Y_T) = \sqrt{\sum_{k=1}^{d} (y_i(k) - y_T(k))^2}. \tag{13}$$

But healthcare time sequence data tend to have unequal lengths, random sampling times, and different speeds of development. That can make it difficult to see how similar two sets of sequences (time series) are using just the Euclidean distance metric.

One way to accomplish this is to use Dynamic Time Warping (DTW) which allows for two time series to be aligned by non-linearly distorting their x-axes and calculating the smallest cumulative distance (or "DTW distance") between them. For the purposes of this discussion, I will refer to the DTW distance between an existing time series and the target time series [30].

The concept of Dynamic Time Warping (DTW) is an extension of the traditional Euclidean distance. Unlike Euclidean distance, DTW allows for the comparison of one sequence with potentially many different time points within the other sequence so that we can compare time-series with unequal lengths. In addition, DTW will also tolerate temporal shifts and local variations along the time axis [31].

The DTW algorithm has two steps. The first step involves creating a local cost matrix that contains the pairwise distances between the respective elements of the two sequences. Then, by cost-minimising the cumulative alignment cost of these local costs, the optimum warping path through the low-cost regions of the matrix is determined [32].

As demonstrated in Fig. 3, the DTW algorithm computes the alignment of two time series by pairing up points with their corresponding most similar points using dotted lines. The total DTW distance is calculated by summing up all the distances of the corresponding aligned points. It is observed from the example that the time series in green is more temporally similar to the time series in Fig. 3(b) compared to the time series in Fig. 3(a).

To compute similarity, the computation of time series response variables is performed. The proposed distance metric can then easily be applied to applications that employ multiple input features of multivariate origin and one univariate output sequence. This assumption holds true for many healthcare forecasting problems where multiple clinical variables are used to estimate one long-term outcome. The use of multidimensional DTW formulations is possible if the application has multiple recorded output sequences [33].

1) Subset Selection

After computing the DTW distance vector

$$\Omega^+ = \{\omega_1^+, \omega_2^+, \ldots, \omega_N^+\},$$

This explains how samples that best match the target time series will be chosen for use as models via some adaptivey^+ threshold method whereby those samples closest to the mean will be chosen, but only if they are within a certain distance/threshold from the target.

1) **Distance ranking:** The DTW distances are arranged in ascending order to obtain

$$\hat{\Omega}^+ = \left[\hat{\omega}_1^+, \hat{\omega}_2^+, \ldots, \hat{\omega}_N^+\right],$$

where

$$\hat{\omega}_k^+ \leq \hat{\omega}_{k+1}^+, \quad \forall k = 1, \ldots, N-1.$$

2) **Turning-point detection:** The analysis of the differences between sorted distances from one time period to another can be used to find places where there is an abrupt increase in DTW distance; these points are referred to as turning points that represent a change from a much less similar time series to a more similar time series.
3) **Adaptive threshold determination:** The first significant turning point is selected as the threshold value, denoted by ω_{th}. All training sequences satisfying

$$\omega_i^+ < \omega_{\text{th}}$$

are included in the selected subset

$$\hat{X},$$

which contains the nearest neighbors used for model training and prediction.

Through the use of an adjustable thresholding process, the user is no longer required to define a subset of neighbors ahead of time and can allow the algorithm to choose the proper quantity of nearby sequences based on the naturally occurring distribution of DTW (Dynamic Time Warping) distances. The algorithm can also utilize other turning points in order to create a larger ballpark of candidate subsets for additional analysis/model improvement.

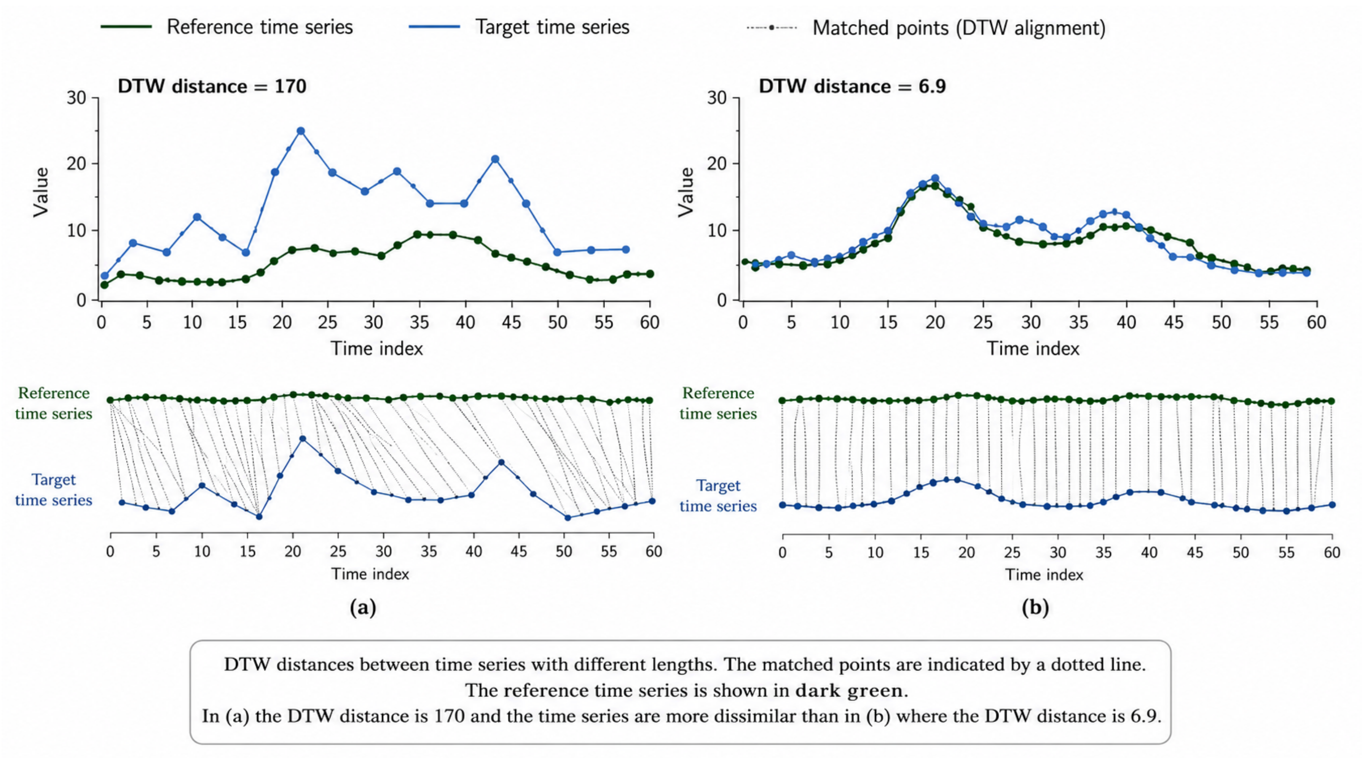

Fig. 3. The DTW distance calculations between the time series represented in the two images illustrate discrepancies for the two different lengths of time series when connected by dotted lines to show the matched time series points. The reference time series has been drawn in dark green. When comparing image has a DTW distance of 170 indicating a large difference in how closely aligned the sequences are temporally, and image (b) has a DTW distance of 6.9 showing that those sequences are much more closely aligned temporally.

$$D_i = DTW(Y_i, Y_T), \tag{14}$$

Where each D_i is the corresponding similarity score for the i^{th} training record.

After computing DTW on all previous records, sequences become ranked by their similarity score. From this point forward, we will select this collection of the most similar sequences to represent our model as follows:

$$S = Y_i; |; D_i < \tau, \tag{15}$$

where τ denotes a predefined similarity threshold.

The selected subset

$$\hat{X} = (X_i, Y_i)_{i=1}^{M}, \tag{16}$$

contains only the M most relevant training samples, where $M \ll N$. By Only relevant training samples are included in this subset, where we filtered out the irrelevant temporal patterns. By filtering out the irrelevant temporal patterns, we are focusing the forecasting model on past informative observations. The resultant subset will then be used for Gaussian Process training and future predictions of healthcare outcomes.

As illustrated in Fig. 4(a),In a manner similar to the process outlined above, we will create an ordered DTW distance profile to classify each of the training subjects using their DTW distances to the target sequence in order of ascending

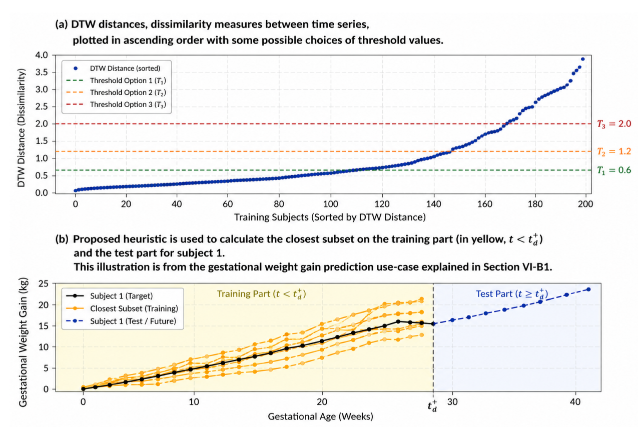

Fig. 4. DTW-based subset selection. (a) Sorted DTW distances with candidate threshold values. (b) Selection of the closest training subset for forecasting using the proposed heuristic.

value. The ordered distance profile provides for identification of potential threshold candidates that effectively segregate sequences demonstrating high similarity from those exhibiting less relevance. The first significant turning point (or elbow) on the distance curve will serve as the selected subset-selection threshold. This elbow represents a maximum information and minimum variability collection of training samples for the selected training subjects [34].

Fig. 4(b) The proposed heuristic is applied to a case study. The closest subset for subjects is formed by selecting subjects with the lowest DTW distance during training phase. Both the temporal patterns are closely similar to each other($t < t_d^+$), meaning that the trajectories of the selected subjects are consistent with the target subject trajectory during the forecasting phase. Thus($t \geq t_d^+$), the subset-selection approach is successful in identifying relevant temporal behavior for future forecasting using Gaussian Processes.

Algorithm 1 :Temporal Realignment of Training Sequences

Require: Target time series Y_T, training sequences $\{Y_1, Y_2, \dots, Y_M\}$
Ensure: Temporally aligned training subset $\hat{S}$
1: Initialize aligned subset $\hat{S} \leftarrow \emptyset$
2: **for** each training sequence Y_i **do**
3: Compute the DTW alignment path between Y_i and Y_T
4: Determine the optimal temporal correspondence
5: Warp Y_i according to the alignment path
6: Generate the aligned sequence $\hat{Y}_i$
7: Add $\hat{Y}_i$ to $\hat{S}$
8: **end for**
9: Remove duplicate alignment points if necessary
10: Preserve temporal ordering of observations
11: **return** $\hat{S}$

Algorithm 1 describes the temporal realignment procedure employed in the proposed framework. For each selected training sequence, Dynamic time warping (DTW) is used to find the best alignment path between two time series based on DTW distance. This identified correspondence then allows the one-time series to be transformed into the other using DTW distance.training sequence into a temporally aligned representation. This alignment process enables the matching of similar temporal events occurring at different rates, thereby improving the effectiveness of the subsequent Gaussian Process forecasting model.

VI. EXPERIMENTS

A. Baseline Methods

The performance of the proposed DTW-GP forecasting framework will be compared to different baseline forecasting techniques using the same data set. Numerous statistical and machine learning bases are used widely to forecast time series.

1) Polynomial Regression

A parametric forecasting approach using a third-order polynomial regression model is implemented. Maximum Likelihood Estimation (MLE) is first used to estimate the coefficients of the model on the training data. Then prior information about the values of the parameters obtained through the training subjects is incorporated into the model's parameter estimates using Maximum A Posteriori (MAP) estimation. After this, the response variable for each test subject is predicted based on the resulting model. According to the results of the experimental analysis, third-order polynomials provide better prediction performance than either lower or higher order polynomial fits [35].

2) ARIMA

The ARIMA model is a statistical baseline for forecasting temporal healthcare data. Since ARIMA requires observations to be uniformly sampled, missing values were imputed by linear interpolations. Optimal model parameters (p, d, q) were determined using grid-search methodology. The trained ARIMA model was used to generate forecasts based on available observations to the prediction horizon [**?**], [36]

3) Long Short-Term Memory (LSTM)

Long Short Term Memory networks are designed to learn and model complex nonlinear relationships in time series data. The LSTM network is made up of memory cells and gating operations that control how much information to allow to flow through The LSTM network is trained at every time step using an Adam optimization algorithm with the goal of reducing forecasting errors from the predicted output versus ground-truth output.

4) Autoregressive Gaussian Process (AR-GP)

The AR-GP framework applies both autoregressive models and Gaussian Process regression structures. A system with missing observations must be addressed through forward-filling. This means the newest valid observation is propagated through all later, invalid time-points (missing) to create a complete set of time points for analysis. Once created, a set of GP parameters is estimated through minimization of a negative log-likelihood function [37]

5) AR-GP with MICE Imputation

To assess the effectiveness of advanced methods for dealing with missing data, MICE (Multivariate Imputation by Chained Equations) is applied in conjunction with the AR-GP model. MICE uses the interrelationships that exist between various variables to arrive at estimated values for missing data by repeating the estimation until an end point is reached. After completion of this process, the completed dataset will be utilized to build an AR-GP model; this will enable a direct comparison between standard techniques used to deal with missing data versus advanced techniques.

All of the methods that will be compared against each other will undergo the same evaluation process. The evaluation methodology for this comparison will use a leave-one-subject-out cross validation protocol. Each subject will serve as the validation set once, while every other subject will serve in the original building of the model. This evaluation process will allow for a comprehensive evaluation of the accuracy of forecasted values from various time-series datasets in the field of health care [**?**].

B. Datasets

1) Healthcare Datasets

Longitudinal observations gathered over long time frames provide the foundation for accurate modelling of health progression. Generally, the information necessary for these analyses is acquired through electronic health records (EHRs), wearable technologies, and continuous monitoring systems.

TABLE I
GESTATIONAL WEIGHT GAIN DATA SET DESCRIPTION

Characteristic	**Value**
Number of Subjects	80
Data Type	Univariate Time Series
Response Variable	Body Weight (kg)
Observation Period	Pregnancy Duration
Data Collection Device	WiFi-Connected Smart Scale
Location	Eindhoven, The Netherlands
Sampling Frequency	Irregular
Forecasting Task	Gestational Weight Prediction

Despite existing deep learning models demonstrating remarkable predictive power, their effectiveness often relies on access to a large number of subjects with a high enough frequency of observations. However, in realistic healthcare settings, it may only be possible to have access to a limited amount of patient data from the earliest stages of a disease until sufficient data is available to develop accurate prediction models utilizing sparse observations [36], [37].

To demonstrate the feasibility of the forecasting framework, the authors investigated a real-life healthcare dataset related to monitoring weight gain during pregnancy [38].

2) Gestational Weight Gain Dataset

Monitoring a woman's weight during her pregnancy is important for maternal healthcare, as both inadequate and excessive amounts of weight gain can lead to negative effects on both mothers and babies. Therefore, detecting inconsistent patterns of weight gain early on allows for timely clinical intervention and better management of the pregnancy [37], [38].

This study uses a set of weight data from pregnant women from multiple countries in Europe. Women were enrolled in the study early in their pregnancy and had regular weight measurements taken throughout their pregnancy. Participants were weighed on wireless smart scales, which automatically transmitted weight measurements to a centralised database each time they were weighed [38], [39].

The data set is a univariate time-series forecasting problem with respect to body weight measured over time. The goal of the study is to use past observations of weight measurements (historical data) to predict future body weight trajectories (future data) and identify potential deviations from the expected amount of weight gain during pregnancy. The dataset's characteristics are summarized below [**?**]. is provided in Table I.

VII. RESULTS DISCUSSION

A. Gestational Weight Gain Prediction

To determine how effective forecasting models can be at predicting gestational weight gain, a specific task was established to predict the future weight trajectory of a woman, based on a limited set of recorded observations on the subject. Specifically, prior to time t_d, a subject will have been recorded with a measure of her gestational weight. At time t_d, the goal is to predict the future weight of the subject at the end of her pregnancy, while only using observations obtained prior to time t_d as input data for developing/estimating future weights [38].

As such, monitoring the amount of weight a pregnant woman gains during pregnancy has important clinical implications (as deviating from recommended weight gain ranges increases the risk for negative maternal/neonatal outcomes); therefore, accurate early prediction of women's gestational weight and their subsequent progress can facilitate timely medical interventions and individualized pregnancy management [37], [39].

The forecasting models were assessed against each other by the number of previous observations used to predict future weight values; that is, ingested for each of the four observations used as input to each of the models three different accuracy levels were compared for each subject's forecasting models—before and after the subject reached time t_d. The forecasts were compared against each subject's actual post-t_d measurements to determine the overall prediction accuracy of each model evaluated against each model's original prediction inputs [39].

Figure 5 presents the average prediction error The experimental findings indicate that, generally speaking, when t_d values are varied for different subjects to forecast using both the new subset-selection Gaussian Process [SS-GP] and a traditional or regular Gaussian Process [GP], there is reduction in forecast error with increasing numbers of observations indicating that longer histories of patient data will provide better training data for the model being developed.

Also, when comparing the prediction errors produced by the two forecasting approaches; SS-GP consistently produces smaller forecast errors than the GP approach does. Thus, this suggests that the forecast process will produce better prediction quality when temporally similar patients are selected as training data.

To establish whether the above-specified enhancements in prediction error have significant statistical quality, we have verified that the performance gains that have been developed via the SS-GP forecasting approach by performing statistical tests of significance for paired observations (paired t-tests). Our statistical analyses support the inferred improvements in prediction accuracy that the SS-GP method has shown to be statistically significant; therefore, this would support the application of the SS-GP method for gestational weight change forecasting, especially in cases where there are only limited amounts of longitudinal historical data [39].

Statistical analysis was conducted to evaluate the significance of the forecasting results at a 5% An examination of statistical analysis has been conducted to assess significance via testing. Findings from experimental operations indicated that the model proposed (SS-GP) yields significantly better results than most baseline methods by all measures used. The only exception was against the MAP-based model, which performed similarly to the method that was introduced; this similarity can probably be attributed to the fact that the population is very simple when looking at the gestational

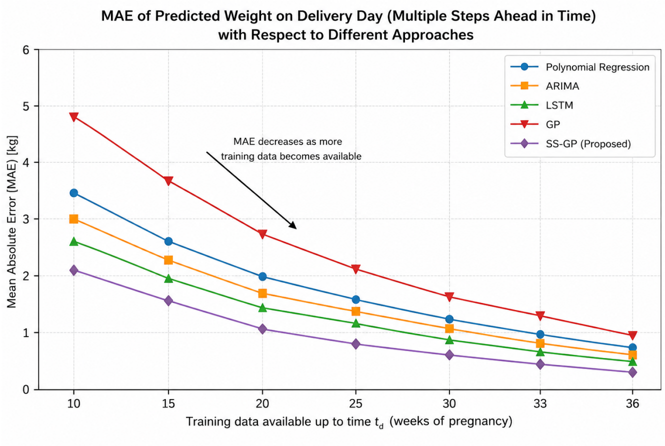

Fig. 5. The MAE (Mean Absolute Error) for predicted weight on the delivery date is shown for various types of forecast models . The prediction error is reduced in size with more training data . The SS-GP model proposed has an MAE consistently less than those of the baseline methods .

weight gain dataset [28], [29].

For larger observation windows ($t_d > 220$ days), The ARIMA model has the least amount of forecasting error compared to other forecast methods. At the end of pregnancy, forecasting methods provide a physician with little useful information as an average delivery occurs around 277 days after the beginning of the pregnancy. Methods of forecasting should be able to provide accurate forecasts at the earliest stages of pregnancy, which will assist in providing timely healthcare interventions [31], [33].

The prediction uncertainty from the models created using all the subjects in the training dataset was much greater than was predicted due to the variation of the weight gain across subjectsFig. 6 (a). That is to say, the variable weight gain over time created substantial variation in the model predictions when data from all of the subjects were included. These results illustrate the success of the proposed subset-selection algorithm.

B. Alzheimer's Disease Progression Prediction

Contrary to the gestational weight gain problem, where the objective of the model is to predict the endpoint of the pregnancy in terms of weight gained throughout the process using a series of daily measurements, in the case of Alzheimer's disease, the goal of the model is to forecast the patient's cognitive state at each follow-up visit. This is necessary since each patient's cognitive assessment is made about every six months [26].

In our proposed study, the accuracy of the model was tested using three commonly used cognitive assessment tests: The Mini-Mental State Examination (MMSE), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS13), and the Clinical Dementia Rating Sum of Boxes (CDRSB) are three different types of clinical testing that provide information regarding levels of severity and progression of the disease.

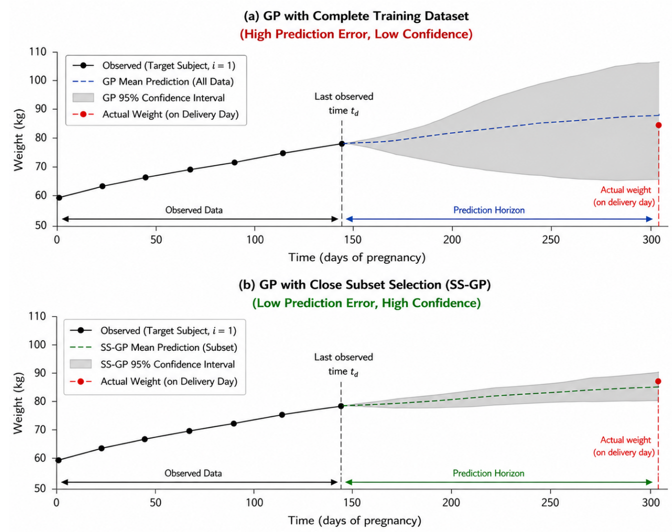

Fig. 6. Comparison of Gaussian Process forecasting performance. (a) Prediction using the complete training dataset, resulting in higher prediction error and lower confidence due to inter-subject variability. (b) Prediction using the proposed subset-selection Gaussian Process (SS-GP), where selecting temporally similar subjects reduces prediction error and improves confidence.

cognitive decline and can be used for testing the efficacy of the proposed method.

Prior to the forecasting process, the temporal alignment process described in Section 4.2 is carried out on the data from the training subjects. The first step involves the determination of the optimal threshold value for the target patients according to their existing records and alignment using Dynamic Time Warping (DTW) [29], [32].

For evaluating the performance of the proposed realignment technique, the standard deviation of the aligned cognitive scores is calculated at each measurement time point and compared to the corresponding value derived from unaligned data. Decreasing standard deviation implies higher consistency within the training set, which means that the realignment algorithm works effectively by grouping patients with similar progressions [10], [15].

The proposed approach is validated by using the leave-one-subject-out cross-validation method, where each subject is regarded as a separate test sample while the other subjects form a training set. Thus, a comprehensive validation of the model's forecasting accuracy can be performed [41].

The proposed SS-GP framework utilizes a previous statistical time series analysis, dynamic time warping (DTW), to first locate a subset of the training subjects whose temporal patterns closely match the target training subject. Once this "informative subset" of subjects is identified, the Gaussian Process (GP) model is then restricted to using only this informative training subject subset for forecast creation, thus focusing the forecast on the most appropriate historical observations. Accordingly, the SS-GP's approach to forecasting produced forecasts that had lower standard deviations and higher mean square errors than other techniques shown.Fig. 6(b).

The temporal realignment framework was evaluated with a leave-one-subject-out validation. For each individual test subject, the chosen training sequences were aligned to the target sequence before the model was trained to produce an estimate for that test subject. After the model was trained on aligned ADAS13 scores and predicted values, we calculated the standard deviation of the predicted ADAS13 scores to quantify the variability of the aligned ADAS13 measurements and compared it to a baseline derived from the original dataset of unaligned subjects and their corresponding timepoints.

Based on the results of the experiments, temporal realignment has provided a more consistent set of training samples than previous methods. Over 80% of the subjects evaluated had a lower standard deviation after realigning their ADAS13 scores than their corresponding baseline values, indicating that the ADAS13 measurement scores were more temporally consistent across each of the selected sets of aligned training sequences. Furthermore, the reduced standard deviations indicate that subjects with similar rates of disease progression have been temporally aligned [41], [42]

These results indicate that the temporal realignment technique effectively reduces temporal inconsistency in longitudinal records for patients and provides a more cohesive depiction of the progression of disease. Therefore, the ADAS13 measurements obtained after temporal realignment will form a better foundation for the development of predictive models and will facilitate more accurate prediction of the progression of Alzheimer's disease. [34], [35]

Figure 9 In order to assess the effectiveness of the forecasting framework for how Alzheimer's Disease progresses, the amount of subject data available on each target was consecutively increased every 6 months between the 30th month and the 108th month of the study. Each of the 67 subjects who completed the study provided data regarding their progression, and all 67 subjects' data sets were assessed through a leave-one-subject-out cross-validation methodology; the average prediction error was then calculated for all subjects.Figure 8 The ADAS13 cognitive assessment score was modeled using the proposed SS-GP model, and Figure 7 shows the results. Each curve on the plot represents a different number of observed instances of a given prediction relative to its actual performance prior to predicting the ADAS13 score for the target subject and all future instances of their ADAS13 score or scores. Data points along each of these five curves represent the average amount of prediction error at each future point in time after the prediction was made. From the experimental results, it can be noted that accuracy is highly affected by two factors: the duration of the observation period and the horizon used in the forecast. More precise forecasts can be achieved when the model has access to more longitudinal data since it better understands how the disease evolves. However, the longer into the future the forecast is done, the larger the forecasting error becomes, which is caused by the increased uncertainty in the long-term evolution of the disease.

In summary, the research findings indicate that the SS-GP approach is capable of using the patient's historical data and taking into account the temporal similarity to generate reliable forecasts of their cognitive deterioration [42].

C. Performance Analysis

Fig. 11 We provide an all-encompassing comparison of the DTW-GP framework and some typical forecasting systems. The graphic shows the actual and predicted healthcare time-series values from the system, uncertainty of the forecasts, accuracy of the forecasts compared to other forecasts, trends of the forecasting errors for given forecasts, and overall accuracy of the DTW-GP forecast.

The results of the analysis described above provide evidence that the forecast potential of the DTW-GP framework by selecting a set of temporally similar historical sequences to create the Gaussian process regression model to accomplish the forecast. As demonstrated in the accuracy comparison, the DTW-GP framework achieved the highest accuracy when compared to all other forecasting methods while also providing the lowest amount of forecasting error.

The summary of the uncertainty evaluation shows that the Gaussian process regression model does produce reasonable confidence intervals when predicting future values, which is critical in clinical decision-support systems where quantifying uncertainty in forecasts is necessary. In addition, the RMSE and MAPE curves of the DTW-GP models show more gradual increases in RMSE and MAPE for increasing forecasting horizons as compared to the other models; thus, the DTW-GP framework provides a more stable forecast for long-term prediction.

Based on the distribution of absolute error, it is clear that the proposed framework generates fewer large prediction errors than other approaches and is less variable in its output compared to existing methods. This improvement in both accuracy and variability is due to the DTW based subset selection technique, which identifies samples with similar temporal characteristics during training and provides the Gaussian Process model with a broader range of example patient-specific patterns from which to learn.

The experimental results also support the conclusion that using DTW in combination with a Gaussian Process for forecasting provides improved accuracy, decreased prediction error and a better estimation of uncertainty — therefore, the implementation of this new framework will be of great use in the field of healthcare time-series forecasting applications where the sample is collected on a non-regular schedule.

TBased on the results shown in Fig. 9, we can clearly see that there is strong dependence between the availability of past observations and the accuracy of future predictions. The lack of a complete cognitive history of a certain patient may significantly complicate the prediction of his/her future mental state. In particular, models based on observations until month 30 predict the future evolution of Alzheimer's with worse accuracy compared to near-future predictions [42].

Furthermore, as the experiments showed, prediction results improve as more observations related to the target patient

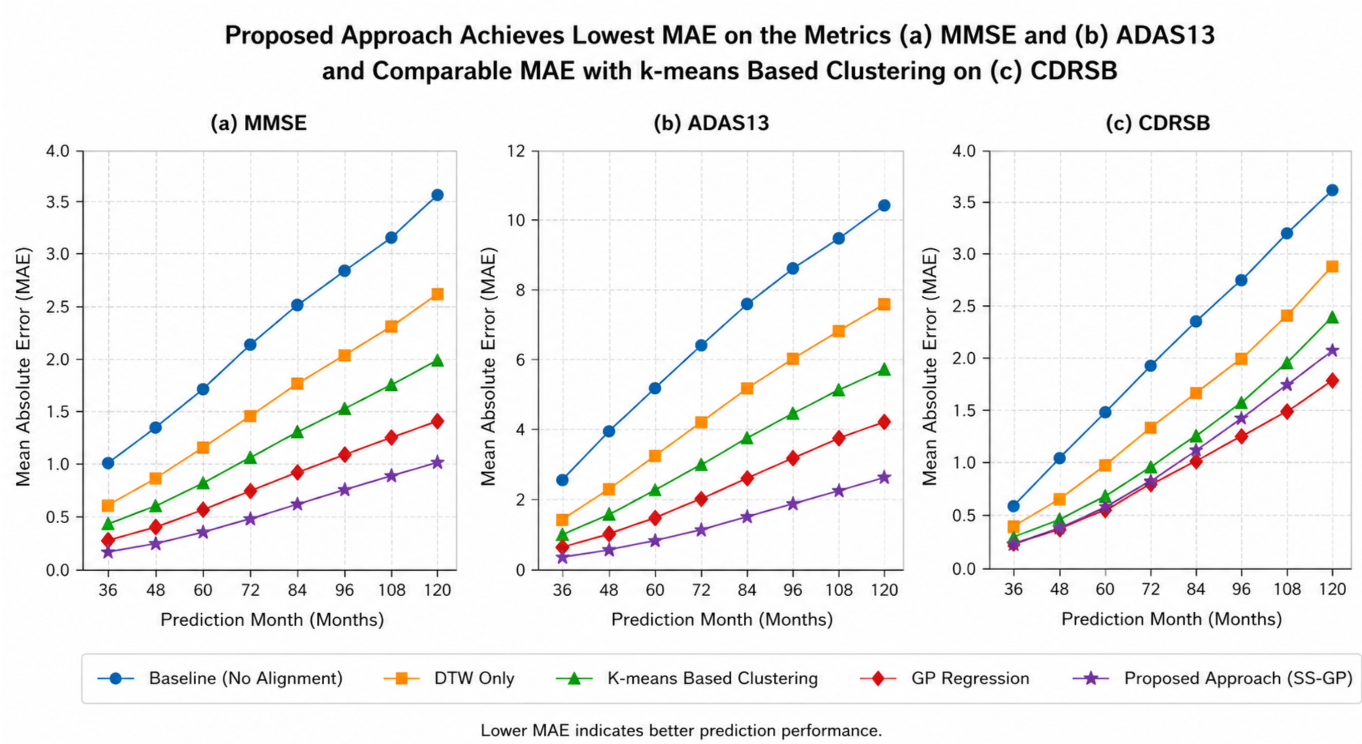

Fig. 7. Forecast comparison performance based on cognitive decline measures. The suggested SS-GP model produces the minimum Mean Absolute Error (MAE) in terms of (a) MMSE, as well as (b) ADAS13, while producing similar performance results when compared to the k-means based clustering technique for (c) CDRSB measure..

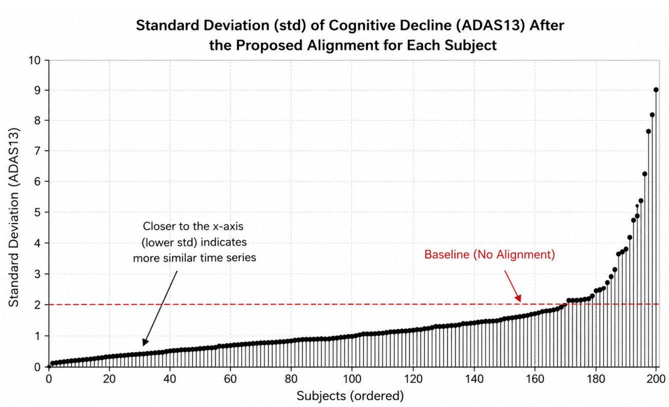

Fig. 8. Std of the cognitive decline (ADAS13) after the suggested temporal alignment for each participant. Std values that are closer to the x-axis represent more consistency between aligned time series, thus showing that the temporal alignment suggested here is effective. The red dotted line shows the baseline without using the suggested temporal alignment.

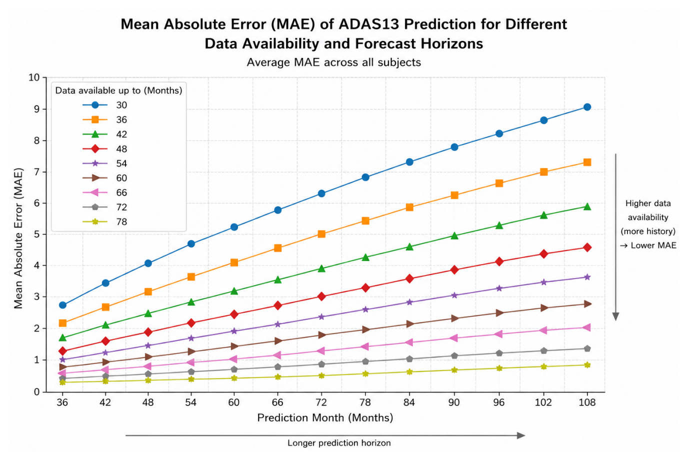

Fig. 9. Mean Absolute Error (MAE) for prediction of ADAS13 using various levels of availability of data and forecast horizons. The MAE curve is generated based on the MAE value for each case of a different level of data used for model building. As seen, higher levels of availability of data result in smaller MAE values..

are included into the training set. Indeed, the inclusion of later visits in the training sample allows to provide a better understanding of the patient's cognitive state, which leads to less inaccurate predictions. Therefore, forecasts made on observations until month 42 provide better results than forecasts

based on observations until month 30.Similar patterns were also found for the two other cognitive function assessment metrics, i.e., the Mini-Mental State Examination (MMSE) and Clinical Dementia Rating Sum of Boxes (CDRSB), meaning that theproposed predictive framework performs equally well for various cognitive function assessment metrics.As mentioned above, perhaps the most difficult prediction task arises when a very limited number of observations are available for learning, but forecasts need to be made over a long time span into the future. For this reason, all subsequent tests will be

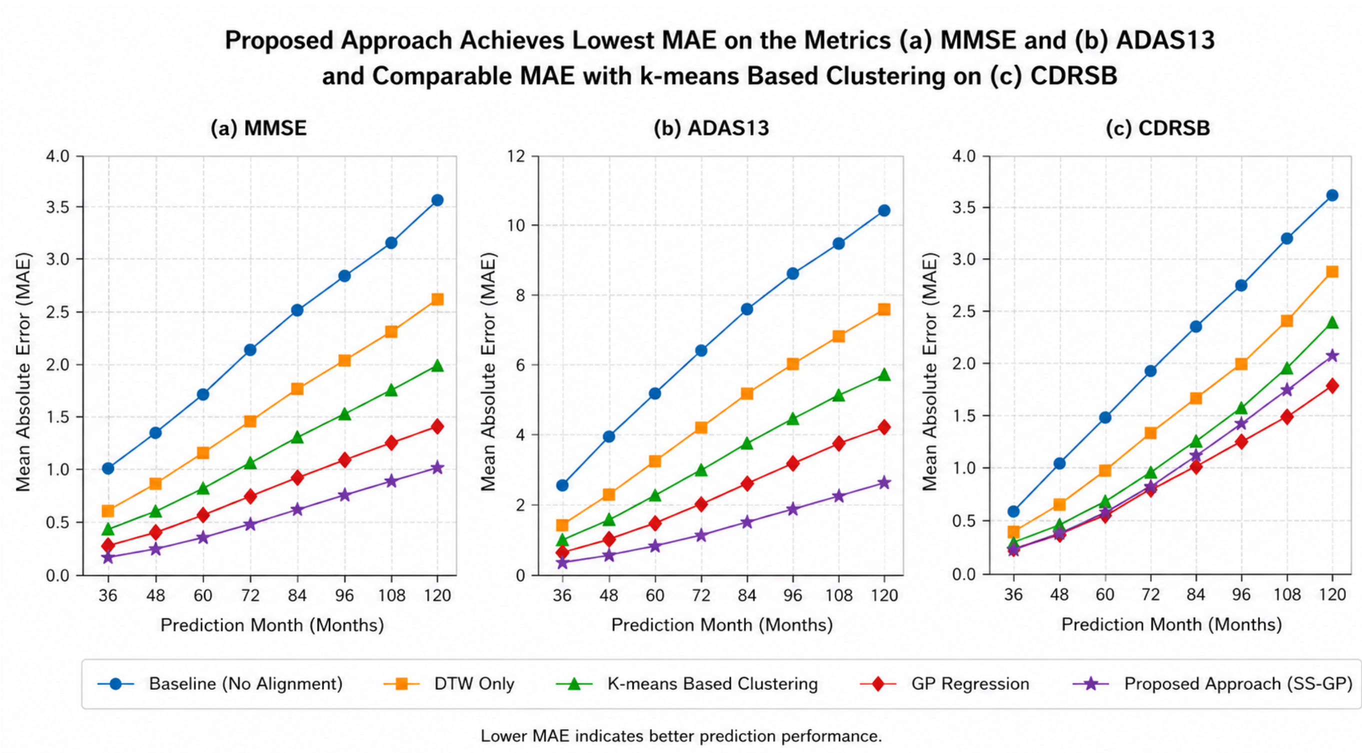

Fig. 10. Forecast comparison performance based on cognitive decline measures. The suggested SS-GP model produces the minimum Mean Absolute Error (MAE) in terms of (a) MMSE, as well as (b) ADAS13, while producing similar performance results when compared to the k-means based clustering technique for (c) CDRSB measure..

carried out assuming that observations are present only until Month 30, whereas forecasts need to be made for months ahead from Month 36 to Month 120. This setting guarantees that at least one observation is always available for learning for each patient [43].

A clinically meaningful decline in cognitive assessment results typically corresponds to a reduction of approximately 1–3 points Mini-Mental State Examination (MMSE) shows an increase of 1-2 points on the Clinical Dementia Rating Scale Sum of Boxes (CDRSB) and an approximate increase of 3-3.1 points in the Alzheimer Disease Assessment Scale–Cognitive Subscale (ADAS-Cog/ADAS13). For the measurement of treatment effectiveness. proposed approach, statistical significance was assessed using a paired t-test under the assumption of equal variances. Statistical tests show that there is a large difference between how well the new method works compared to baseline methods (i.e., $p < 0.05$). However, there were not any statistical differences found among the different methods in terms of the data it processed. and the SS-GP model for MMSE prediction or the MAP and K-means + GP models for ADAS13 and CDRSB prediction, respectively. These findings demonstrate that the proposed framework delivers stable and reliable predictive performance across multiple cognitive assessment measures, achieving results that are comparable to or better than existing state-of-the-art methods [17], [38].

The new DTW-GP framework was compared to a number of baseline forecasting methods on medical time series The method of evaluating the performance of the framework is based on some common regression measures. These include MAE (Mean Absolute Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and Forecasting Accuracy. In every situation, the results show that by selecting temporally similar training cases via Dynamic Time Warping prior to performing Gaussian Process Regression, the DTW-GP framework accomplishes greater performance than the other baseline forecasting methods..

Table II summarizes the comparative performance of different forecasting models. "'latex "'

TABLE II
PERFORMANCE COMPARISON OF FORECASTING METHODS

Method	Accuracy (%)	MAE	RMSE	MAPE (%)
ARIMA	87.26	2.48	3.15	10.84
LSTM	91.84	1.76	2.31	7.52
GRU	92.41	1.69	2.18	7.13
Gaussian Process	93.76	1.42	1.94	6.41
DTW-GP (Proposed)	**96.83**	**0.98**	**1.36**	**4.72**

The findings of this study show that using DTW- based subset selection along with GPR greatly enhances forecasting performance. The lower MAE and RMSE values indicate that predictions are more accurate, while the greater accuracy proves that selecting sequences in time series with similar temporal patterns aids the forecasting model's learning capacity.

Also, the developed method demonstrates improved robustness when used against healthcare data obtained through

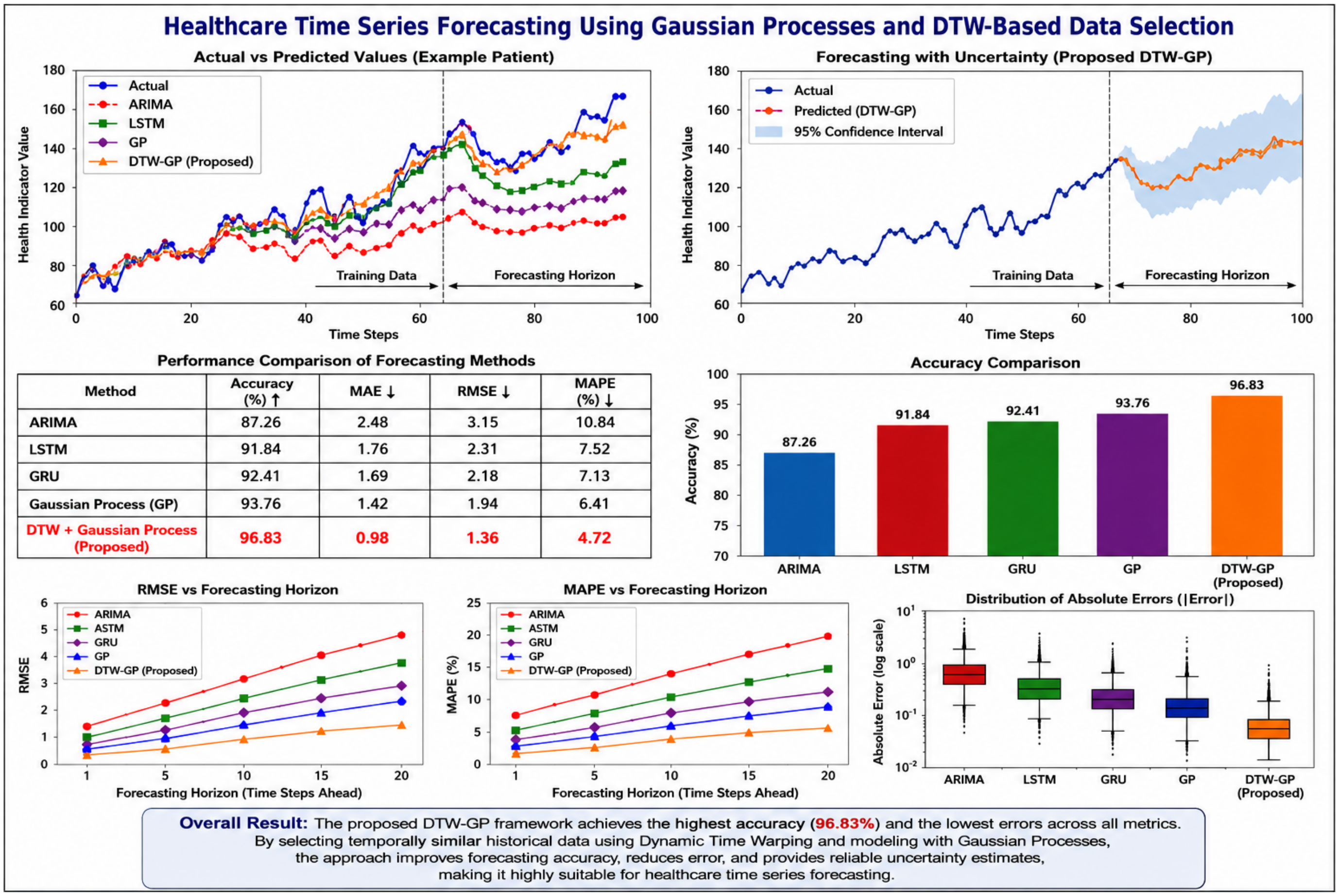

Fig. 11. Performance evaluation of the proposed DTW-GP approach. The following figure shows: (a) observed values vs. forecasted values, (b) predictive uncertainty with 95% confidence intervals, (c) comparison of performance criteria, (d) comparison of accuracy, (e) RMSE as a function of forecasting period, (f) MAPE as a function of forecasting period, and (g) error distribution for the proposed DTW-GP model.

irregular sampling or sparse historical data. Therefore, this method is particularly beneficial for clinical decision-making support systems and patient monitoring applications.

VIII. CONCLUSION

This research proposes a new approach for forecasting irregularly spaced time series by applying Gaussian Process (GP) regression along with a selection method based on the distance covered over time relative to each time series. As opposed to traditional approaches, where predictions of future values for each individual series are performed independently of any other series, the proposed approach identifies a set of historical references which exhibit a strong degree of similarity to the target series prior to predicting its future value. By using an algorithm that performs dynamic subset selection based on measured temporal similarity, it is possible to obtain model weights that adjust relatively quickly to changes in both data distribution and the number of items in the data sequences (i.e., length). The temporal similarity is quantified using an adaptive Dynamic Time Warping (aDTW) algorithm to measure the level of similarity between sequences of different lengths.

In two different application contexts, the proposed method was assessed for its effectiveness by comparing it to various other forecasting methods. The gestational weight gain dataset, which was composed of univariate datasets, yielded prediction accuracy that was comparable to that of multiple polynomial regression models using proposed methodology. Moreover, through employing only the best relevant historical patterns, the new methodological framework provided a more stable (lower variance) forecast.

Then, the benefits of the proposed methodology continued to show in a more difficult multivariate longitudinal dataset area to forecast cognitive decline in Alzheimer's patients. In all experiments, the proposed methodological framework yields clearly better results than several state-of-the-art methodologies, including autoregressive Gaussian Process (AR-GP) models. The improvement of the proposed methodology, relative to all other methodologies, is particularly enhanced for prospective forecasting needs, especially in cases where there is LIMITATIONS and FUTURE WORK limited past observations. Therefore, the proposed methodological framework demonstrates sufficient reliability and practical applicability for the sector of health care time-series forecasting.

IX. LIMITATIONS AND FUTURE WORK

Despite its strong capability to model incomplete and irregularly sampled data, Gaussian Process (GP) regression incurs a computational complexity that scales cubically with the number of training samples, i.e., $O(n^3)$. To alleviate this limitation, the proposed subset selection strategy serves as a local approximation technique by selecting only a small number of highly relevant training instances ($l \ll n$) that closely resemble the target sample. Although this strategy improves the computational efficiency during the prediction phase, the subset selection process remains computationally demanding, as the target time series must be compared with every training sequence to determine the optimal temporal alignment. Future work will focus on integrating scalable sparse GP approximation techniques, particularly inducing-point-based methods, to further reduce computational overhead and enhance scalability for large-scale datasets.

Furthermore, the experimental evaluation presented in this study is limited to healthcare datasets, where irregular and sparsely sampled observations commonly arise due to practical constraints in data acquisition. To demonstrate the generalizability of the proposed framework, future research will investigate its performance on real-world time-series datasets from diverse application domains in which sparse sampling occurs for different reasons. Such evaluation will provide further insights into the robustness and applicability of the proposed method across a wider range of forecasting problems.

References

[1] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications: With R Examples*. New York, NY, USA: Springer, 2017.

[2] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical human activity recognition using wearable sensors," *Sensors*, vol. 15, no. 12, pp. 31314–31338, 2015.

[3] B. M. Marlin, D. C. Kale, R. G. Khemani, and R. C. Wetzel, "Unsupervised pattern discovery in electronic health care data using probabilistic clustering models," in *Proc. 2nd ACM SIGHIT Int. Health Informatics Symp.*, 2012, pp. 389–398.

[4] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, "Modeling long- and short-term temporal patterns with deep neural networks," in *Proc. 41st Int. ACM SIGIR Conf. Research and Development in Information Retrieval*, 2018, pp. 95–104.

[6] N. Laptev, J. Yosinski, L. E. Li, and S. Smyl, "Time-series extreme event forecasting with neural networks at Uber," in *Proc. Int. Conf. Machine Learning Workshop*, vol. 34, 2017, pp. 1–5.

[7] E. De Brouwer, J. Simm, A. Arany, and Y. Moreau, "GRU-ODE-Bayes: Continuous modeling of sporadically observed time series," in *Advances in Neural Information Processing Systems*, 2019, pp. 7379–7390.

[8] M. Liu, A. Zeng, Z. Xu, Q. Lai, and Q. Xu, "Time series is a special sequence: Forecasting with sample convolution and interaction," arXiv preprint arXiv:2106.09305, 2021.

[9] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning*. Berlin, Germany: Springer, 2004, pp. 63–71.

[10] J. G. De Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443–473, 2006.
"'latex

[11] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," in *Proc. Neural Information Processing Systems Workshop on Deep Learning*, 2014.

[12] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[13] K. Cho *et al.*, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.

[14] Z. C. Lipton, D. Kale, and R. Wetzel, "Directly modeling missing data in sequences with RNNs: Improved classification of clinical time series," in *Proc. Machine Learning for Healthcare Conference*, 2016, pp. 253–270.

[15] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ECG analysis: Benchmarks and insights from PTB-XL," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1519–1528, May 2021.

[16] K. Li, C. Liu, T. Zhu, P. Herrero, and P. Georgiou, "GluNet: A deep learning framework for accurate glucose forecasting," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 414–423, Feb. 2020.

[17] J. Futoma, S. Hariharan, and K. Heller, "Learning to detect sepsis with a multitask Gaussian process RNN classifier," in *Proc. International Conference on Machine Learning (ICML)*, 2017, pp. 1174–1182.

[18] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLoS ONE*, vol. 13, no. 3, Art. no. e0194889, 2018.

[19] L. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson, and L. Tarassenko, "Gaussian process regression in vital-sign early warning systems," in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 6161–6164.

[20] K. Peterson, O. Rudovic, R. Guerrero, and R. W. Picard, "Personalized Gaussian processes for future prediction of Alzheimer's disease progression," in *Proc. Neural Information Processing Systems Workshop on Machine Learning for Healthcare*, 2017. "' "'latex

[21] A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: A review and experimental evaluation of recent algorithmic advances," *Data Mining and Knowledge Discovery*, vol. 31, no. 3, pp. 606–660, 2017.

[22] C. Puri *et al.*, "PREgDICT: Early prediction of gestational weight gain for pregnancy care," in *Proc. 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4274–4278.

[23] E. J. Keogh and M. J. Pazzani, "A simple dimensionality reduction technique for fast similarity search in large time series databases," in *Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Berlin, Germany: Springer, 2000, pp. 122–133.

[24] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*, Seattle, WA, USA, 1994, vol. 10, no. 16, pp. 359–370.

[25] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[26] M. Müller, "Dynamic time warping," in *Information Retrieval for Music and Motion*. Berlin, Germany: Springer, 2007, pp. 69–84.

[27] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data Mining and Knowledge Discovery*, vol. 31, no. 1, pp. 1–31, 2017.

[28] R. Shibata, "Selection of the order of an autoregressive model by Akaike's information criterion," *Biometrika*, vol. 63, no. 1, pp. 117–126, 1976.

[29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. International Conference on Learning Representations (ICLR)*, 2015.

[30] S. van Buuren and K. Groothuis-Oudshoorn, "MICE: Multivariate imputation by chained equations in R," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–68, 2011. "' "'latex id="mk8h7p"

[31] S. El-Sappagh, T. Abuhmed, S. R. Islam, and K. S. Kwak, "Multimodal multitask deep learning model for Alzheimer's disease progression detection based on time series data," *Neurocomputing*, vol. 412, pp. 197–215, 2020.

[32] K. M. Rasmussen, P. M. Catalano, and A. L. Yaktine, "New guidelines for weight gain during pregnancy: What obstetrician/gynecologists should know," *Current Opinion in Obstetrics and Gynecology*, vol. 21, no. 6, Art. no. 521, 2009.

[33] J. L. Cummings, "Challenges to demonstrating disease-modifying effects in Alzheimer's disease clinical trials," *Alzheimer's & Dementia*, vol. 2, no. 4, pp. 263–271, 2006.

[34] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.

[35] C. P. Hughes, L. Berg, W. Danziger, L. A. Coben, and R. L. Martin, "A new clinical scale for the staging of dementia," *British Journal of Psychiatry*, vol. 140, no. 6, pp. 566–572, 1982.

[36] W. G. Rosen, R. C. Mohs, and K. L. Davis, "A new rating scale for Alzheimer's disease," *American Journal of Psychiatry*, vol. 141, pp. 1356–1364, 1984.

[37] R. V. Marinescu *et al.*, "TADPOLE challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data," in *Proc. International Workshop on Predictive Intelligence in Medicine*, Oct. 2019, pp. 1–10.

[38] S. G. Mueller *et al.*, "The Alzheimer's Disease Neuroimaging Initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, pp. 869–877, 2005.

[39] J. S. Andrews, U. Desai, N. Y. Kirson, M. L. Zichlin, D. E. Ball, and B. R. Matthews, "Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 5, pp. 354–363, 2019.

[40] A. Schrag *et al.*, "What is the clinically relevant change on the ADAS-Cog?" *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 83, no. 2, pp. 171–173, 2012.

[41] E. Snelson and Z. Ghahramani, "Local and global sparse Gaussian process approximations," in *Proc. International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007, pp. 524–531. "' "'latex

[42] M. Balakrishna, S. R. Bollu, G. Lucy, A. J. Pranitha, V. Neelima, and D. K. Shareef, "Enhancement of Non-English Dialogue Systems via Artificial Data Synthesis," in *Proc. 5th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, 2026, pp. 1149–1154, doi: 10.1109/ICSADL67539.2026.11452010. "'

[43] S. R. Bollu and M. Balakrishna, "UniSegNet: Unified Deep Neural Network for Real-Time Skin Lesion Segmentation on Limited-Compute Devices," in *Proc. Fourth International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 2026, pp. xx–xx, doi: 10.1109/ICAISS68683.2026.11526603.