OPEN ACCESS
Journal of Multidisciplinary Applied Natural Science Vol. 5 No. 2 (2025) https://doi.org/10.47352/jmans.2774-3047.275
Check for updates
Research Article

Traumatic Physiological Vital Sign Fusion: Insight from Composite Spatial Similarity Measure Modelling

David Kwamena Mensah, Micheal Arthur Ofori*, George Otieno Orwa, and Paul Hewson

Received : November 5, 2024	Revised : April 11, 2025	Accepted : May 4, 2025	Online : May 30, 2025

Abstract
This paper develops a non-linear composite similarity-based framework for generating univariate physiological vital signs data from an input multivariate counterpart. The framework is built on mixture random variate using information provided by the inter-relationships among variables. This allows the latent one-dimensional data to be generated as a weighted linear combination of the multivariate data, providing an easy way to model the weights in terms of desirable data features of interest. Using variable specific non-linear composite similarity statistic to handle short, medium- and long-term auto-relationships, the framework provides a unified context for easy quantification and assessment of both vital sign and observation level relative relevance. With the above formulation, better calibration and indication of key vital signs in traumatic events is presented. An illustrative example using real physiological vital sign datasets on trauma and non-traumatic patients provides evidence on its utility in handling both key informative incident and non-incident vital sign-specific features, events and patterns for development of pragmatic health monitoring indicators.

Keywords: composite random variable, physiological vital sign, fusion weight, health deterioration, multivariate vital signs

1. INTRODUCTION

In the rapidly evolving field of medical diagnostics and patient monitoring, the analysis of physiological vital signs plays a crucial role in the assessment and prediction of patient health outcomes. This requires fast and accurate models with the ability to leverage information from different data sources related to health. Traditional approaches often focus on individual vital signs or use simple composite scores, potentially overlooking complex interactions between different physiological parameters. This paper introduces a composite similarity-based data fusion, with a specific application to physiological vital signs [1]-[4].

The human body's intricate systems are deeply interconnected, with vital organs working together to maintain homeostasis. This interconnectedness suggests that a more holistic approach to vital sign analysis could yield deeper insights into a patient's overall health status. Nevertheless, modelling of physiological vital sign data has been focused on variable-specific models and sometimes, with their relationship to either a set of predictors or composite scoring of multiple vital signs. The utility of composite scoring of multiple vital signs is registered in the prioritization of Intensive Care Unit (ICU) patients for urgent treatment based on Early Warning (EW) score derived from timely monitored vital signs [1][5]. It seems this convention, which has been adopted worldwide, is being widely used.

In the area of variable-specific models for vital signs, literature has witnessed notable proposals spanning empirical and Gaussian process regression methods within both the univariate and multivariate dimensions. Clifton et al. [6] applied Kernel density estimation method proposed by Scott [7] for physiological vital sign data of upper-gastrointestinal surgery patients to quantify deterioration in the health. Subjects were put into normal and abnormal groups using patient-specific admission information which aided the calibration of normal and abnormal patient models for the development of novelty scores. A standard threshold statistic was then derived from the normal model. Previous work developed univariate and multivariate Kernel density quantification schemes based on vital sign deterioration for prioritizing patients with chronic obstructive pulmonary disease for clinical review [8].

Substantiating the Gaussian process application

Table 1. Correlation structure of traumatic data.

	RR	HR	SBP	DBP	TEP	SPO2	RB	MAP
RR	1.000	0.216	-0.017	0.032	0.039	-0.154	0.022	0.123
HR	0.216	1.000	-0.050	0.049	0.039	-0.099	0.055	0.484
SBP	-0.017	-0.050	1.000	0.784	-0.003	0.020	0.021	0.665
DBP	0.032	0.049	0.784	1.000	0.012	0.030	0.001	0.898
TEP	0.039	0.039	-0.003	0.012	1.000	-0.021	-0.010	0.027
SPO2	-0.154	-0.099	0.020	0.030	-0.021	1.000	-0.564	-0.017
RB	0.022	0.055	0.021	0.001	-0.010	-0.564	1.000	0.024
MAP	0.123	0.484	0.665	0.898	0.027	-0.017	0.024	1.000

Table 2. Correlation structure of non-traumatic data.

	SBP	DBP	MAP	PP	HR
SBP	1.000	0.704	0.821	0.589	0.265
DBP	0.704	1.000	0.901	-0.159	0.269
MAP	0.821	0.901	1.000	0.116	0.275
PP	0.589	-0.159	0.116	1.000	0.061
HR	0.265	0.269	0.275	0.061	1.000

to the quantification of deterioration of vital signs in health monitoring, the following notable works can be identified. Application in the classification of ICU patients using acuity of ill-ness within the multi-task Gaussian process regression framework [6][9]-[13]. For a counterpart application within the single-task Gaussian process framework, Khalid et al. is a good candidate reference [14]. A comprehensive definition of a Gaussian process is provided in Seeger [15]. A Gaussian process (GP) as a generalization of multivariate normal distribution in infinite dimension. In line with the above definition, it is easy to understand that a Gaussian process regression (GPR) models functional data using a Gaussian process with mean and covariance functions. With this view, the perspective of multi-task Gaussian process regression becomes clear as an extension of GPR to model jointly, multiple vital signs with composite covariance defined for all related tasks. This allows for the information contained in interrelationships among the multiple vital signs to be considered if the observation times are different according to the multiple vital signs. This application breaks down when vital signs are observed over a common period time. For such data, drawing information from the interrelationships among the variables can provide baseline information for developing one-dimensional framework in which the multi-dimensional variables can be treated using their latent univariate counterpart based on composite random variable [16]. By developing a method that can effectively fuse multiple vital signs into a composite measure, we aim to capture these complex relationships and provide a more comprehensive view of a patient's physiological state. Our proposed method builds similarity-based framework for composite random variables, extending these concepts to the realm of physiological data. By leveraging empirical data recovery techniques, we address common challenges in medical data analysis, such as extreme data points, point-wisely interrelationships, varying measurement frequencies, and the need to account for individual patient baselines automatically.

2. MATERIALS AND METHODS

2.1. Data

We consider two real data examples to illustrate the utility of proposed method in handling both incident and non-incident datasets. The first data example will consider traumatic physiological vital sign data to test the applicability of the method in

calibrating traumatic events. In second example, an application to physiological vital sign data from healthy subjects will be utilized. These two examples will provide insights into the practical application of the methods in addressing public health data issues.

2.1.1. Data 1

Data 1 a traumatic vital signs data from trauma patients at the Komfo Anokye Teaching Hospital (KATH). A di-identified data of dimension 5011×12 on vital signs variables RR, HR, SBP, DBP, TEMP, SPO2, RBS, and MAP was obtained. Data pre-processing was conducted to exclude infants, babies and missing observations resulting in a cross-section data on adults in the age group 18–44, comprising a subset of dimension 4064×8.

2.1.2. Data 2

The second data is a real de-identified physiological vital signs data obtained from Biofourmis, Pte. Ltd., a Health Data Analytic Company in Singapore based on their collaborative research with Singapore Heart Foundation. The data was made up of vital signs variables Systolic blood pressure (SBP), Diastolic blood pressure (DBP), Mean arterial pressure (MAP), Pulse Rate (PP), and Heart Rate (HR) with dimension 848 by 5. The collection process followed a structured study in which healthy participants were continuously monitored over a period via wearable devices spanning varied physiological states and activities defined for the research such as sleeping, walking exercising, siting, etc.

2.2. Model

Let, physiological vital signs be denoted by $D=\left(t,y_1,y_2,\ldots,y_p\right)$, $y_j=\left(y_{1j},y_{2j},\ldots,y_{nj}\right)$, $j=1,2,\ldots,p$ with $t=\left(t_1,t_2,\ldots,t_n\right)$, as the time period common to all observations. Let D be associated with a set of covariates say $X=\left(x_1,x_2,\ldots,x_p\right)$, $x_j=\left(x_{1j},x_{2j},\ldots,x_{mj}\right)$, which is not consistent with $t, m\neq n$ or otherwise $(m=n)$. Suppose there exists interrelationships among the *ys* as well as the *X*s. The application of mainstream statistical methods becomes non-trivial due to the inherent data issues, especially when computational savings is of essence. In this case, we propose a modelling framework in which the multivariate data can be fused into univariate data based on the interrelationships among the defining variables. This framework is built on the concept of mixture (composite) random variable. We illustrate the nature of the frameworks before proposing the models. The framework for the case where $m\neq n$ becomes

$$\begin{aligned}&\left(t,y_1,y_2,\ldots,y_p\right)\Leftrightarrow\left(t,\tilde{y}\right)\\&\left(X_1,X_2,\ldots,X_p\right)\Leftrightarrow\left(\tilde{X}\right)\end{aligned}\qquad(1)$$

and

$$\left(t,y_1,y_2,\ldots,y_p,X_1,X_2,\ldots,X_p\right)\Leftrightarrow\left(t,\tilde{y},\tilde{X}\right)\qquad(2)$$

In Eq. (1), t and y are consistent but not with X, while Eq. (2) exhibits consistency across t, y, and X. Thus, data challenges associated with Eq. (1)

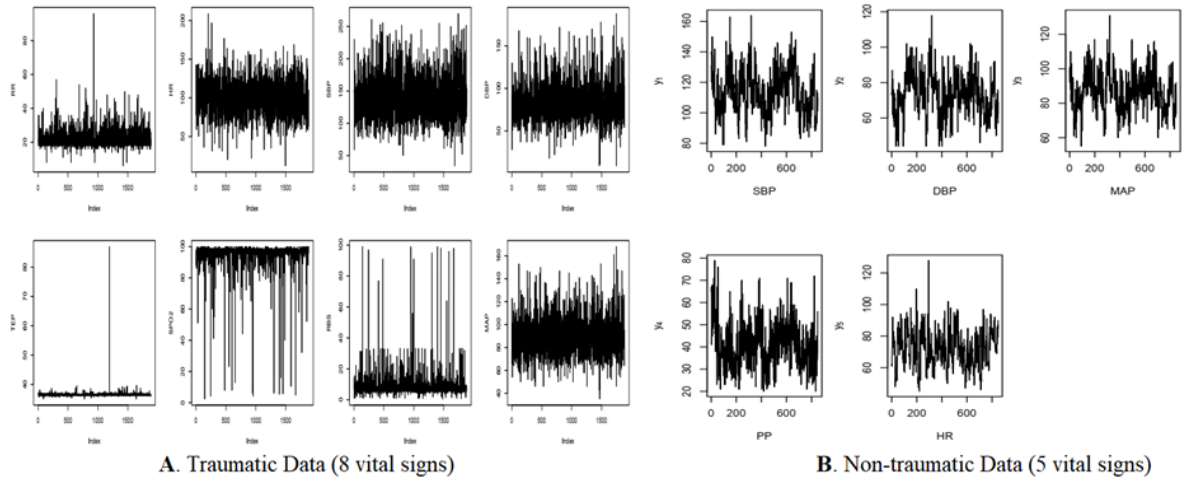

Figure 1. Data pattern of (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

Table 3. Traumatic data. Empirical values of spatial covariance parameters.

Parameter	Empirical Estimates							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
a_1	0.4426	0.3291	0.3217	0.3139	0.2599	0.4897	0.2923	0.3192
a_2	0.4885	0.3538	0.3417	0.3526	0.2557	0.4690	0.2851	0.3418
a_3	0.8125	0.5315	0.5619	0.5509	3.2094	1.6008	1.2660	0.5493
b_1	1.8416	1.6428	1.7644	1.7607	12.3985	3.3410	4.3378	1.7292
b_2	1.8230	1.6428	1.7644	1.7607	1.8009	3.3410	4.3378	1.7292
b_3	1.8230	1.6428	1.7644	1.7607	1.8009	2.9412	4.3378	1.7292

will be more than Eq. (2). Framework (1) proposes generating univariate version of $y_1, y_2, \ldots, y_p$ as $\tilde{y}$ and that of $X_1, X_2, \ldots, X_p$ as $(\tilde{x})$. It can be seen that this framework reduces $(p+1)$ dimensional data with lots of missingness created in the predictor spaces into just 3 dimensions. Modeling of $(t, \tilde{y})$ becomes relatively easy since a simple hierarchical model where $\tilde{y}$ is modeled in terms of t at the first level and second level linking $\tilde{x}$. For framework (2), there is reduction of $(2p+1)$ into just 3 dimensions as a data variable space $(t, \tilde{y}, \tilde{x})$. Thus, within this new data space, a model of the form $\tilde{y} = f(t, \tilde{x}) + \varepsilon$, where ε gives a vector of error terms. This illustrates how the framework works in ensuring computational savings. With this, the proposed framework must ensure flexibility of movement between the two data spaces as the arrow depicts. This in a way indicates the nature of potential candidate models to be considered for the development of the framework. The focus of this paper is in line with the development of the framework.

Now focusing on Eq. (1), and considering the response, and with motivation from mixture (composite) random variables, we propose a univariate data generative model $\tilde{y}$ according to model (3);

$$\tilde{y} = \theta \otimes G(y) \tag{3}$$

where θ and $G(y)$ are $n \times p$ matrices of weights and fusion statistics respectively and $\otimes$ defines an element-wise multiplication. The observations of $\tilde{y}$ are of the form:

$$\tilde{y}_i = \sum_{j=1}^{p} \theta_{ij} g(y_{ij})$$

making the entries of θ and $G(y)$ as θ_{ij} which are vital sign specific together with $g(y_{ij})$. The fusion weights are allowed to satisfy the following property $\sum_{j=1}^{p} \theta_{ij} = 1$. This ensures that the resulting random variable is valid or follows a valid probability distribution by statistical theory.

2.3. Spatial Composite Similarity Model for θ

Adopting the Gaussian process concept of modelling and applying the idea of covariance function, we can model θ in terms of the point-wise spatial inter-relationships existing within variables to capture all levels of auto-relationships. This allows easy quantification of short-, medium-, and long-term trends present in the data. This is where the flexibility of Gaussian process covariance functions can be explored [15]. This property is well exhibited in the case of compound covariance functions. Based on the above information, we model θ_{ij} as the underlying variable specific spatial relationships via a composite similarity function, define point-wisely $k(a_1, a_2, a_3, b_1, b_2, b_3, \delta_{(1,2)})$ as for y_{11} and y_{21} :

$$k(a_1, a_2, a_3, b_1, b_2, b_3, \delta_{(1,2)}) = k_1(a_1, b_1, \delta_{(1,2)}) + k_2(a_2, b_2, \delta_{(1,2)}) + k_3(a_3, b_3, \delta_{(1,2)}) \tag{4}$$

$$k_1\left(a_1,b_1,\delta_{(1,2)}\right)=\sigma_1^2\exp\left(-b_1^2\delta_{(1,2)}^2\right)$$

$$k_2\left(a_2,b_2,\delta_{(1,2)}\right)=\sigma_2^2\left(1+\frac{\sqrt{5}\delta_{(1,2)}}{b_2}+\frac{\sqrt{5}\delta_{(1,2)}^2}{3b_2^2}\right)\times \exp\left(-\frac{\sqrt{5}\delta_{(1,2)}}{b_2}\right)$$

$$k_3\left(a_3,b_3,\delta_{(1,2)}\right)=\sigma_3^2\left(1+\frac{\sqrt{3}\delta_{(1,2)}}{b_3}\right)\exp\left(-\frac{\sqrt{3}\delta_{(1,2)}}{b_3}\right),$$

with $\delta_{(1,2)}=|y_{11}-y_{21}|$, with $a_1,a_2,a_3,b_1,b_2,b_3>0$. we adopted the deterministic treatment approach for the parameters of similarity measure (4), by relating them to key features of the underlying probability density functions (pdfs), $f(y_j)$ of the variables. Particularly, they were estimated using functions of moments of $f(y_j),j=1,2,..,p.$ The kth non-central moment of a random variable y is defined as:

$$E\left[Y^k\right]=\int y^k g(y)dy \quad \textbf{(5)}$$

where $g(y)$ is the pdf of y. Let also define the statistic $D(y)$ as below:

$$D(y)=\sqrt{y^k g(y)} \quad \textbf{(6)}$$

Setting $k = 2$ to ensure only positive statistics to be consistent with parameter restrictions, we compute b_1,b_2,b_3 as:

$$b_1=\overline{D(y)},\ b_2=\tilde{D}(y),\ b_3=\delta_2-\delta_1$$

where $\overline{D(y)}$, $\tilde{D}(y)$, δ_1 and δ_2 are the mean, median, minimum and maximum of $D(y)$ respectively. The motivation for using $D(y)$ is its capability to assess the contributions of y values to a common measure of center, scaling them accordingly with respect to their proximity to the center. The $g(y)$ is taken to be the kernel density estimate defined point-wisely as:

$$\hat{g}(y_l)=\frac{1}{nh}\sum_{l=1}^{n}K\left(\frac{y_l-y}{h}\right), \quad \textbf{(7)}$$

where $K\left(\frac{y_i-y_j}{h}\right)$ is symmetric kernel with smoothing parameter h and $y_1,y_2,\ldots,y_n$ gives the observation on y [7]. On the other hand, a_1,a_2,a_3 are set to be the first three smallest order statistics of $\delta_r=\frac{D(y)}{d_m}$, $d_m=\max(D(y))$.

2.4. Statistics for θ and g(y)

By Eq. (4), each $y_j, j=1,2,\ldots,p$generates an $n \times n$ similarity matrix which needs to be summarized into a vector to aid the generation of θ utilized in model (3). We consider the mean, median and Orthogonalized Gnanadesikan-Ketterning (OGK) statistics in this regard [16][17]. Physiological vital sign data often contains outliers and may follow non-normal distributions. Mean weights are optimal for normally distributed data, median weights are more robust against outliers, and OGK weights can offer multiple protection against outliers than its counterparts namely mean and median with the nature of statistics (Eq. (4)) adopted in this paper. They are computed as follows;

$$\bar{\theta}=\left(\bar{\theta}_1,\bar{\theta}_2,\ldots,\bar{\theta}_p\right),\ \tilde{\theta}=\left(\tilde{\theta}_1,\tilde{\theta}_2,\ldots,\tilde{\theta}_p\right),\ \theta_{OGK}=\left(\theta_1^o,\theta_2^o,\ldots,\theta_p^o\right) \quad \textbf{(8)}$$

Table 4: Non-traumatic data. Empirical values of spatial covariance parameters.

Parameter	Empirical Estimates				
	y_1	y_2	y_3	y_4	y_5
a_1	0.3400	0.3320	0.3320	0.3399	0.3502
a_2	0.3646	0.3640	0.3424	0.3654	0.3856
a_3	0.5712	0.5277	0.5264	0.5674	0.5870
b_1	1.7016	1.6302	1.6630	1.7438	1.7254
b_2	1.7016	1.6320	1.630	1.7438	1.7254
b_3	1.7016	1.6320	1.630	1.7438	1.7254

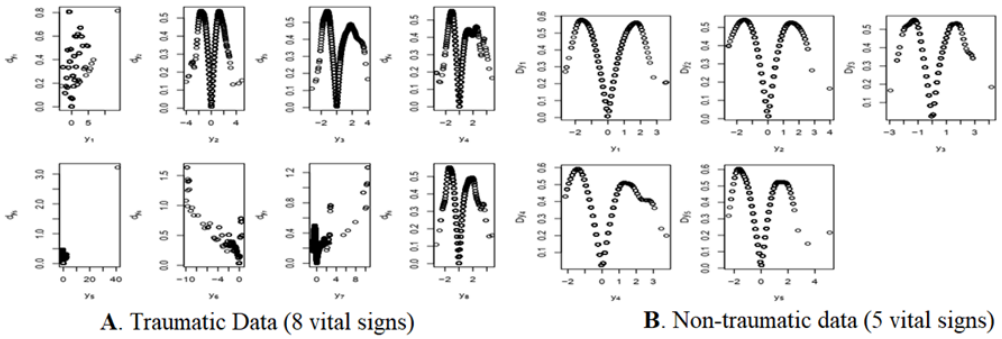

Figure 2. Nature of the fusion statistics $D(y)$ for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

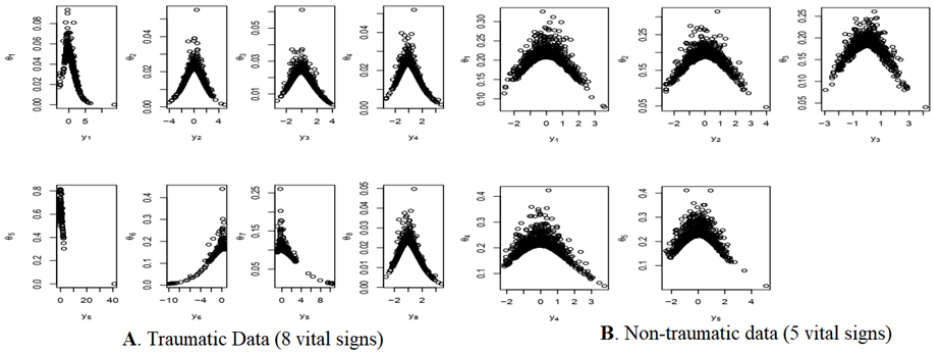

Figure 3. Original data: Nature of fusion weight θ based on mean statistic for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

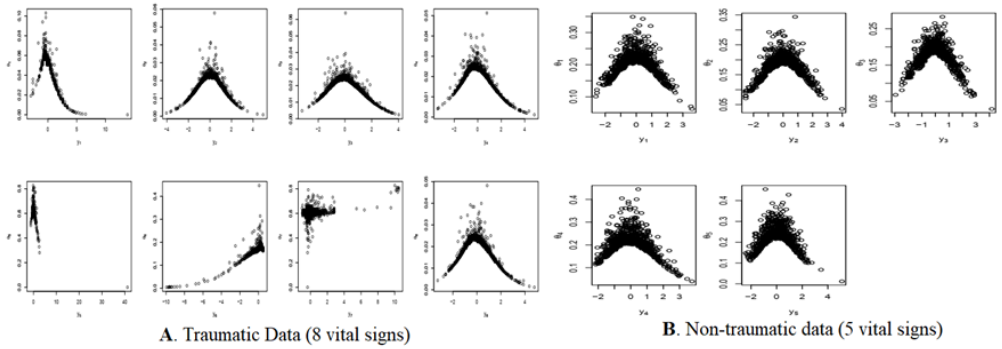

Figure 4. Original data: Nature of fusion weight θ based on median statistic for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

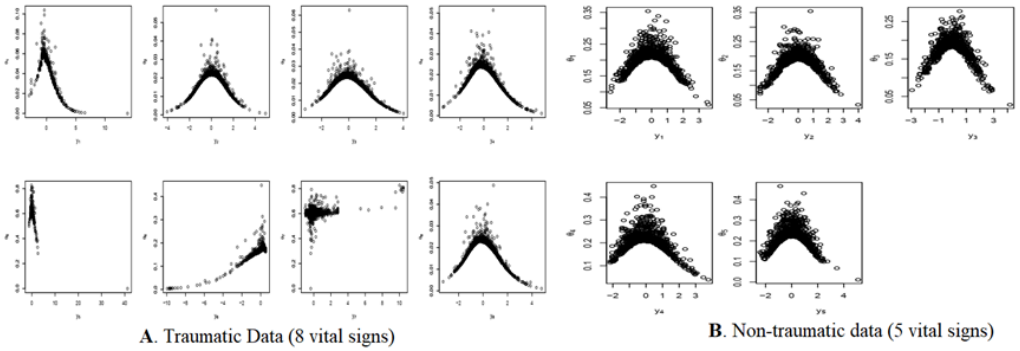

Figure 5. Original data: Nature of fusion weight θ based on OGK statistic for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

where $\bar{\theta}_j = (\bar{\theta}_{1j},\ldots,\bar{\theta}_{nj})$, $\tilde{\theta} = (\tilde{\theta}_{1j},\ldots\tilde{\theta}_{nj})$, $j = 1,2,\ldots,p$ with

$$\bar{\theta}_{1j} = \frac{1}{n}\sum_{i=2}^{n} k(a_1,a_2,a_3,b_1,b_2,b_3,\delta_{(1j)})$$

$$\tilde{\theta}_{1j} = \text{median}\begin{pmatrix} k(a_1,a_2,a_3,b_1,b_2,b_3,\delta_{(2j)}),\ldots, \\ k(a_1,a_2,a_3,b_1,b_2,b_3,\delta_{(nj)}) \end{pmatrix}$$

and

$$\theta_{OGK} = \sum_{i=1}^{n} c_i v(w_i)\left(\sum_{i=1}^{n} v(w_i)\right)^{-1}$$

$$w_i = \frac{y_i - \tilde{\mu}_0}{\tilde{\sigma}_0}$$

$$v(w_i) = \left[1 - G^2(w_i,b)^2 I_{(|w_i|\le b)}\right]$$

$$G^2(w_i,b^*) = \frac{w_i}{b^*}$$

where $b^* = 4.5$, $\tilde{\mu}_0$ and $\tilde{\sigma}_0$ denote the median and median absolute deviation (MAD) of $c_i = \theta_i$. respectively. We considered y_{ij} as the fusion statistics for $g(y_{ij})$ in model (3).

2.5. Vital Sign Specific Contributions in $\tilde{y}$

Treating $\tilde{y}$ as the common data pattern presented by the entire *p*-dimensional data, both variable and observation level contributions in $\tilde{y}$ can be quantified and assessed using simple statistics. This helps to assess fusion components and quantify their relevance to the common pattern $\tilde{y}$. Define vital sign specific and total contribution statistics for y_j as follows:

$$w_j = \frac{\tilde{y}}{\hat{y}_j} \quad \textbf{(9)}$$

$$\bar{w}_j = \frac{\sum_{i=1}^{n_j} w_{ij}}{n_j}, j = 1,2,\ldots,p \quad \textbf{(10)}$$

where $\hat{y}_j$ denotes the *j*th component of $\hat{y}$. If we define the within vital sign specific contribution statistics as ϑ_{ij}, then the relevance of *i*th observation in the *j*th vital sign can be written as follows.

$$\vartheta_{ij} = \frac{\tilde{y}_j}{y_{ij}}, \quad i = 1,2,\ldots,n;\ j = 1,2,\ldots,p. \quad \textbf{(11)}$$

2.5.1. Utility of Contribution Statistics

These statistics can be put to many useful uses such as pragmatic monitoring and assessment of influential observations by examining contribution to change-points in data pattern, which will eventually manifest in change-point in health condition such as deterioration. This can provide information on crucial vital signs to consider during traumatic events. They can be automated and put to a similar use as in the Early Warning Scores track and score approach to vital sign monitoring [1][5]. In the case of mean contribution statistic, its utility is seen in quantification of the overall impact a vital sign has with a measure of the entire vital signs considered in the fusion, defined in terms of either health, deterioration, indicator of emergency in the case of traumatic events etc.

2.6. Implementation

The methods were fully implemented in the R statistical software. Codes were scripted using *TinnR* and executed in R. The empirical density *g* (*y*) was estimated using the kernel smoothing package, "ks" [8]. We considered the smoothed cross-validation estimator for h_i. The execution of scripted algorithms was done using the Intel (R) Core (TM) *i*7, 6700 processor Windows PC 3.40 GHz workstation. The vital sign variables $y_1, y_2, \ldots y_n$ were standardized to remove the variable units of measure and allow them to have a common reference family of sampling distributions. That is for the *j*th variable, its standardized version, z_j is defined as below.

$$z_j = \frac{y_j - \mu_j}{\sigma_j}, j = 1,2,\ldots,8. \quad \textbf{(12)}$$

2.7. Performance Assessment

We define the following statistical measures of fit for the assessment of fusion performance, mean squared fusion error (MSFE), mean absolute fusion error (MAFE) and standard mean absolute fusion error (SMAFE);

$$\text{MSFE} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\tilde{y}_i - \tilde{y}_{cij}\right)^2}{m}$$

$$\text{MAFE} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left|\tilde{y}_i - \tilde{y}_{cij}\right|}{m}$$

$$\text{SMAFE} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|Q(\tilde{y},\tilde{y}_c)\right|}{m}, Q(\tilde{y},\tilde{y}_c) = \frac{\tilde{y}_i - y_{cij}}{\tilde{y}_i},$$

where $\tilde{y}$ and $\tilde{y}_c$ define the fused and variable specific components of $\tilde{y}$ and $m = np$ gives the total sample size.

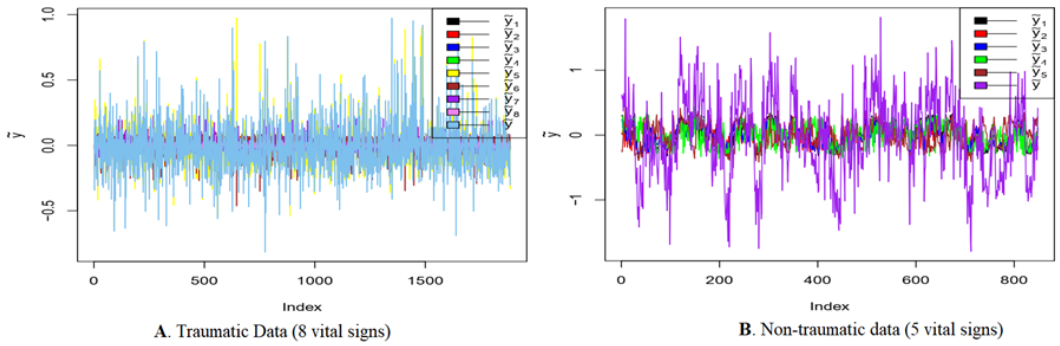

Figure 6. Characteristics of fused vital signs using mean weights for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

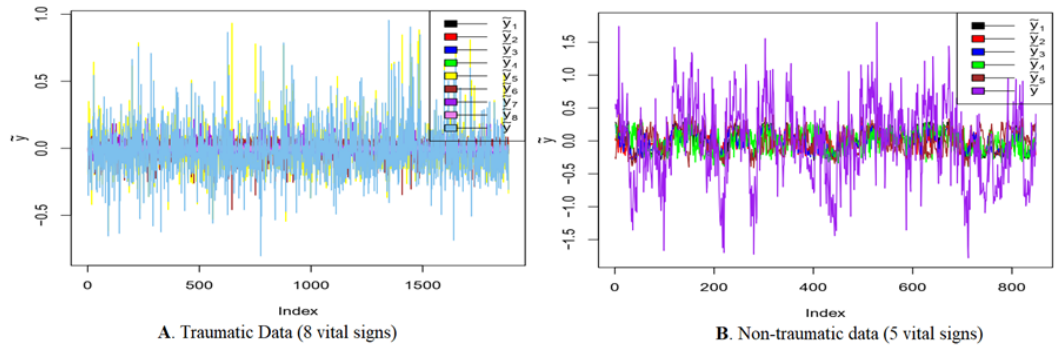

Figure 7. Characteristics of fused vital signs using median weights for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

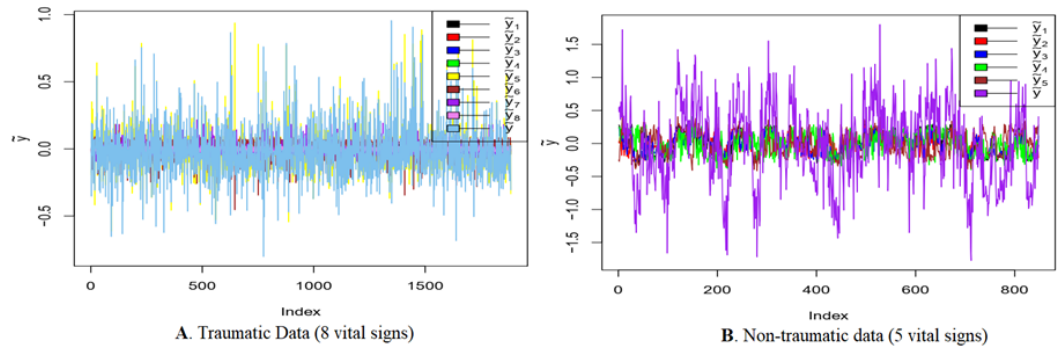

Figure 8. Characteristics of fused vital signs using OGK weights for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

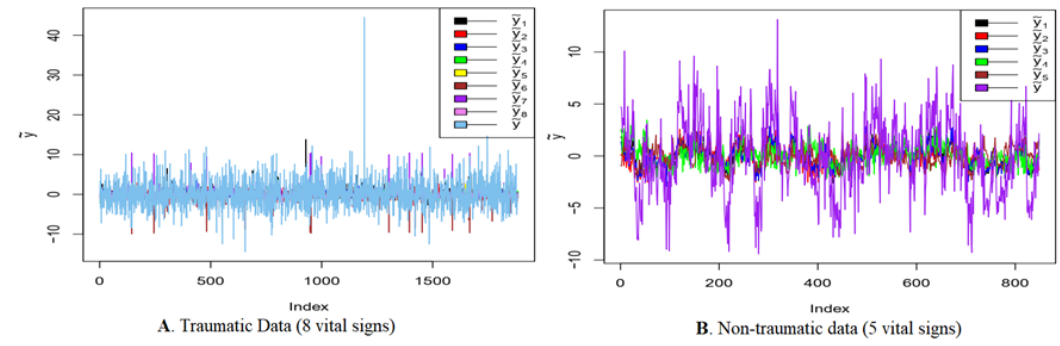

Figure 9. Characteristics of fused vital signs using common weights for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

3. RESULTS AND DISCUSSIONS

We examine the evaluation of our proposed composite similarity-based fusion method for physiological vital signs. We look at the method's performance on real patient data using the empirical values of spatial covariance parameters. First, we examine the nature of the underpinning assumption of the proposed methods for the datasets for the illustrative examples. Tables 1 and 2 report the pair-wise interrelationships existing among the defining variables of the datasets 1 and 2, quantified using the Peasrson Correlation measure respectively. It can be seen that physiological vital signs variables are naturally interrelated at varied degrees in magnitude and direction. In particular, the traumatic vital signs variables exhibit relatively low interrelationships than its non-traumatic counterpart data. This suggest that data fusion methods for such datasets should be designed to inherit the natural underlying features to ensure its practicality in real life applications.

Next, we examine the nature of data generated by the variables of the two datasets. Figure 1 shows the plot of the vital signs for both traumatic data A (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and Non-traumatic data B (SBP, DBP, MAP, PP, and HR).

The differences in patterns underlying the vital sign variables of trauma patients are evident, with RR and HR exhibiting variant trends in comparison with the rest. Also, within the non-trauma data, PP and SBP exhibit different trends in comparison with the others. It is expected to obtain similar feature patterns if the data is rich across the vital sign variables.

Table 3 and Table 4 report the empirical estimates of the parameters of the spatial similarity measure defined for quantifying the underlying interrelationships among the vital sign variables for traumatic and non-traumatic datasets respectively. Clearly, there exists differences among the estimates obtained for the two datasets. Also, estimates for $a_1, a_2,$ and a_3 exhibit variability across and within vital sign variables. In the case of $b_1, b_2,$ and $b_3,$ the variability is only seen across vital sign variables. This evident for both the traumatic and non-traumatic datasets.

Figure 2 shows the nature of the $D(y)$ statistics obtained from the vital signs data. From left to right for A are statistics corresponding to RR, HR, SBP, DBP, RBS, TEP, SPO_2, and MAP. Also, for B, are statistics corresponding to SBP, DBP, MAP, PP, and HR. The differences in pattern underlying the vital sign variables of the trauma patients are evident, with HR, SBP and HR exhibiting variant trends in comparison with the rest. The non-trauma data exhibits similar feature patterns with minimal deviations. It is expected to obtain similar feature patterns with vital sign specific data, if the data is similar across all the vital sign variables.

The typical relationships that exist between the fusion weights θ and the vital sign observations are illustrated in Figures 3 – 5 for θ based on the mean, median and OGK weighting statistics respectively. This was done for both the traumatic and non-traumatic data. Apparently, a common general trend is exhibited by all the weighting statistics with varying discriminating capabilities across both datasets. Particularly, the mean weighting statistics shows that there are few observations that deviate from a common pattern underlying the data. This

Table 5. Traumatic data: Vital sign specific average contribution statistics $\bar{\omega}$.

Parameter	Vital signs							
	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
$\bar{\omega}_M$	15.9381	3.4775	1.4774	-0.6401	0.9433	0.2056	-2.3609	1.0678
$\bar{\omega}_{Med}$	15.9381	3.4775	1.4774	-0.6401	0.9433	0.2056	-2.3609	1.0678
$\bar{\omega}_{OGK}$	15.9381	3.4775	1.4774	-0.6401	0.9433	0.2056	-2.3609	1.0678
$\bar{\omega}_c$	14.6590	2.9355	2.6037	-0.7206	-0.1374	-1.3325	-5.4360	4.2350

Table 6. Non-traumatic data: Vital sign specific average contribution statistics $\bar{\omega}$.

Parameter	Vital signs				
	y_1	y_2	y_3	y_4	y_5
$\bar{\omega}_M$	2.5182	2.8018	3.0090	1.6988	2.3029
$\bar{\omega}_{Med}$	2.4706	2.7056	2.9576	1.6273	2.2177
$\bar{\omega}_{OGK}$	2.7140	3.4863	3.2294	2.0694	3.2183
$\bar{\omega}_c$	2.9447	3.3114	2.9747	2.0952	3.3836

may be due to its non-resistance to extreme observations. On the hand, the median and OGK weighting statistics yield improved data pattern, for clearer identification of potential observations with issues than those reported by the mean weighting.

The vital sign-specific fused components and overall composite data pattern underlying the multivariate vital sign data are shown in Figures 6 – 8 respectively, for mean, median and OGK weighting schemes for the two illustrative data examples. The results obtained based on equal weights (natural weight of 1) for all vital signs is presented in Figure 9. There is obvious adaption of the methods to dataset with differences at varied levels of frequencies according to observed events. Also, obvious significant differences in latent data patterns are evident among those generated by the mean, median and OGK and their equal weight counterpart across both datasets. Furthermore, in terms of preservation of intrinsic functional nature of data, it can be observed that ensuring variable weighting yields improvement in localized adjustment to vital sign conditions. The ability to generate the same fusion results using the three weighting schemes is particularly of great value in public health in terms of pragmatic intervention delivery. It is of great value in the sense that there would not be any extra need for model selection which may require additional resources that might not support in the case of emergency.

Tables 5 and 6 report the overall performance statistics of the vital signs in terms of average contribution to the formation of composite vital sign from the multivariate data based on the two illustrative datasets. Contribution efforts of the vital signs are highly different across statistics considered for building the weights. This is better seen with variable weighting scheme than its common or fixed counterparts on the average. Particularly, ranking the vital signs based on the weighting scheme, we have $y_1, y_2, y_3, y_8, y_5, y_6, y_4$, and y_7. Also, ranking the vital signs based on common weighting produces $y_1, y_8, y_2, y_3, y_5, y_4, y_6$, and y_7. for the traumatic data. Furthermore, there exists consistency in calibration of contributions of vital signs based on the contribution statistics using the weighting schemes in comparison with the equal weights across vital signs. In the case of the non-traumatic data, it can be observed that contributon efforts of vital signs exhibits variation across weighting schemes both within and between variables. There seems to be clearer illustration of consistency in calibration reported by the weighting schemes in terms of ordering the associated vital signs in comparison with their equal weight counterpart. This inconsistencies can lead to issues in identication of crucial vital signs in the provision of pragmatic intervesions in the case of ermegencies.

The effect of vital sign specific derivatives of fused components on the fused vital sign based on mean, median, OGK, and common weight (weights = 1) weighting schemes are shown in Figures 10 – 13 respectively. The impact of variable weights-observation level in comparison with common weights, is clear in the key differences in data pattern exhibited in the above figures from both data perspective. Also, the same trend showing across the mean, median and OGK weighting schemes is a reflection of the authenticity of the true information that underlies the two multivariate vital sign datasets.

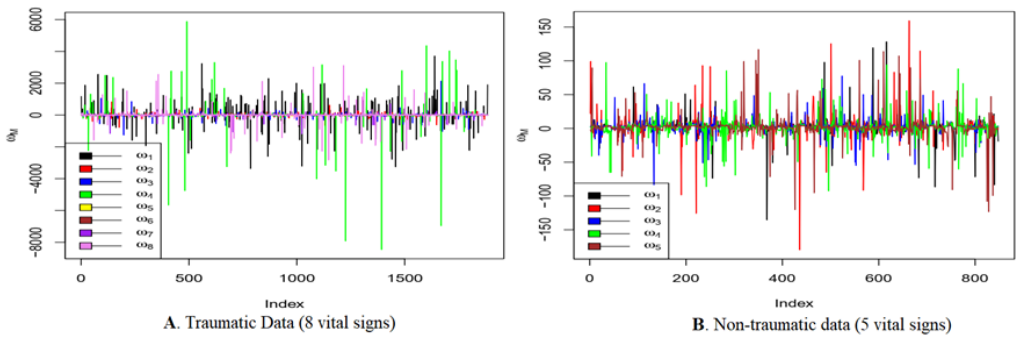

Figure 10. Vital sign fusion contribution based on mean weighting for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

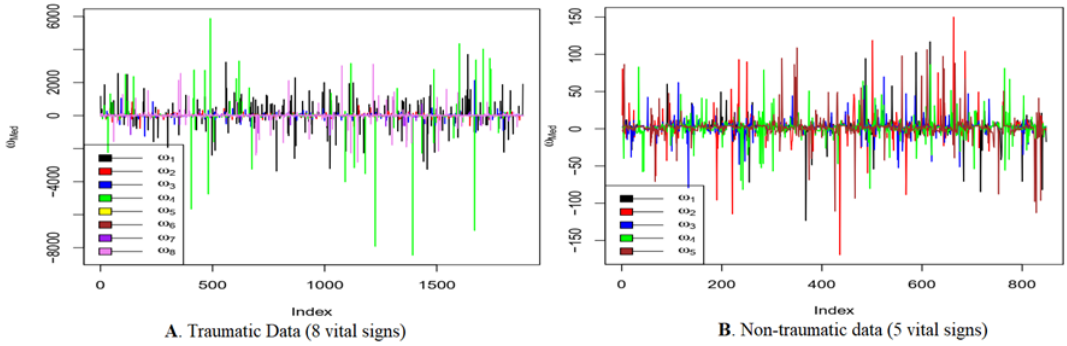

Figure 11. Vital sign fusion contribution based on median weighting for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

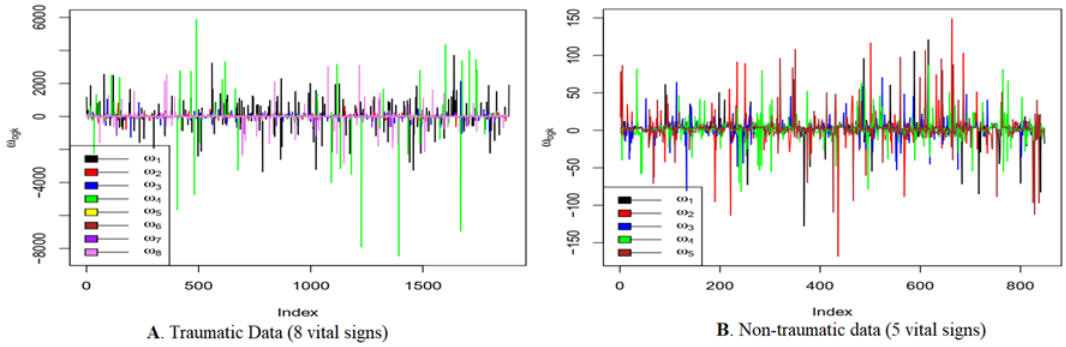

Figure 12. Vital sign fusion contribution based on OGK weighting for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

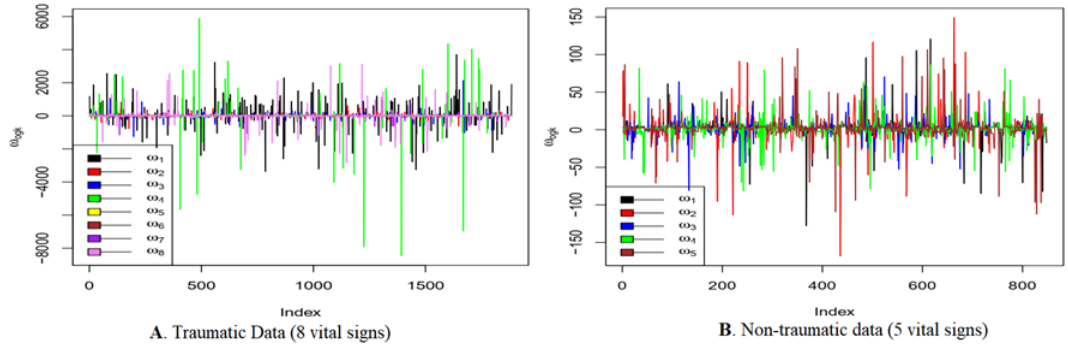

Figure 13. Vital sign fusion contribution based common unit weight for (A) traumatic data (RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP) and (B) non-traumatic data (SBP, DBP, MAP, PP, and HR).

Figure 14 show the vital sign observation level fusion contributions in vital sign fusion component with Median. The fifth vital sign, temperature (yellow line) shows the highest contribution consistently across all three fusion statistics methods in the vital sign observation level fusion contributions in vital sign fusion component. Physiologically, temperature regulation is essential for proper enzyme function and metabolic processes with it changes significantly affecting heart rate, breathing rate, and overall cellular function. Also, the sixth vital sign, oxygen saturation (SPO2) (brown line) appears as the second most influential parameter, with values typically around 0.15–0.2 and occasional spikes reaching 0.40– 0.45 measures the percentage of hemoglobin in your blood that's carrying oxygen with normal range from 95–100%. Below 90% is considered low and may indicate hypoxemia (low blood oxygen). SpO2 monitors how efficiently your lungs transfer oxygen to your blood and how well your circulatory system delivers that oxygen throughout your body. Lastly, vital sign seven, random blood sugar (RBS) (purple line) shows consistent moderate contribution around 0.1, positioning it as the third most influential vital sign. RBS measures the blood glucose level at any random time, regardless of when you last ate. The normal range is less than 140 mg/dL (7.8 mmol/L) with higher values indicating prediabetes or diabetes RBS Helps assess glucose metabolism and screen for diabetes. Clinically, there is the need for prioritize monitoring. The analysis suggests that clinicians should prioritize monitoring temperature and SpO2 as they contribute most significantly to the overall physiological assessment of trauma patients. The minimal contribution of several vital signs suggests that in resource-constrained environments, monitoring efforts could potentially be streamlined to focus on the most informative parameters.

Tables 7 and 8 compare the fusion error statistics for different weighting methods namely the Mean, Median, and OGK weights as well as a common weighting for both traumatic and non-traumatic datasets. The error statistics are reported using MSFE, MAFE and SMAFE. Interestingly, the Mean, Median, and OGK weights all produce identical error statistics, while the common weight produces much higher errors, particularly for MSFE for the traumatic vital sign data. In the case of the non-traumatic vital sign data, there exists variability with across error measure and weighting scheme. It can be seen that three weighting schemes tends to overperform their common weight counterpart in MSFE and MAFE. Interestingly, it can be observed that MSFE and MAFE seem to support vital sign data fusion as overall fitting error statistics than SMAFE.

In general, it can be seen that the proposed method demonstrates a strong ability to fuse multiple physiological vital signs into a single composite measure. This is evidenced by the low error statistics (MSFE, MAFE, SMAFE) for the mean, median, and OGK weighting methods shown in Tables 7 and 8. The fact that these three weighting methods produce identical error statistics

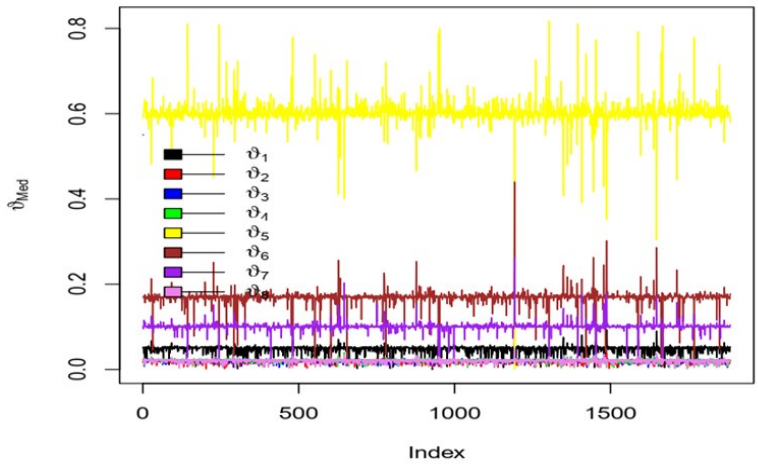

Figure 14. Vital sign observation level fusion contribution in fused component based on median statistic (Theta 1 to 8 are RR, HR, SBP, DBP, TEP, SPO2, RBS, and MAP).

Table 7. Traumatic data, Fusion error statistics across varied weights.

Error statistic	Fusion Statistics			
	Mean	Median	OGK	Common Weight
MSFE	0.03143	0.03143	0.03143	10.90361
MAFE	0.12953	0.12953	0.12953	2.46013
SMAFE	2.03833	2.03833	2.03833	1.26431

Table 8. Non-traumatic data, Fusion error statistics across varied weights.

Error statistic	Fusion Statistics			
	Mean	Median	OGK	Common Weight
MSFE	0.25199	0.23503	0.34250	8.60034
MAFE	0.40385	0.38980	0.46810	2.33706
SMAFE	2.15498	5.99621	2.26192	2.10886

suggests a robust fusion process that is not overly sensitive to the choice of weighting method [18]-[20]. The similar performance of mean, median, and OGK weighting methods for the traumatic data suggests that the fusion approach is flexible and can adapt to different data characteristics relating to traumatic events. This could be particularly valuable when dealing with physiological data that may have outliers or non-normal distributions [21]-[23]. The relevance of observations within a given vital sign to a component can be quantified, so that their impact on data specific events (e.g. traumatic events) can be assessed. These statistics can be put to many useful uses such as pragmatic monitoring and assessment of influential observations by examining contribution to change-points in data pattern, which will eventually manifest in change-point in health condition such as deterioration [24]. This can provide information on crucial vital signs to consider during traumatic events. They can be automated and put to a similar use as in the EW scores track and score approach to vital sign monitoring [1][5]. In the case of mean contribution statistic, its utility is seen in quantification of the overall impact a vital sign has with a measure of the entire vital signs considered in the fusion, defined in terms of either health, deterioration, indicator of emergency in the case of traumatic events, and so on.

4. CONCLUSIONS

In this paper, a novel framework for calibrating the relative significance of physiological vital signs has been proposed and implemented. The framework creates a one-dimension data out of an input multivariate physiological vital sign data based on the theory of mixture (composite) random variables. By this, the one-dimensional data is modelled as empirical mixture of underlying p dimensional data variables. Mixing weights are computed using non-linear composite similarity statistic to capture the available interrelationships within a variable, so that it is variable-specific. Adopting the Gaussian process nature of covariance functions, the composite similarity statistic is defined to handle short-, medium- and long-term auto relationships, providing a unified context for easy quantification and assessment of both vital sign and observation level relative relevance. This yields a better calibration and indication of key vital signs in traumatic and non-traumatic situations. An implementation of the framework using real physiological vital sign data of both trauma and non-trauma, using both common and variable-specific weights illustrates its utility in ensuring improved quantification and calibration of relevance of vital signs in traumatic conditions. Thus, providing informed indication on crucial vital signs for pragmatic planning of treatments of traumatic and non-traumatic patients.

AUTHOR INFORMATION

Corresponding Author

Micheal Arthur Ofori — Department of Mathematical Sciences, Pan African University Institute for Basic Sciences, Technology and Innovation, Juja-62000 (Kenya);
orcid.org/0000-0002-4983-540X
Email: arthur.michael@students.jkuat.ac.ke

Authors

David Kwamena Mensah — Department of Statistics, University of Cape Coast, Cape Coast-CC1070 (Ghana);
orcid.org/0000-0003-0109-8110

George Otieno Orwa — Department of Mathematics, Jomo Kenyatta University of Agriculture and Technology, Juja-62000 (Kenya);
orcid.org/0000-0002-2093-229X

Paul Hewson — Department of Statistics, University of Exeter, Exeter-EX44PY (United Kingdom);
orcid.org/0000-0002-4990-5498

Author Contributions

All the authors contributed immensely towards this work.

Conflicts of Interest

The authors have declared that no competing interests exist.

ACKNOWLEDGEMENT

The first author is Indebted to the University of Cape Coast for support in term of facility for the conduct of research. The second author wants to thank the African Union through the Pan African University Institute for Basic Sciences, Technology, and Innovation (Pan African University Scholarship) for their support in the form of Scholarship for my studies.

REFERENCES

[1] J. Gardner-Thorpe, N. Love, J. Wrightson, S. Walsh, and N. Keeling. (2006). "The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study". *The Annals of The Royal College of Surgeons of England.* **88** (6): 571-5. 10.1308/003588406X130615.

[2] P. H. Charlton, M. Pimentel, and S. Lokhandwala. (2016). In: "Secondary Analysis of Electronic Health Records". Cham (CH): Springer. 325-38. 10.1007/978-3-319-43742-2_22.

[3] N. Skyttberg, J. Vicente, R. Chen, H. Blomqvist, and S. Koch. (2016). "How to improve vital sign data quality for use in clinical decision support systems? A qualitative study in nine Swedish emergency departments". *BMC Medical Informatics and Decision Making.* **16** : 61. 10.1186/s12911-016-0305-4.

[4] S. Wang, M. E. Celebi, Y.-D. Zhang, X. Yu, S. Lu, X. Yao, Q. Zhou, M.-G. Miguel, Y. Tian, J. M. Gorriz, and I. Tyukin. (2021). "Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges, and Prospects". *Information Fusion.* **76** : 376-421. 10.1016/j.inffus.2021.07.001.

[5] D. Evans, B. Hodgkinson, and J. Berry. (2001). "Vital signs in hospital patients: a systematic review". *International Journal of Nursing Studies.* **38** (6): 643-50. 10.1016/s0020-7489(00)00119-x.

[6] L. Clifton, D. A. Clifton, M. A. Pimentel, P. J. Watkinson, and L. Tarassenko. (2013). "Gaussian processes for personalized e-health monitoring with wearable sensors". *IEEE Transactions on Biomedical Engineering.* **60** (1): 193-7. 10.1109/TBME.2012.2208459.

[7] D. W. Scott. (2015). "Multivariate Density Estimation". 10.1002/9781118575574.

[8] C. Velardo, S. A. Shah, O. Gibson, H. Rutter, A. Farmer, and L. Tarassenko. (2014). "Automatic generation of personalised alert thresholds for patients with copd". *Signal Processing Conference.* 1990-1994.

[9] M. Ghassemi, M. Pimentel, T. Naumann, T. Brennan, D. Clifton, P. Szolovits, and M. Feng. (2015). "A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data".

Proceedings of the AAAI conference on artificial intelligence. **29** (1).

[10] B. M. Marlin, D. C. Kale, R. G. Khemani, and R. C. Wetzel. (2012). "Unsupervised pattern discovery in electronic health care data using probabilistic clustering models". *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium.* 389-398. 10.1145/2110363.2110408.

[11] T. A. Lasko. (2014). "Efficient inference of Gaussian-process-modulated renewal processes with application to medical event data". *Conference on Uncertainty in Artificial Intelligence.* **2014** : 469.

[12] P. Schulam and S. Saria. (2015). "A framework for individualizing predictions of disease trajectories by exploiting multi-resolution structure". *Advances in Neural Information Processing Systems.* **28** : 748-756.

[13] M. Lichman and P. Smyth. (2014). "Modeling human location data with mixtures of kernel densities". *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining.* 35-44. 10.1145/2623330.2623681.

[14] S. Khalid. (2014). "Data fusion models for detection of vital-sign deterioration in acutely ill patients". *University of Oxford.*

[15] M. Seeger. (2004). "Gaussian processes for machine learning". *International Journal of Neural Systems.* **14** (2): 69-106. 10.1142/S0129065704001899.

[16] D. K. Mensah, M. A. Ofori, and N. Howard. (2022). "Traumatic Systolic Blood Pressure Modeling: A Spectral Gaussian Process Regression Approach with Robust Sample Covariates". *Mathematics and Statistics.* **10** (3): 562-574. 10.13189/ms.2022.100312.

[17] D. E. Tyler. (2008). "Robust statistics: Theory and methods". 888-889.

[18] Z. Liu, H. A. Song, V. Zadorozhny, C. Faloutsos, and N. Sidiropoulos. (2017). "H-fuse: Efficient fusion of aggregated historical data". *Proceedings of the 2017 SIAM International Conference on Data Mining.* 786-794.

[19] H. Lee, K. Park, B. Lee, J. Choi, and R. Elmasri. (2008). "Issues in data fusion for healthcare monitoring". *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments.* 1-8. 10.1145/1389586.1389590.

[20] R. C. King, E. Villeneuve, R. J. White, R. S. Sherratt, W. Holderbaum, and W. S. Harwin. (2017). "Application of data fusion techniques and technologies for wearable health monitoring". *Medical Engineering & Physics.* **42** 1-12. 10.1016/j.medengphy.2016.12.011.

[21] D. M. Titterington and B. Wang. (2006). "Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model". *Bayesian Analysis.* **1** (3). 10.1214/06-ba121.

[22] D. Labate, F. L. Foresta, G. Occhiuto, F. C. Morabito, A. Lay-Ekuakille, and P. Vergallo. (2013). "Empirical Mode Decomposition vs. Wavelet Decomposition for the Extraction of Respiratory Signal From Single-Channel ECG: A Comparison". *IEEE Sensors Journal.* **13** (7): 2666-2674. 10.1109/jsen.2013.2257742.

[23] H. Khan, M. T. Rasheed, H. Liu, and S. Zhang. (2024). "High-order polynomial interpolation with CNN: A robust approach for missing data imputation". *Computers and Electrical Engineering.* **119**. 10.1016/j.compeleceng.2024.109524.

[24] A. Jazayeri, O. S. Liang, and C. C. Yang. (2020). "Imputation of Missing Data in Electronic Health Records Based on Patients' Similarities". *Journal of Healthcare Informatics Research.* **4** (3): 295-307. 10.1007/s41666-020-00073-5.