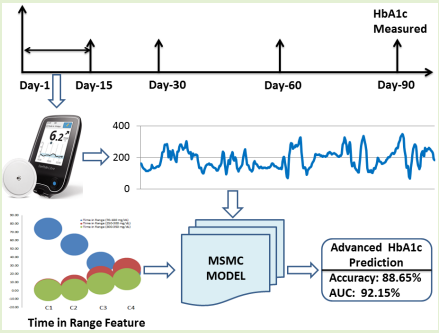
Long Term HbA1c Prediction Using Multi-Stage CGM Data Analysis

Md Shafiqul Islam, Marwa Khalid Qaraqe, *Member, IEEE*, SamirBrahim Belhaouari, *Senior Member, IEEE*, and Goran Petrovski

***Abstract*—The glycated hemoglobin (HbA1c) is regarded as an essential biomarker for diabetes management. Having an elevated HbA1c level significantly increases the risk of developing diabetes-related health complications. Accurate *prediction* of HbA1c can greatly improve the way diabetic patients are treated and can potentially avoid related consequences. This study devises a framework to predict HbA1c levels 2–3 months in advance by using blood glucose data collected through continuous glucose monitoring (CGM) sensors and leveraging advanced feature extraction and machine learning techniques. The CGM data may often contain missing values due to sensor issues or not wearing the sensor for some period. Thus, in the paper, a novel missing data estimation method has been proposed for a single data point, multiple data points, and entire day CGM data imputation. The CGM data have been rigorously investigated, and pertinent features were created along with a multi-stage multi-class (MSMC) classification model to predict futuristic HbA1c levels. To evaluate the developed framework, a total of 150 patients' data were sourced from Sidra Medicine, Doha, Qatar, for analysis. The proposed three-staged and five-staged MSMC models predicted HbA1c levels 2–3 months in advance and obtained overall classification accuracies of 88.65% and 83.41%, respectively.**



***Index Terms*—CGM sensor, diabetes management, feature extraction, HbA1c prediction, missing data estimation.**

I. Introduction

THE global incidence rate of diabetes mellitus (DM) is overgrowing. The DM is described by an increase in blood sugar concentrations. A hormone known as insulin, generated in the human body's pancreas, regulates blood sugar levels. Insufficient insulin production or the inactive response of the body cells to insulin causes the disease [1]. Hyperglycemia, a significant hallmark of diabetes, can damage different body organs and cause severe health complications such as cardiovascular diseases if it goes uninterrupted for a long period. The ramification of DM overloads the health-care system, influences economic growth, and increases the cost of the treatment. Diabetes patients are required to keep their glycemic profile under control to extend a healthier living. One of the significant biomarkers used for diabetes monitoring is the glycated hemoglobin (HbA1c). The HbA1c is a particular hemoglobin structure that makes a bonding with glucose in the bloodstream [2]. The HbA1c test provides an estimation of average glucose concentrations for the previous 90—120 days. The proper management of diabetes significantly depends on the periodical assessment of the HbA1c levels. The HbA1c test is performed in the lab by measuring the percentage of hemoglobin attached to the blood sample. The test is often performed to classify diabetes severity and to forecast upcoming complexities [3]. The American Diabetes Association (ADA) defines a test value of HbA1c $<5.7\%$ as non-diabetic. The test values between 5.7% and 6.4% are regarded as pre-diabetes states while HbA1c value $\geq 6.5\%$ is considered as the subject developed diabetes.

Advanced prediction of HbA1c is significant for proper monitoring of diabetes. Studies [4]–[6] demonstrated that lower-levels of HbA1c play an essential role in reducing or late triggering of microvascular difficulties arise from diabetes. However, there is an association between elevated HbA1c levels and the development of diabetes-related comorbidities. The prediction of HbA1c given current blood glucose trends allows patients and physicians to make changes to treatment plans, lifestyle, foods, to avoid elevated HbA1c levels. Consequently, an advanced interference will facilitate avoiding complications, and thus better diabetes management can be

Manuscript received January 11, 2021; revised April 4, 2021; accepted April 13, 2021. Date of publication April 19, 2021; date of current version June 30, 2021. This work was supported in part by a scholarship from Hamad Bin Khalifa University (HBKU), a member of Qatar Foundation for Education, Science, and Community Development. The associate editor coordinating the review of this article and approving it for publication was Prof. Chao Tan. *(Corresponding author: Md Shafiqul Islam.)*

Md Shafiqul Islam, Marwa Khalid Qaraqe, and SamirBrahim Belhaouari are with the College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar (e-mail: mislam@hbku.edu.qa; mqaraqe@hbku.edu.qa; sbelhaouari@hbku.edu.qa).

Goran Petrovski is with the Sidra Medicine, Doha, Qatar (e-mail: gpetrovski@sidra.org).

Digital Object Identifier 10.1109/JSEN.2021.3073974

TABLE I
CGM DATA COLLECTION SUMMARY

Total subjects	Mean age ± SD (years)	Age range (years)	CGM Data Duration	CGM device	Samples per day	Mean HbA1c ± SD (%)	HbA1c range (%)
150	12.7 ± 4.5	6-22	90 Days	Free Style Libre	96	8.99 ± 2.13	5.2-14.5

ensured. Several studies concluded that HbA1c levels can be used to infer the future progression of diseases such as cardiovascular disease (CVD), nerve, eye, and kidney damage [4]. The population study of East Asian by Sakurai *et al.* [7] found high values of HbA1c increase the likelihood of mortality and death from CVD. Another study [8] correlated HbA1c with mortality and found a resilient connection between elevated HbA1c and mortality among the subjects without a previously known history of diabetes. Diabetic retinopathy is another health complication that arises from diabetes. The researchers investigated the association between HbA1c and retinopathy and found that 10% reduction of HbA1c reduces 43% of retinopathy development risk [9].

Researchers attempted to estimate a patient's current HbA1c level from blood glucose (BG) values. The studies [10]–[14] estimated current HbA1c levels using self-monitoring average blood glucose (μ_{BG}) measurements and reported the coefficient of determination (R^2) values around 0.80. All studies **estimate** current HbA1c that provides information about the patients' present HbA1c levels and their past BG values. The previous studies didn't use advanced continuous glucose monitoring (CGM) sensor data to predict HbA1c. The recent advancements in medical sensor technologies provide a massive amount of information in electronic health records (EHR) [15]. The unprocessed EHR data may provide limited intuition into patients' health. Therefore, it is necessary to develop an automated framework that utilizes advanced techniques to detect disease progression in the early stages, limiting the potential damages. In this study, a multi-stage multi-class (MSMC) machine learning (ML) framework is devised for advanced prediction of HbA1c levels. The present study offers the following significant contributions:

1) A new clinical trail is designed for advance prediction of HbA1c and data from 150 diabetes subjects have been collected using CGM sensor.
2) A novel method has been proposed for missing data estimation.
3) Seven new techniques are introduced to derive pertinent features utilizing CGM sensor data.
4) The extracted significant features are selected, and redundant components are discarded using a new feature ranking method.
5) An MSMC framework is proposed for advanced prediction of HbA1c.

II. RESEARCH DESIGN AND METHODS

The workflow of the proposed methodology to predict HbA1c has been outlined in Fig. 1. The CGM sensor data collected from Sidra Medicine, Qatar, have been utilized to develop the framework. A total of 15 days of data from every patient have been used to build the model. The missing CGM data points are imputed using a novel data imputation procedure. Pertinent features are extracted by implementing seven new methods. The features significantly related to HbA1c have been nominated by assigning a feature importance value calculated using the Pearson correlation. Finally, the selected features are used to build the MSMC model for long-term HbA1c prediction.

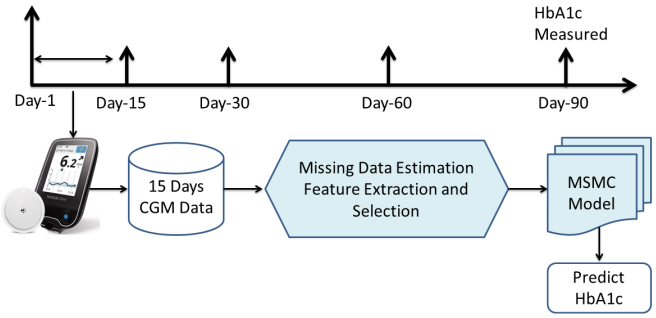


Fig. 1. The proposed methodology of HbA1c prediction.

A. Data Model

The study enrolled one hundred and fifty subjects (mean age 12.7 ± 4.5 years; range 6-22 years) with T1DM during 2019 and 2020. The institutional review board (IRB) of the Sidra Medicine, Doha, Qatar, has approved the research plan (IRB Number, 1536761-1). The pediatric division at Sidra Medicine specializes in children's general care and offers pediatric patients clinical care. All recruited subjects wore CGM sensors, Freestyle Libre, for 90–120 days. The CGM device comprises a glucose sensor implanted into the body's subcutaneous tissue. The sensor measures interstitial fluid glucose levels every 15 minutes and gives 96 measurements per day. The CGM sensor has a lifetime of 14 days, and then it was replaced with a new one. The 14 days CGM data from the sensor were collected and saved to a secured memory disk. All the subjects continued using the CGM sensors for 90 days. The HbA1c level was measured for each subject on the 90th day of data collection at the Sidra Medicine laboratory. The data collection summary is provided in Table I.

B. Data Preparation

This study analyzes 150 patients' 2250 days CGM sensor data. The patients are split into six and ten classes based on their HbA1c control levels, as outlined in Table II and III. The class, C1, consists of 47 subjects with HbA1c levels ≤ 7.5%. The subjects in C1 have HbA1c values in the expected

TABLE II
SPLIT OF 150 PATIENTS INTO SIX (C1-C6) CLASSES BASED ON THEIR HbA1c LEVELS

Class	HbA1c range (%)	Number of subjects
C1	HbA1c≤7.5	47
C2	HbA1c>7.5	103
C3	7.5<HbA1c≤9	42
C4	HbA1c>9	61
C5	9<HbA1c≤12.5	35
C6	HbA1c>12.5	26

TABLE III
SPLIT OF 150 PATIENTS INTO TEN (S1-S10) CLASSES BASED ON THEIR HbA1c LEVELS

Class	HbA1c range(%)	Number of subjects
S1	HbA1c≤6.5	25
S2	HbA1c>6.5	125
S3	6.5<HbA1c≤7.5	22
S4	HbA1c>7.5	103
S5	7.5<HbA1c≤8.25	20
S6	HbA1c>8.25	83
S7	8.25%<HbA1c≤9	22
S8	HbA1c>9	61
S9	9<HbA1c≤10.5	20
S10	HbA1c>10.5	41

range and therefore they are referred as the good control group. Contrarily, the class C2 consists of 103 subjects whose HbA1c values >7.5%. The subjects with HbA1c levels in the range (7.5%–9%) are assigned to the class C3. The patients in C3 have their HbA1c values above the expected levels and therefore, they are defined as medium control group. However, the class C4 includes subjects with HbA1c values >9%. The subjects with HbA1c levels between 9% and 12.5% are grouped together in class C5. The patients belongs to the class C5 have their HbA1c values significantly higher than the expected levels. Therefore, the subjects in C5 are defined as poor control group. Finally, the subjects with HbA1c levels >12.5% are grouped together in the class C6. The patients in the C6 have their HbA1c values very high as compared to the expected levels. Therefore, the subjects in C6 are referred as uncontrolled group.

Furthermore, the subjects have been split into ten classes (S1–S10) to evaluate the proposed MSMC model's efficacy in predicting a narrow range of HbA1c levels. The class S1 consists of 25 subjects with HbA1c levels ≤ 6.5%, while a total of 125 subjects whose HbA1c values >6.5% are assigned to the class S2. The remaining classes formed by including subjects based on different HbA1c ranges are S3, S4, S5, S6, S7, S8, S9, and S10, as outlined in Table III.

III. MISSING DATA ESTIMATION

As CGM devices are worn and need to be replaced every two weeks, there are missing data instances. To address this, a data processing stage is incorporated to i) evaluate the amount of missing data and ii) develop a missing data estimation method to impute short spans of missing data.

A. Single Point Estimation

In cases where there is only one missing BG value, and the nearest values are available, the missing BG values are estimated by taking into consideration the previous and next available data points. A line is drawn, as shown in Fig. 2 by connecting the two nearest data points, and the line's slope is measured. If the slope is positive, the difference between the two nearest data points is divided by two and added with the immediate nearest previous BG value to estimate the missing value. However, the missing value is replaced by dividing the difference between the two adjacent points for the negative slope, followed by adding the immediate next nearest BG value. Finally, if the slope is zero, the missing data point is replaced by the immediate nearest previous BG value. Consequently, the equations used to estimate the missing BG values are:

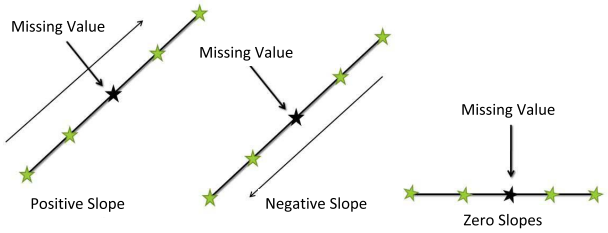


Fig. 2. The estimation of point based missing CGM data-based of values of slope.

$$\mathbf{x_i} = x_{i-1} + \frac{x_{i-1} - x_{i+1}}{2} \quad (slope > 0) \tag{1}$$

$$\mathbf{x_i} = x_{i-1} - \frac{x_{i-1} - x_{i+1}}{2} \quad (slope < 0) \tag{2}$$

$$\mathbf{x_i} = x_{i-1} \quad (slope = 0) \tag{3}$$

where x is the individual BG value.

B. Multiple Points Estimation

A nearest neighbors approach has been implemented for missing data estimation when there are two or more missing data points. The BG values from eight neighbors are considered to replace the missing values, as outlined in Table IV. The missing BG value (x) is estimated by taking the eight neighbors' average. The inter and intra-day BG values are included for multiple missing data points estimation. Some random single and multiple data points have been eliminated to assess the efficacy of the estimations. The missing values are then estimated by the mentioned nearest neighbors approach. An R^2 value of 0.82 ($\pm$0.13) is observed in the estimations.

C. Whole Days Estimation

Some patients do not wear their sensors for a day, or the sensor may have expired in some cases. As we predict HbA1c in advance by using short term CGM data (15 days), it is crucial that missing days are accounted for. Thus, in the event that data is missing for 24 hours, we assume that the subject follows a similar daily routine in food intake and fills the missing day with the previous day. We also implemented other algorithms for the entire day missing data estimation. In the case of 2 or more missing days of CGM data, an autoregressive

TABLE IV
THE ESTIMATION OF MISSING CGM DATA POINTS USING EIGHT NEAREST NEIGHBOURS

Before pre-processing

Day	$BG_{12:00AM}$	$BG_{12:15AM}$	$BG_{12:30AM}$
1	175	176	159
2	234	**x**	255
3	132	128	126
4	78	89	96
5	167	145	139

After pre-processing

Day	$BG_{12:00AM}$	$BG_{12:15AM}$	$BG_{12:30AM}$
1	175	176	159
2	234	**173**	255
3	132	128	126
4	78	89	96
5	167	145	139

moving average (ARMA) method is used to estimate the data as given by:

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{4}$$

where x_t is a missing data point. For a whole missing day, there are 96 such missing data points, t is the timestamp at which the data point is missing. $\varphi_1, \ldots, \varphi_p, \vartheta_1, \ldots, \vartheta_q$ are parameters, c is a constant, and the random variable ε_t is white noise. These parameters of ARMA model are estimated using maximum likelihood function in Matlab toolbox. The estimations of a full day CGM data for a randomly selected subject are compared with the true BG values as shown in Figure 3. The estimations are close to the actual BG measurements (R^2, 0.76 ±0.15).

IV. FEATURE EXTRACTION AND SELECTION

Feature extraction is one of the fundamental steps in ML-based classification tasks. The collected raw BG data are transformed to extract pertinent features to have improved model performance. This study introduces seven different feature extraction methods for advanced HbA1c prediction.

A. Fractional Derivative Feature

A person's reaction to food consumption indicates their glucose metabolizing capacity (GMC). The glucose levels will be higher for the person with poor GMC [16]. Consequently, their HbA1c levels will also be higher as it summarizes the average glucose present in the bloodstream. A new set of GMC features are derived by adapting the fractional derivatives (FD) method [17]. The k^{th} order FD of a function $g(x)$ is defined by:

$$\mathrm{g}^{(\mathrm{k})(\mathrm{x})} \approx \lim_{\mathrm{h}\to 0} \frac{\mathrm{g(x)} - \mathrm{kg(x-h)} + \frac{\mathrm{k(k-1)}}{2}\mathrm{g(x-2h)} + \ldots}{h^k} \tag{5}$$

The above expression is simplified by considering the first two components only and dividing by the time difference as:

$$\mathrm{g}^{(\mathrm{k})(\mathrm{x})} = \frac{\mathrm{g(x+h)} - \mathrm{kg(x)}}{(\mathrm{t(x+h)} - \mathrm{t(x)})^k} \tag{6}$$

This study extracted different GMC biomarkers based on:

$$\mathrm{GMC}^{(\mathrm{k})} = \frac{\mathrm{BG}_i - \mathrm{kBG}_j}{(t_i - t_j)^k} \tag{7}$$

where BG is blood glucose, t_i and t_j are times at which the BG levels have been collected through CGM sensor, i and j are different time indices ($i \neq j$). For each value of k = 1, 2, 0.5, −1, and 0.1, total 95 GMC biomarkers have been derived.

B. Time Range Feature

Time in range (TIR) is defined by the proportion of time a patient passes in a specific range over the total time of data collection. The typical range for a diabetic should be within 70–180 mg/dL. TIR and HbA1c have been found to exhibit high correlation [18]. The study leverages this correlation and introduces novel TIR features to detect fluctuations in BG levels that are significantly interrelated with HbA1c. In particular, seven TIR features are defined as shown in the following equations:

$$\mathrm{TBR}_{54} = \frac{\sum_{i=1}^{N}(C(x_i) \leq 54}{)} N \tag{8}$$

$$\mathrm{TBR}_{70} = \frac{\sum_{i=1}^{N}(C(x_i) \leq 70}{)} N \tag{9}$$

$$\mathrm{TIR}_{70-180} = \frac{\sum_{i=1}^{N}(C(x_i) \geq 70 \wedge \leq 180}{)} N \tag{10}$$

$$\mathrm{TIR}_{180-250} = \frac{\sum_{i=1}^{N}(C(x_i) \geq 180 \wedge \leq 250}{)} N \tag{11}$$

$$\mathrm{TIR}_{250-300} = \frac{\sum_{i=1}^{N}(C(x_i) \geq 250 \wedge \leq 300}{)} N \tag{12}$$

$$\mathrm{TIR}_{300-350} = \frac{\sum_{i=1}^{N}(C(x_i) \geq 300 \wedge \leq 350}{)} N \tag{13}$$

$$\mathrm{TAR}_{350} = \frac{\sum_{i=1}^{N}(C(x_i) \geq 350}{)} N \tag{14}$$

where C represents overall counts, TBR stands for time below range, TAR is the time above range, x stands for individual BG values, and N represents the sample size.

C. Cyclostationary Feature

A cyclostationary signal has statistical properties that fluctuate with time. It is represented as multiple interleaved stationary signals. For example, hourly BG measurement variation can be modeled as a cyclostationary process because today's hourly BG value at noon will be significantly different than the BG values in the morning for a specific subject; however, it is a realistic approximation that for a particular subject, the daily BG values at 6 am will have similar statistics. Thus CGM data can be incorporated as the random signal composed of 24 interleaved stationary processes (representing 24 hours of a day), each taking on a new value once per day. The CGM sensor provides a BG measurement in every 15 minutes. The hourly BG values are derived by taking the average of four BG measurements for the corresponding hour. An example of extracted cyclostationary features BG values is outlined in Table IV for illustrative purposes.

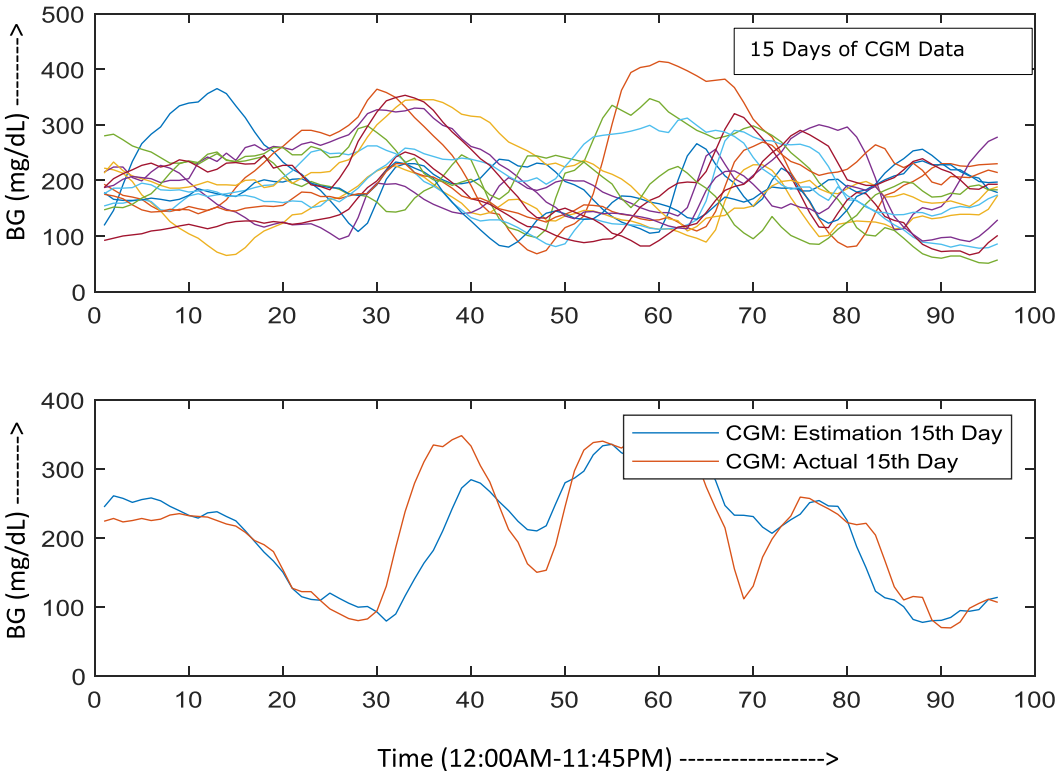


Fig. 3. The estimation of whole day missing CGM data using ARMA model.

Time	06:00AM	06:15AM	06:30AM	06:45AM	07:00AM	07:15AM	07:30AM	07:45AM
BG (mg/dL)	125	120	117	110	130	143	157	180

Time	06:00AM	07:00AM
BG (mg/dL)	**118**	**152.5**

Fig. 4. Extracted cyclostationary feature of a randomly selected subject.

D. Glucose Variability Feature

Glucose variability (GV) represents the measure of oscillations in BG levels for a defined time such as during a day or among days. The GV is considered one of the fundamental indexes used to assess the patient's overall glucose profile. Different GV features have been extracted using CGM data adapted from [19], as outlined in Table VI. The coefficient of variation (CV) is expressed as a percentage whose high value indicates greater dispersion around the mean. The CV is often preferred over SD as a GV feature because the mean highly influences SD. Data with a high mean value usually have a high SD. Thus, to normalize the variability, the SD is divided by the mean while calculating the CV. The GV index M100 provides a measure of the variation of glucose values around 100 mg/dL. Another important GV index, J-index, is a measure of glucose variability used to assess the patient's glycemic profile calculated from average and SD. Mean amplitude of glycemic excursion (MAGE) is another important metric used for evaluating a patients' glycemic variation. The MAGE is derived by calculating the deviations between the successive top and bottom values larger than one SD of average BG. The mean of daily differences (MODD) indicates glucose fluctuations between days. MODD is derived as the average of absolute differences among the BG levels of consecutive days. Continuous overall net glycemic action (CONGA) measures glycemic variability within a defined time window. The CONGA is computed by taking the differences among the BG data points, and then SD is calculated on these differences. The glycemic risk assessment diabetes equation (GRADE) score expresses the associated risk for observed BG levels. The GRADE score is described as proportions: <70 mg/dL, 70–180 mg/dL, and >180 mg/dL are refer to hypoglycemia, euglycemia, hyperglycemia, respectively.

E. Wavelet Decomposition Feature

The features extracted from the wavelet decomposition (WD) technique are extensively used for healthcare applications [20]. This study incorporated Haar WD techniques for feature extraction from CGM data. One of the naïve but extensively used WD techniques is the Haar basis [21]. The Haar basis coefficients are obtained using the points' pairwise average and then subtracting the average value from the pair's first component. In the next steps, averages are computed but differences remain unchanged. This study implemented similar addition and subtraction approaches to generate WD features and the derived WD features for different classes are compared.

F. Power Spectral Density Feature

The power spectral density (PSD) defines power distribution into frequency components composing that signal. Welch's

method is used for estimating spectral density at different frequency levels. It uses to convert a time series signal into frequency domain components. Welch's method decreases the noise while estimating the power spectrum by sacrificing the frequency resolution. It is the advanced method for calculating power spectra as compared to standard periodogram power estimation. The average power, P, of a signal $x(t)$ is derived based on:

$$\mathrm{P} = \lim_{T\to\infty} \frac{1}{T} \int_0^T |x(t)|^2 \, dt \tag{15}$$

where T is the total time duration of the signal $x(t)$. The power spectral density, $S_{xx}(\omega)$, is derived as:

$$S_{xx}(\omega) = \lim_{T\to\infty} \mathbf{E}\left[\left|\hat{x}(\omega)\right|^2\right] \tag{16}$$

where $E[\hat{x}(\omega)]$ is the expected value of the signal $x(t)$ in frequency domain, and ω (rad/sec) is the frequency of the signal. The extracted power spectral density features for the individual classes are used to developed the proposed framework.

G. Time Series Feature

Time series feature extraction is considered as one of the preliminary steps of the ML framework. It is a complex task as it involves domain knowledge and coding an implementation. In this study, different time series features have been extracted from CGM data using a Python package titled Time Series Feature Extraction (tsfresh). The tsfresh implements 63-time series characterization to extract temporal, statistical, and spectral features. In this study, the following time series features have been extracted from the CGM data. The absolute energy (E) of the CGM is the sum over the squared BG values calculated based on:

$$E = \sum_{i=1,\ldots,n} x_i^2 \tag{17}$$

where x is the individual BG value. The absolute sum of changes (ASC) yields the summation over the absolute value of successive variations in the BG values, and it is calculated using:

$$ASC = \sum_{i=1,\ldots,n-1} | \, x_{i+1} - x_i \, | \tag{18}$$

The autocorrelation R(l) of BG values for lag l is derived as:

$$R(l) = \frac{1}{(n-l)\sigma^2} \sum_{t=1}^{n-l} (x_i - \mu)(x_{i+l} - \mu) \tag{19}$$

where n is overall observations, σ^2 and μ are variance and mean of BG values. The autoregressive coefficient (ARC) of CGM is extracted using the maximum likelihood of an autoregressive system:

$$x_i = \varphi_0 + \sum_{n=1}^{k} \varphi_i x_{i-n} + \varepsilon_i \tag{20}$$

where k is the maximum lag. The process returns AR coefficients φ_i. A more complicated time series has more peaks, valleys, etc. The time series complexity (TSC) feature for BG values is estimated based on the following equation.

$$TSC = \sqrt{\sum_{i=1}^{n-1} (x_i - x_{i-1})^2} \tag{21}$$

The coefficients of Ricker wavelet (RW), a continuous wavelet transform, are derived from:

$$RW = \frac{2}{\sqrt{3a}\pi^{\frac{1}{4}}} (1 - \frac{x^2}{a^2}) exp(-\frac{x^2}{2a^2}) \tag{22}$$

where a is the width of the RW function. The Fourier coefficients for BG values are extracted by using a fast Fourier transformation algorithm:

$$A_k = \sum_{i=0}^{n-1} x_i \exp\left\{-2\pi j \frac{mk}{n}\right\}, \quad k = 0, \ldots, n-1 \tag{23}$$

which returns the complex coefficients, and j is the imaginary unit. The only real part of the coefficient is extracted as a feature. The entropy is calculated by split data into bins that are as pure as possible: most of the values in a bin belong to the same class. The bin entropy (E) is derived for CGM data based on:

$$E = \sum_{k=0}^{min(max\_bins, len(x))} p_k log(p_k) \cdot \mathbf{1}_{(p_k>0)} \tag{24}$$

where p_k is the percentage of samples in bin k.

H. Feature Selection and Fusion

Feature selection is considered a crucial step during ML architecture development. It facilitates the removal of redundant features. Integrating all the features generated from individual techniques into a compact feature vector is defined as feature fusion. The fused set of pertinent features can enhance model performance. All the extracted features are merged to generate the ultimate feature set of size 1050, and the finalized data size is 150×1050. This study applies the filter method, a correlation-based feature selection technique, to find pertinent features significantly related with the outcome variable HbA1c. A statistical measure known as the Pearson correlation is used to rank features based on their values. Pearson correlation measures the linear dependence between two variables, lies between -1 and 1, is calculated using:

$$\rho_{X_1,X_2} = \frac{\mathrm{cov}_{X_1,X_2}}{\sigma_{X_1}\sigma_{X_2}} \tag{25}$$

where cov is the covariance, σ_{X_1} and σ_{X_2} are standard deviation of the feature vector X_1 and X_2 respectively.

Furthermore, the extracted features are ranked based on their values in a decreasing order. The highly discriminating features are selected by visual inspection and redundant features are discarded. The WD feature selection results for four classes (C1, C3, C5, and C6) are shown in Fig. 5 as an example. A total of 102 WD features are ranked for each class in descending order based on their wavelet coefficient values. It is observed that the 1-12 ranked feature values of class C1 are significantly lower than class C6. For class

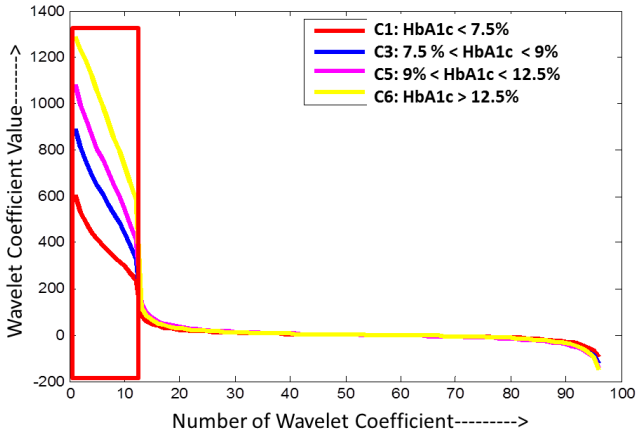


Fig. 5. The extracted WD feature selection based on their coefficient values.

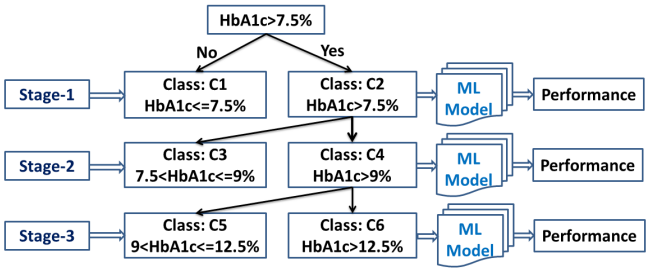


Fig. 6. Proposed three-staged MSMC model for HbA1c prediction.

C3 and C5, there are noticeable differences for 1-12 positioned feature values. However, there is no significant difference for 13–102 ranked feature values among all four classes. Therefore, those 13–102 ranked features are discarded during feature selection as those features poorly correlate with the outcome variable HbA1c.

V. MSMC Machine Learning Framework

The proposed three-staged and five-staged MSMC ML framework for HbA1c prediction is summarized in Fig. 6 and 7. The model's outcome is the 2–3 months advanced prediction of HbA1c ranges using the past 15 days of CGM data. The three-staged MSMC model involves a total of three stages to accomplish the classification tasks. In stage 1, the aim is to develop an optimized ML model to differentiate between C1 (HbA1c $\leq$ 7.5%) and C2 (>7.5%). Another optimal ML model classifies instances into class C3 (7.5% < HbA1c $\leq$ 9%) and class C4 (>9%) in stage 2. In the final stage, the aim is to distinguish C5 (9% < HbA1c $\leq$ 12.5%) from C6 (>12.5%). However, the five-staged MSMC model consists a total of five classification stages. The ML model in stage 1 differentiate between S1 (HbA1c $\leq$ 6.5%) and S2 (>6.5%). In the second stage, another optimal ML model classifies instances into class S3 (6.5% < HbA1c $\leq$ 7.5%) and class S4 (>7.5%). The third stage distinguishes S5 (7.5% < HbA1c $\leq$ 8.25%) from S6 (>8.25%). In the subsequent stage, separate ML models are developed and optimized to distinguish between the classes: S5 vs. S6, S7 vs. S8, and S9 vs. S10.

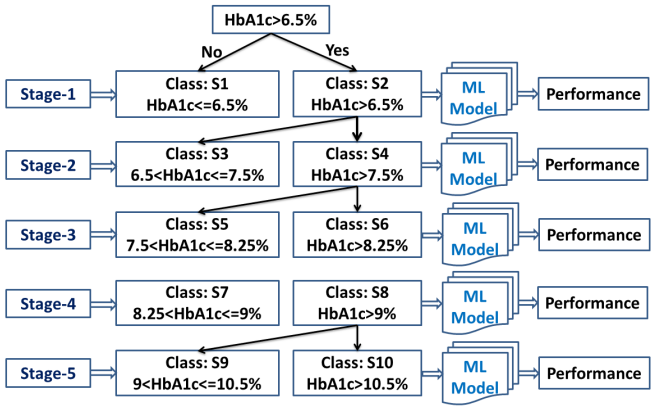


Fig. 7. Proposed five-staged MSMC model for HbA1c prediction.

The three-staged classification approach ultimately divides HbA1c ranges into four distinct patients groups. These patient groups belong to classes C1, C3, C5, and C6, respectively. Conversely, the five-staged MSMC model has distinguished six unique patient classes. These classes are S1, S3, S5, S7, S9, and S10.

A. Support Vector Machine

The SVM uses the Lagrangian optimization technique to find the best plane that maximizes the margin among the classes [22]. It implements a kernel trick to differentiate the outcomes when the data are not linearly distinguishable. The data are transformed into higher dimensional space using a kernel trick that ensures linear separation. The study adapted polynomial SVM for advanced HbA1c prediction. The hyper-parameters C and Γ are optimized in a brute-force manner. The hyperparameter C is a regularization factor responsible for margin flexibility, while the Γ value controls the position of the hyperplane. The optimal value of the C parameter is 100, while a value of 10 is found to be optimal for the Γ parameter.

B. Naive Bayes

The Naive Bayes (NB) classifiers are probabilistic models developed applying Bayes' theorem. The model, coupled with kernel density function, achieves higher performance with a strong assumption of independence between the features. This study ensembles three classifiers, namely, the NB, averaged one-dependence estimators (A1DE), and averaged two-dependence estimators (A2DE), to predict HbA1c levels. The NB assumes complete feature independence. However, the A1DE and A2DE models relax the assumption and apply weaker independence among the features and achieve higher accuracy than the NB model [23]. The ensembling steps are: first, data have been split into ten folds, then three models are developed using nine folds data, and the tenth rest fold is used to test the models. Each model provides its class probability on the test samples. Finally, probabilities are combined from the individual classifiers' decision to finalize the test samples' class.

C. Random Forest

The ML techniques often encounter a bias-variance trade-off property. To reduce bias and variance of the model, a method

known as boosting is used. In boosting, several classifiers' outcomes are combined to come up with the final decision. It uses an iterative approach where misclassified tuples are given more attention in the next step by increasing weight. On the contrary, the bagging approach avoids over-fitting by randomly sampling data while building trees and, thus, improves the model's stability and accuracy. The random forest (RF) is one such bagging model that randomly samples data and selects a subset of the features to add more randomness [24]. This study adapted and optimized the RF model for long-term HbA1c prediction, previously used in the study [25]. A total of three hyperparameters, namely, split criterion, the number of estimators, and the minimum samples split, are searched for optimal values. The highest performance is achieved for the gini impurity criterion, 200 estimators, and minimum samples split of 10.

VI. Result and Discussion

The results of feature extraction and selection are highlighted and evaluated. The 10-folds cross-validation (CV_{10}) results of the developed MSMC framework is discussed and compared.

A. Performance Evaluation

The metrics used to assess the efficacy of the developed MSMC framework are: Accuracy $= (tp+tn)/(tp+tn+fp+fn)$, Sensitivity $= tp/(tp+fn)$, Specificity $= tn/(tn+fp)$, and AUC $= p(Score(tp) > Score(tn)$, where tp, tn, fp, and fn stand for true positive, true negative, false positive, and false negative instances. The accuracy is the number of cases correclty predicted out of all cases. The sensitivity measures the proportion of tp cases accurately detected, while the specificity measures the numbers of tn cases correctly identified. The area under the curve (AUC) indicates the probability of a tp event will be ranked higher as compared to a tn event.

B. Feature Evaluation

The results of extracted TIR characteristics for classes C1, C3, C5, and C6 are summarized using a bar graph, as shown in Fig. 8. The subjects' 15 days of CGM data have been investigated to find the association between TIR and HbA1c. The feature values for TIR (70–180 mg/dL), TBR (<70mg/dL), and TIR (180–250 mg/dL) of the class C1 were found 74.15%, 1.82%, and 21.46%, respectively. However, the value for TIR (70–180 mg/dL) feature of class C6 was much lower (22.19%) as compared to C1 (74.15%). The features TBR (<70mg/dL) and TIR (180–250 mg/dL) also followed a linear association with the outcome variable. From the analysis in Fig. 8, it can be inferred that the proposed TIR features are strong predictors of the future HbA1c levels.

The extracted PSD features for classes C1, C3, C5, and C6 are presented in Fig. 11. It shows the differences in PSD values (dB) for different classes. The observation is that PSD values of class C1 are significantly lower than class C6 throughout the frequency spectrum. For class C3 and C5, there are noticeable differences in feature values. However,

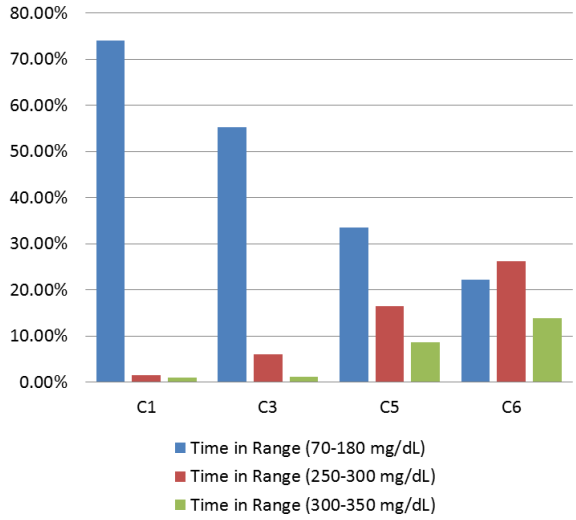


Fig. 8. Time in range features comparison for different classes.

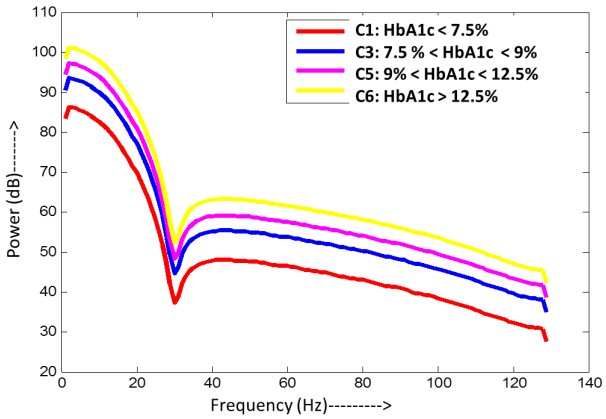


Fig. 9. The PSD features with their frequency components.

to reduce the redundancy, top 20 PSD features based on their power values are selected for model development and evaluation.

C. HbA1c Prediction Model Performance

The cross-validation results of the proposed three-staged MSMC models are presented in Fig. 10 and outlined in Table V. The ensembling approach obtained 90.67% accuracy, 91.48% sensitivity, 90.29% specificity, and 92.15% AUC score in stage 1 while classifying classes C1 and C2. The SVM achieved 83.49% accuracy, 85.71% sensitivity, 81.96% specificity, and an AUC score of 86.23% in stage 2 while differentiating between classes C3 and C4. In the final stage, the proposed A1DE model distinguished class C5 from C6 with 91.80% accuracy, 89.54% sensitivity, 88.19% specificity, and 90.72% AUC. The developed model displayed an overall accuracy of 88.65% when tested with the entire dataset.

Furthermore, the HbA1c prediction performance for the proposed five-staged MSMC model is summarized in Table VIII. The RF model manages 90% accuracy, 88% sensitivity, 90.4% specificity, and 92.37% AUC score in stage 1 while classifying class S1 and S2. Accuracy has dropped to 84.8% during classification of classes S3 and S4 in stage 2 using the SVM. The lowest accuracy of 79% is observed in the final stage of the MSMC model while separating class S9 from S10.

TABLE V
LIST OF GLUCOSE VARIABILITY FEATURES EXTRACTED

Feature	Formula	Description
$\bar{x}$	$\bar{x} = \frac{\sum_{k=1}^{n} x_i}{n}$	$\bar{x}$= mean of the BG values n= number of observations
SD	$SD = \sqrt{\frac{\sum_{k=1}^{n}(x_i - \bar{x})^2}{n}}$	SD= Standard deviation of the BG values x_i= individual BG value $\bar{x}$= mean of the BG values n= number of observations
CV	$CV[\%] = \frac{SD}{\bar{x}} * 100\%$	SD= Standard deviation of the BG values $\bar{x}$= mean of the BG values
LBGI and HBGI	$f(x) = 10 * (1.509 * (\ln(x)^{1.084} - 5.381))^2$ $rl(x) = \begin{cases} f(x), \text{ if } f(x) < 0 \\ 0, \text{ if } f(x) \geq 0 \end{cases}$ $rh(x) = \begin{cases} 0, \text{ if } f(x) \leq 0 \\ f(x), \text{ if } f(x) > 0 \end{cases}$ $LBGI = \frac{\sum_{k=1}^{n} rl(x_i)}{n}$ $HBGI = \frac{\sum_{k=1}^{n} rh(x_i)}{n}$	LBGI= low blood glucose index HBGI= high blood glucose index x= BG value x_i= individual BG value n= number of observations
M100	$M100 = \frac{\sum_{k=1}^{n} 1000 * \left\lvert \log\left(\frac{x_i[\frac{mg}{dl}]}{100}\right)\right\rvert}{n}$	x_i= individual BG value n= number of observations
J-index	$J\text{-index} = 0.001 * (\bar{x} + SD)^2$	SD= Standard deviation of the BG values $\bar{x}$= mean of the BG values
MAGE	$MAGE = \sum\left(\frac{\lambda}{n}\right), for\ each\ \lambda > SD, where\ \lambda\ >\ SD.$	SD= Standard deviation of the BG values n= number of observations λ= difference between peak and nadir of BG values
MODD	$MODD = \frac{\sum_{i=24h}^{n}(x_i - x_{i-24h})}{n}$	x_i= individual BG value n= number of observations
CONGA	$CONGA(t) = \sqrt{\frac{\left(\sum_{i=t}^{n}(x_i - x_{i-t}) - \frac{\sum_{l=t}^{n} x_l - x_{l-t}}{n-t}\right)^2}{n-t-1}}$	x_i= individual BG value n= number of observations t= past hours
GRADE	$GRADE = \frac{\sum^{n}\left(425 * \left(\left(\log_{10}\left(\log_{10}(x_i[\frac{mmol}{l}])\right)\right) + 0.16\right)^2\right)}{n}$	x_i= individual BG value n= number of observations

The performance comparison for three-staged and five-staged architecture is outlined in Fig. 11. For the proposed three-staged model, the highest accuracy of 88.65% has been achieved. However, for the five-staged model, the accuracy is dropped to 83.41%. The overall sensitivity of 89.54% has been observed for the three-staged approach. However, the overall sensitivity of the developed five-staged model is significantly lower than the three-staged model. The

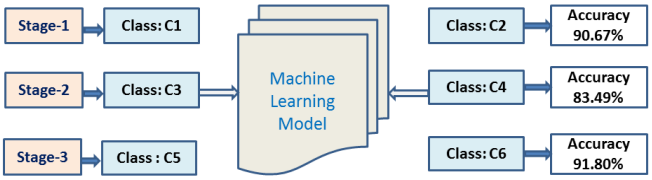


Fig. 10. HbA1c classification result for the proposed three-staged MSMC model.

TABLE VI
HbA1c CLASSIFICATION RESULT FOR **THREE-STAGED** MSMC MODEL

Stage	Model	Sen (%)	Spec (%)	AUC
1	Ensemble	91.48	90.29	92.15
2	SVM	85.71	81.96	86.23
3	A1DE	91.43	92.31	93.78
Overall		89.54	88.19	90.72

Sen=Sensitivity, Spec=Specificity.

TABLE VII
HbA1c CLASSIFICATION RESULT FOR **FIVE-STAGED** MSMC MODEL

Stage	Model	Acc (%)	Sen (%)	Spec (%)	AUC
1	RF	90	88	90.4	92.37
2	SVM	84.8	86.36	84.67	90.25
3	A1DE	82.52	85	81.92	87.43
4	NB	80.72	81.82	80.32	84.12
5	Ensemble	79	80	78.05	83.45
Overall		83.41	84.24	83.07	87.52

Acc=Accuracy, Sen=Sensitivity, Spec=Specificity.

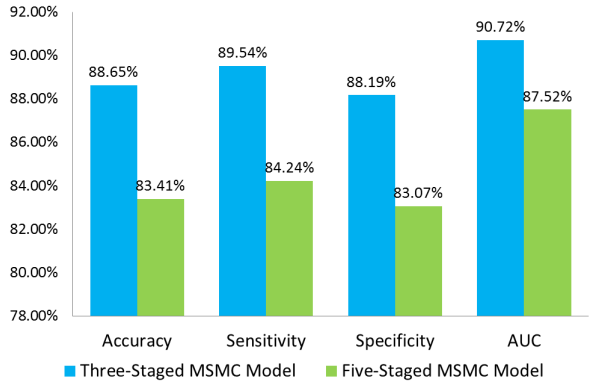


Fig. 11. The performances comparison for three-staged and five-staged model.

specificity and AUC score results for the three-staged model are significantly higher than the five-staged model. This is because the finer grain the classification becomes, the more complex the prediction becomes as HbA1C is highly variable an accurate and precise prediction is very complex.

There was a tradeoff between the number of stages of the MSMC model and its performance. It was observed that the overall performance of the three-staged MSMC model was significantly better as compared to the five-staged model. The discrepancy in the performance is that the five-staged model predicted smaller HbA1c ranges (margin, $\sim$ 1%). Contrarily, the margin of HbA1c ranges for the three-staged model was higher ($\sim$ 2%). To correctly predict smaller HbA1c ranges, the five-staged model compromised its performance from overall accuracy of 88.65% to 83.41%. Longer term prediction of HbA1c is a challenging task as it depends on the subjects' lifestyle and biological factors [26]. The work done in this research aims to predict the HbA1c of a user 2-3 month in advance. This is the first time this concept is investigated in the literature. All previous works have attempted to only estimate

TABLE VIII
COMPARISON OF LITERATURE ON HbA1c ESTIMATION

Study	Data, Model	Feature	Prediction category	R^2
[10]	SMBG, DCCT 1441 instances	μ_{PG}	Current estimation	0.82
[11]	SMBG, ADAG 507 instances	μ_{PG}	Current estimation	0.84
[12]	SMBG, TIR 1137 instances	μ_{PG}	Current estimation	0.71
[13]	SBGM, DNN 1543 instances	–	Current estimation	0.71
Ours	CGM, 150 Subjects 2225 instances	FD,TIR, GV, WD PSD	advanced prediction 2–3 months	Accuracy 88.65%

the current and instantaneous HbA1c. The estimated HbA1c levels are sometimes way off from the actual HbA1c values. This estimation, with large deviation, may often misguided healthcare professionals while taking necessary preventive interventions. However, predicting accurate HbA1c values into a specific range, such as between 7.5% and 9% as an example, appears to be more beneficial for diabetes management [27]. In the literature, advanced estimation of HbA1c values utilizing CGM data were not investigated. The studies [10], [11] as outlined in Table VIII calculated the present HbA1c values using the current BG data. The HbA1c values are significantly related with recent BG values as compared to the previous values. Furthermore, the extraction of pertinent features utilizing CGM sensor data to forecast HbA1c haven't been explored. This is the first time in literature that HbA1c *prediction* is attempted by applying an MSMC classification framework. The missing data treatment, feature extraction, selection, and fusion, combined with the MSMC framework, obtained an overall accuracy of 88.65% and 83.41% for the three-staged and five-staged classification tasks, respectively. The developed framework has an excellent perspective for both doctors and patients to arrange preemptive actions as they are now well-informed of a person's future HbA1c levels and infer the possibility of developing diabetes-related difficulties. The interventions or treatment can be started early to avoid complications and prolong healthier living.

CONCLUSION

The present research work devised a novel approach for the prediction of HbA1c levels. The prediction (not current estimation) of HbA1C levels has never been investigated in the literature despite the significance it holds. The proposed model is composed of data collection using CGM sensor, new methods for missing data estimation, seven feature extraction techniques have been utilized to extract representative features, then implementing an MSMC model for the advanced prediction task. The developed framework achieved 88.65% and 83.41% accuracy for the three-staged and five-staged classification models, respectively. One of the challenges we faced during model evaluation is the lack of a publicly available CGM dataset to test our model's applicability for HbA1c prediction. The public CGM datasets from Diabetes Research in Children Network (DirecNet) have limited HbA1c levels (6.7–9%), while the developed MSMC models classify HbA1c in the range 5.2–14.5%. This limitation is addressed by

embarking on a data collection initiative from Qatar's national children's hospital for collection of CGM diabetic children.

ACKNOWLEDGMENT

The authors would like to thank the Qatar National Library for providing Open Access publication fee.

REFERENCES

[1] K. G. M. M. Alberti, P. Z. Zimmet, and W. Consultation, "Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: Diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation," *Diabetic Med.*, vol. 15, no. 7, pp. 539–553, Jul. 1998.

[2] V. S. Freeman, "Glucose and hemoglobin A1C," *Lab. Med.*, vol. 45, no. 1, pp. e21–e24, Feb. 2014.

[3] S. I. Sherwani, H. A. Khan, A. Ekhzaimy, A. Masood, and M. K. Sakharkar, "Significance of HbA_{1c} test in diagnosis and prognosis of diabetic patients," *Biomarker Insights*, vol. 11, Jan. 2016, Art. no. BMI.S38440.

[4] E. W. Gregg, N. Sattar, and M. K. Ali, "The changing face of diabetes complications," *Lancet Diabetes Endocrinol.*, vol. 4, no. 6, pp. 537–547, Jun. 2016.

[5] L. Agrawal *et al.*, "Long-term follow-up of intensive glycaemic control on renal outcomes in the veterans affairs diabetes trial (VADT)," *Diabetologia*, vol. 61, no. 2, pp. 295–299, Feb. 2018.

[6] A. Murray *et al.*, "Action to control cardiovascular risk in diabetes follow-on memory in diabetes (accordion mind) investigators. Accordion mind: Results of the observational extension of the accord mind randomised trial," *Diabetologia*, vol. 60, no. 1, pp. 69–80, 2017.

[7] M. Sakurai *et al.*, "HbA_{1c} and the risks for all-cause and cardiovascular mortality in the general Japanese population: NIPPON DATA90," *Diabetes Care*, vol. 36, no. 11, pp. 3759–3765, Nov. 2013.

[8] N. Brewer *et al.*, "A New Zealand linkage study examining the associations between A1C concentration and mortality," *Diabetes Care*, vol. 31, no. 6, pp. 1144–1149, Jun. 2008.

[9] D. Control *et al.*, "The relationship of glycemic exposure (HbA_{1c}) to the risk of development and progression of retinopathy in the diabetes control and complications trial," *Diabetes*, vol. 44, no. 8, pp. 968–983, 1995.

[10] C. L. Rohlfing, H.-M. Wiedmeyer, R. R. Little, J. D. England, A. Tennill, and D. E. Goldstein, "Defining the relationship between plasma glucose and HbA_{1c}: Analysis of glucose profiles and HbA_{1c} in the diabetes control and complications trial," *Diabetes Care*, vol. 25, no. 2, pp. 275–278, Feb. 2002.

[11] D. M. Nathan *et al.*, "Translating the A1C assay into estimated average glucose values," *Diabetes Care*, vol. 31, no. 8, pp. 1473–1478, Aug. 2008.

[12] R. A. Vigersky and C. Mcmahon, "The relationship of hemoglobin A1C to time-in-range in patients with diabetes," *Diabetes Technol. Therapeutics*, vol. 21, no. 2, pp. 81–85, Feb. 2019.

[13] A. Zaitcev, M. R. Eissa, Z. Hui, T. Good, J. Elliott, and M. Benaissa, "A deep neural network application for improved prediction of HbA_{1c} in type 1 diabetes," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 10, pp. 2932–2941, Oct. 2020.

[14] Z. Alhassan, D. Budgen, A. Alessa, R. Alshammari, T. Daghstani, and N. Al Moubayed, "Collaborative denoising autoencoder for high glycated haemoglobin prediction," in *Proc. Int. Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2019, pp. 338–350.

[15] M. R. Cowie *et al.*, "Electronic health records to facilitate clinical research," *Clin. Res. Cardiol.*, vol. 106, no. 1, pp. 1–9, 2017.

[16] R. R. Simon, V. Marks, A. R. Leeds, and J. W. Anderson, "A comprehensive review of oral glucosamine use and effects on glucose metabolism in normal and diabetic individuals," *Diabetes/Metabolism Res. Rev.*, vol. 27, no. 1, pp. 14–27, Jan. 2011.

[17] A. Atangana and S. B. Belhaouari, "Solving partial differential equation with space- and time-fractional derivatives via homotopy decomposition method," *Math. Problems Eng.*, vol. 2013, Nov. 2013, Art. no. 318590.

[18] R. W. Beck *et al.*, "The relationships between time in range, hyperglycemia metrics, and HbA_{1c}," *J. Diabetes Sci. Technol.*, vol. 13, no. 4, pp. 614–626, Jul. 2019.

[19] D. Czerwoniuk, W. Fendler, L. Walenciak, and W. Mlynarski, "GlyCulator: A glycemic variability calculation tool for continuous glucose monitoring data," *J. Diabetes Sci. Technol.*, vol. 5, no. 2, pp. 447–451, Mar. 2011.

[20] H. Garry, B. McGinley, E. Jones, and M. Glavin, "An evaluation of the effects of wavelet coefficient quantisation in transform based EEG compression," *Comput. Biol. Med.*, vol. 43, no. 6, pp. 661–669, Jul. 2013.

[21] R. S. Stanković and B. J. Falkowski, "The Haar wavelet transform: Its status and achievements," *Comput. Electr. Eng.*, vol. 29, no. 1, pp. 25–44, Jan. 2003.

[22] I. Steinwart and A. Christmann, *Support Vector Machines*. Berlin, Germany: Springer, 2008.

[23] G. I. Webb, J. R. Boughton, and Z. Wang, "Not so naïve Bayes: Aggregating one-dependence estimators," *Mach. Learn.*, vol. 58, no. 1, pp. 5–24, 2005.

[24] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[25] S. Brahim-Belhaouari, M. Hassan, N. Walter, and A. Bermak, "Advanced statistical metrics for gas identification system with quantification feedback," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1705–1715, Mar. 2015.

[26] L. Al-Eitan, B. Almomani, A. Nassar, B. Elsaqa, and N. Saadeh, "Metformin pharmacogenetics: Effects of SLC22A1, SLC22A2, and SLC22A3 polymorphisms on glycemic control and HbA_{1c} levels," *J. Personalized Med.*, vol. 9, no. 1, p. 17, Mar. 2019.

[27] C. Fabris, L. Heinemann, R. Beck, C. Cobelli, and B. Kovatchev, "Estimation of hemoglobin A1C from continuous glucose monitoring data in individuals with type 1 diabetes: Is time in range all we need?" *Diabetes Technol. Therapeutics*, vol. 22, no. 7, pp. 501–508, Jul. 2020.

Md Shafiqul Islam received the B.Sc. degree in electrical and electronic engineering (EEE) from the Rajshahi University of Engineering and Technology (RUET), Bangladesh, in 2011, the M.E. degree in electrical and computer engineering (ECE) from the American University of Beirut (AUB), Lebanon. He is currently pursuing the Ph.D. degree in computer science and engineering (CSE) with Hamad Bin Khalifa University (HBKU), Qatar. From 2016 to 2017, he worked as a Graduate Research Assistant (GRA) with ECE Department, AUB. His research interests are machine learning, data analytics, and signal and image processing.

Marwa Khalid Qaraqe (Member, IEEE) graduated the bachelor's *(summa cum laude)* degree from Texas A&M University, Qatar, in May 2010. She received the M.Sc. and Ph.D. degrees in electrical engineering from Texas A&M University, Texas, USA, in May 2016 and August 2012, respectively. She is currently an Assistant Professor with Hamad Bin Khalifa University, Doha, Qatar. Her current research interests lie in the area of early seizure onset detection using EEG and ECG signals. She was a recipient of the Richard E. Wing Award for Excellence in Student Research in April 2012. During her master's, she was awarded Texas A&M Universitys' Diversity Fellowship.

SamirBrahim Belhaouari (Senior Member, IEEE) received the master's degree from the Institut National Polytechnique de Toulouse, France, in 2000, and the Ph.D. degree from the Federal Polytechnic School of Lausanne-Switzerland, in 2006. He is currently an Associate Professor with the Division of Information and Communication Technologies, College of Science and Engineering, HBKU. He also holds several positions with the University of Sharjah, Innopolis University, Petronas University, and EPFL Federal Swiss School.

Goran Petrovski was an Endocrinologist and a Professor with the Medical Faculty, University Clinic of Endocrinology, Skopje, Macedonia. In 2017, he joined the Sidra Medicine, Doha. In the last 15 years, he has analyzed more than 350 000 CGM days in patients with type 1 and type 2 diabetes, pregnancy, and exercise. His research interest focuses on type 1 diabetes, insulin pump, and continuous glucose monitoring. He is a reviewer in well-known diabetes journals, including *Diabetes Care*, *Diabetes UK*, *Acta Dibatologica*, and *Diabetes Technology and Therapeutics*.