Journal of Information Assurance and Security.
ISSN 1554-1010 Volume 16 (2021) pp. 059-072

A Novel Detection Technique using Martingales For Optimal Parameters in Time Series

Jonathan Etumusei[1], Jorge Martinez Carracedo[2] and Sally McClean[3]

[1]Ulster University, School of Computing,
Shore Rd, Newtownabbey BT37 0QB,
Jordanstown, United Kingdom
etumusei-j@ulster

[2]Ulster University, School of Computing,
Shore Rd, Newtownabbey BT37 0QB,
Jordanstown, United Kingdom
j.martinez-carracedo@ulster.ac.uk

[3]Ulster University, School of Computing,
Shore Rd, Newtownabbey BT37 0QB,
Jordanstown, United Kingdom
si.mcclean@ulster.ac.uk

Abstract: **Time series analysis is becoming essential in different areas for the observation and monitoring of time-sequential data sets to extract relevant statistics and predict the series' behaviour. Current approaches effectively detect changes in the data streams. Still, most of these techniques are limited to noise interference and the inability to identify the most significant parameter values for productive abnormality detection in time series. In this paper, we improve on the previous moving median of the martingale sequence and the Gaussian moving average of the martingale sequence approaches by implementing various optimisation algorithms such as G-mean enumeration, genetic algorithms and particle swarm optimisation. The use of these methods allows us to find the optimal parameter set for each algorithm. The proposed system can reduce noise in the data and estimate the change degree in time series scenarios. Results show that the proposed approaches perform better than the previous martingale approaches.**

Keywords: Anomaly detection, time series, martingales, optimisation, electromagnetic, human activity recognition.

I. Introduction

A time series is a sequence of observations that takes place at consistent intervals. Data acquired from observing and monitoring time series are ubiquitous nowadays. For example, in the trading sector, we could observe and record the annual sales amount, gross domestic product, and gross domestic product yearly sales figures for budgeting purposes. Time series [5] can be categorised into two types: univariate (one variable) and multivariate (multiple variables). The univariate time series can be temperature measurement obtained using thermometer while the multivariate could be a tri-axial accelerometer data with x, y and z-axis.

Time series [36][56][55][47][11] characteristics can be classified into auto-correlation, seasonality and stationarity. Auto-correlation is the extent of closeness between a given time sequence and a diminishing rendition of the series over previous periods. It is also a measure of the relationship between present and past values. For example, when the temperatures measured within five years are of similar magnitude and variance. Seasonality represents the alteration that takes place in a time series. This periodic alteration can result from certain conditions, such as climate conditions, vacation and festive season, that occur weekly or monthly within a year. Stationarity refers to the statistical properties of time series that do not alter over time. Analytic tools and statistical tests often rely on stationarity for effective analysis of time series.

In time series, an anomaly or outlier is a data point that does not follow expected trends and is noticeably dissimilar from the rest of the data. Mathematically, an anomaly is the data point that diverges by a specific deviation from the mean. Therefore, anomaly detection in time series can also refer to locating outlier points concerning the rest of the data set. There are three types of anomalies in time series: point, contextual and collective anomalies [23][64]. P A point anomaly happens when a point in the time series is far off compared to the rest of the data set. An example of a point anomaly is the process of discovering credit card fraud based on spending activities or history. This anomaly happens when there are unusual spending and withdrawals. A contextual or conditional anomaly occurs when the abnormality is context-specific. For example, a high temperature in winter would be an anomaly, while a high summer temperature is not. A collective anomaly happens when a collection of data points within a data set is considered abnormal and deviate significantly from the entire data set. Still, the differing data points are not necessarily anomalous. A typical example of a collective anomaly is breaking the rhythm in electrocardiogram readings. The rare happenings in an electromagnetic data set due to seismic activity can be seen as a collective anomaly.

Anomaly detection can discover unusual patterns in time series that do not conform to expected behaviour. There are many applications of anomaly detection [41]. For instance, intrusion detection systems [51] find abnormal behaviour in network traffic resulting from a hack. Health monitoring systems [21] identify a malignant tumour in the body with an MRI scan. Also, fraud detection applications [37] can discover an unfamiliar surge in spending in a controlled environment. Anomaly detection can be achieved using the following approaches: the simple statistical approach and the machine learning-based approach. A simple statistical method [8] can discover irregularities in the data set by picking out the points which deviate from the data statistics. This method is usually quite robust as it relies on data distribution. Machine learning-based approaches can be used to build a statistical model to make predictions [43]. There are three types of machine learning-based [3][52] approaches known as supervised, unsupervised and semi-supervised machine learning. The supervised machine learning-based requires a labelled training data set that contains both standard and abnormal samples for developing the predictive model [11]. The semi-supervised anomaly detection approach assumes that every training instance is acquired from the standard class. The aim is to differentiate the subsequent instances that transpire from a distinct distribution [45]. The unsupervised anomaly detection consists of one sample with both standard and abnormal instances, and the primary aim is to distinguish them [45].

In time series, anomaly detection techniques still face some challenges such as noise interference, inability to differentiate between normal and abnormal boundaries, handling the imbalance of normal and anomalous data, and addressing the variation of peculiar characteristics [7]. In the following paragraphs, we shall discuss both the univariate and multivariate time series.

An example of the univariate data set is the electromagnetic(EM)[66][35] data set contains information that will enable experts to study and predict anomalies such as earthquakes. The electromagnetic field can be influenced by electromagnetic interference, which comes from different sources and can be natural or human-made. The existence of an EM field makes possible the presence of noise [31]. Noise, in this case, refers to the intercession of the communication procedure between satellite and the Earth's surface; however, it can also occur through human or equipment errors. Being able to isolate anomalies in an EM data set has had a big effect on new algorithms' development. This paper will look at anomaly detection in electromagnetic time series.

Human activity recognition (HAR) is an example of a time series gaining popularity in machine learning and artificial intelligence. HAR is the process of predicting the motion of a person based on sensor data obtained from wireless sensors devices or other sources [16][12]. HAR data collection using sensors is sensitive to noise that interferes with the output rendering. Consequently, this interference might produce misleading information that can affect the accuracy and precision of data readings. The study of HAR can assist in establishing a more complete and robust treatment for physically disabled people [58][10]. HAR research enables medical practitioners to manage physical disabilities, such as multiple sclerosis or epilepsy [58][53]. This paper will also analyse the HAR time series for anomaly detection.

In our previous work [18], we proposed two methods that can identify anomalies in the data stream. In the current paper, we intend to improve these approaches by using optimisation techniques. The optimisation techniques can discover the optimal parameters of an algorithm [20][42]. Some examples of optimisation methods, that will be discussed in this paper, are the G-mean enumeration method (GEM), genetic algorithms (GA) and particle swarm optimisation (PSO) approach. GEM manually locates the best G-mean metric value from a range of possible solutions. The G-mean metric will be elaborated in Section IV.

GA [39][30] is a popular heuristic method motivated by natural selection procedures associated with an enhanced class of evolutionary algorithms. PSO [61] [4] is a heuristic algorithm based on the swarming behaviour of living entities. Both GA and PSO are robust search methods that use probabilistic and deterministic rules to improve iterations from a set point to another [54][27]. For more complex problems, GA and PSO are suitable algorithms; we shall explain these optimisation methods further in Section II and IV.

The paper structure is as follows: In Section II, we review the latest work done on identifying changes in different time series; in Section III, we introduce our proposed approach; in Section IV, we tested the proposed optimised algorithms on electromagnetic and HAR time series and compare the result with the previous version of the algorithms; we conclude the paper in Section V discussing the results obtained and the future work of the research.

II. Related work

Many anomaly detection approaches can discover abnormalities in a data stream. This section briefly discusses these methods.

In our previous work [18], we proposed two methods, namely the moving median of the martingale sequence (MMMS) and Gaussian moving average of the martingale sequence (GMAS). These methods use the martingale framework to detect anomalies in time series. The techniques also minimise the noise interference in a data stream. The approaches work using a given window size and computed threshold to enhance the accuracy rate of discovering abnormalities in time series. The parameter set used in these two methods is large and optimisation results based on this parameter space have not been obtained yet.

Time series deals with big and non-stationarity data which lead to uncertainty in obtaining satisfactory results when discovering anomalies using a single-model-based method. To handle this challenge, Zhou et al. [67] proposed a method that combines model-based and similarity measurement approaches for anomaly detection. Firstly, processing of the data is performed to reduce the dimensionality obtaining a new sequence. The original data set and the new sequence are then compared to capture the structure and morphological features. Finally, this comparison leads to a method to identify abnormalities. Experimental results show that the proposed technique can enhance the performance of anomaly detection.

Detecting abnormalities in multivariate time series can be challenging. Li et al. [33] proposed a clustering-based method to find anomalies related to the magnitude and the structure of multivariate sequences. Firstly, they use a sliding window to obtain a set of multivariate sub-sequences and then an extension of fuzzy clustering is applied to discover if some underlying data structure exists in the newly generated dataset. Finally, a criterion is implemented to rebuild the multivariate sub-sequences with an optimal cluster centre and partition matrix. This is preceded by the creation of a confidence index and the optimisation of parameters using particle swarm optimisation. Experimentation conducted on several real-world data set showed that the proposed approach can identify abnormalities and structure patterns in multivariate time series.

Physical impairment can restrict the daily activities of the elderly. Such activities can include walking, standing, sleeping and taking meals. These challenges motivated Fahad et al. [19] to suggest a method that recognises movement performed in a smart home environment. The approach is also able to identify and distinguish normal from abnormal activities carried out daily. This involves the application of a probabilistic neural network on pre-segmented action data retrieved from smart home sensors. H20 autoencoder is used to discover an anomalous sequence of every action. Consequently, these abnormal points are grouped based on conditions such as lost or aberrant activity duration. Analysis and evaluation of public CASAS smart home data sets showed that this method can correctly identify anomalies in activity recognition. The following Section discusses our proposed time series model.

III. Time series models

This Section discusses the proposed method for anomaly detection in univariate and multivariate time series.

A. Univariate time series analysis

This Section explains the proposed approach for analysing the HAR data set.

Let us consider a time series $Z = z_1, ..., z_{i-1}$. The arriving point will be represented as z_i. Let us suppose that the data has been clustered into k disjoint sets $Y_1, ..., Y_k, (k <= i - 1)$ [26].

Definition 1. *The strangeness of z_i is defined as*

$$s_i = s(Z, z_i) = \parallel z_i - C_r \parallel \ , \tag{1}$$

where C_r is the centroid of the cluster Y_r, for some $r \in \{1, ..., k\}$ such that $z_i \in Y_r$. $\parallel . \parallel$ denotes the chosen distance [24].

We use the strangeness of z_i to compute the $\widehat{p_i}$ in equation (2).

Definition 2. *Let $X_1, X_2, ..., X_s$ be a sequence of random variables, the randomised power martingale (RPM) [24] [60] is indexed by $\epsilon \in [0, 1]$ defined at each time-point as*

$$M_n^{(\epsilon)} = \prod_{i=1}^{n} (\epsilon \widehat{p_i}^{\epsilon - 1}), \tag{2}$$

$\widehat{p_i}$ [24] is computed as follows:

$$\widehat{p_i}(Z \cup z_i, \theta_i) = \frac{\sharp\{j : s_j > s_i\} \ + \theta_i \sharp\{j : s_j = s_i\}}{i}, \tag{3}$$

where s_j denotes the strangeness of z_j (see [24]) and θ_i is a fixed number in $[0, 1]$ (see [24] and [60]) for every $i = 1, 2, ..., n$, being the initial Martingale value $M_0^{(\epsilon)} = 1$.

The model (2) will detect a change when

$$M_n^{(\epsilon)} > t, \tag{4}$$

where the threshold t is chosen in a probabilistic way based on Dobb's Inequality [24]. In a multidimensional data set, M_n will be computed for each of the variables. In the following Section, we introduce a method that improves the accuracy, recall and F1 of the previously described martingale approach.

B. Moving median of a martingale sequence

A moving median approach is a robust and effective technique to detect anomalies in a data stream [28]. The moving median method finds the median of a data stream using a sliding window.

Once the martingale sequence for our time series has been computed, we can implement moving medians on the sequence of martingale points [18]. The main reason why we use the median rather than the mean is that it is not affected by individual points, but only their order. This feature builds on the idea that the median can smooth time series by isolating the effect of noise and bringing out the patterns of the data set. Therefore, this model will detect a change when

$$W_k \geq t, \tag{5}$$

where the W_k represents the MMMS points and t the threshold. There are several ways of choosing a threshold t. Ho and Weschler [24] proposed a probabilistic approach to compute it, while Ley et al. [32] proposed a threshold based on outlier detection: $\bar{x} \ \pm 3 * MAD$, where $\bar{x}$ denotes the mean of the data points and the MAD the mean absolute deviation. In this paper, we compute the threshold using $\bar{x} \pm \sigma$, where σ is the standard deviation of $\{W_k \mid k = 1, ..., s\}$ points and μ denotes the mean of $\{W_k \mid k = 1, ..., s\}$.

We use σ rather than the MAD because the absolute deviation is less sensitive to significant outliers than the use of σ, which is more susceptible to substantial outliers [25]. If W_k exceeds the given threshold of t, then an anomaly has been detected. When an abnormality is detected, the computation of W_k terminates, and the algorithm is re-initiated.

C. Gaussian moving average of a martingale sequence

The Gaussian function [34][18] can be used as a smoothing operator to compute a weighted average of the martingale points. The nearest position to the average acquires a more considerable weight while the ones apart from the standard will gain a lesser weight. This weighting process can isolate noise, making the pattern clearer.

We consider a martingale sequence $M = \{M_k : k = 1, ..., s - 1\}$. For every $k \in \{1, ..., s - 1\}$ we can compute the Gaussian function of the martingale point M_i as

$$e_i = e^{-(M_i-\mu)^2/(2\sigma^2)}, \tag{6}$$

where μ and σ are the mean and standard deviation of M [34]. We can compute a new point L_i using the equation:

$$L_i = \frac{e_i}{\sum_{i=0}^{k} e_i} . \tag{7}$$

Finally, we define a final sequence of values as (8):

$$A_k = \sum_{i=0}^{k} L_i M_i , \tag{8}$$

for $k = 1, ..., n$. If the computed A_k value is greater than the threshold t, it is feasible to presume that an abnormality or anomaly occurs. As before, we will be using a threshold of $\bar{x} \pm \sigma$. where $\bar{x}$ and σ are the mean and standard deviation of A_k. Therefore, the model for identifying anomalies in time series is given as:

$$A_k \geq t. \tag{9}$$

D. Multivariate time series model

In this Section, we explain the process of reducing the dimension of the multivariate sequences W_k and A_k to a single variable suitable to study.

Let us consider a data sequence $\{X_1, ..., X_s\}$, where each point X_i is a j vector (being j the number of variables in the study). M_n, W_k and A_k are computed for each variable at any time point. The next step will be to reduce this new multidimensional sequence into a single metric.

Once M_n, W_k and A_k are computed for each variable at a given time point the mean of all these values will be calculated. For the multivariate randomised power martingale (MRPM), the k-th mean element is computed using the following equation:

$$V_k = \frac{M_n(r_1^{(k)}) + M_n(r_2^{(k)}) + ... + M_n(r_j^{(k)})}{j} , \tag{10}$$

where $r_i^{(k)}$ is the i-th variable of X_k.

For the multivariate moving median of the martingale sequence (MMMMS), the k-th mean element is computed using the equation:

$$C_k = \frac{W_k(r_1^{(k)}) + W_k(r_2^{(k)}) + ... + W_k(r_j^{(k)})}{j} , \tag{11}$$

where $r_i^{(k)}$ is the i-th variable of X_k.

For the multivariate Gaussian moving average of the martingale sequence (MGMAS), the k-th mean element is acquired using the equation:

$$E_k = \frac{A_k(r_1^{(k)}) + A_k(r_2^{(k)}) + ... + A_k(r_j^{(k)})}{j} , \tag{12}$$

where $r_i^{(k)}$ is the i-th variable of X_k.

The condition for change, for the different methods, is given as:

$$V_k \geq t \; (MRPM) , \tag{13}$$

$$C_k \geq t \; (MMMMS) , \tag{14}$$

$$E_k \geq t \; (MGMAS). \tag{15}$$

Depending on the method, V_k, C_K, and E_k will be our analysis point. If V_k exceeds the computed threshold t, then a change has been detected. The same condition applies to C_K (14) and E_K (15). In the next Section, we will discuss some effective optimisation techniques used to optimise the parameters of the algorithms. Figure 1 presents the details of the MMMS/GMAS and MMMMS/MGMAS algorithms, respectively. The proposed anomaly detection methodology (Algorithms 1 and 2) is required as a consequence of threshold violation for the martingale. One can observe that, for both the univariate and multivariate data analysis, a change happens when W_k, A_K (8), C_k (14) and E_K (12) exceeds the threshold t. Hence, no re-computation is required when computing M_i.

```
Data: Input (F): EM univariate data set
Result: Output: MMMS/GMAS points
Initialise: M(0) = 1; i = 1; F = {};
Set values for cluster group k, ϵ value, window size;
while do
    A new example of z_i is discovered;
    if F ={ } then
        Set the strangeness of z_i := 0
    else
        Compute the strangeness of z_i and the data points in F
        Compute the p̂_i of z_i;
        Compute the M_i using (1);
        Compute the W_k or A_k points;
        Compute the threshold t
    end
    if W_k or A_k ≥ t then
        Change detected
        Re-initiate M_i = 1
    else
        Add z_i into F ;
    end
    if i = i + 1 ; then
    end
end
Algorithm 1: MMMS/GMAS output
```

```
Data: Input (F): HAR multivariate data set
Result: Output: MMMMS/MGMAS points
Initialise: M(0) = 1; i = 1; F = {};
Set cluster group k, ϵ value, window size;
while do
    A new example of z_i is discovered;
    if F ={ } then
        Set the strangeness of z_i := 0
    else
        Compute strangeness of z_i and the data points in F
        Compute p̂_i of z_i;
        Compute M_i using (1);
        Compute W_k or A_k points;
        Compute C_k or E_k points;
        Compute the threshold t
    end
    if C_k or E_k ≥ t then
        Change detected
        Re-initiate M_i = 1
    else
        Add z_i into F ;
    end
    if i = i + 1 ; then
    end
end
Algorithm 2: MMMMS/MGMAS output
```

Figure. 1: Algorithm 1 and 2

Most applications make use of parameters, which can be optimised, to enhance performance [1]. The rationale of using an optimisation approach is to identify the best solution to the problem by minimising or maximising the objective function of an algorithm. The following Section discusses the proposed optimisation algorithms, results and evaluation.

IV. Experimental results

This Section gives a performance evaluation of the different approaches adopted to discover anomalies in time series. The Section also describes the pre-processing method and analysis of some time series (electromagnetic, electrocardiogram and human activity recognition data set) using the proposed methods.

The evaluation performance for our methods is measured using evaluation metrics such as accuracy, precision, recall, harmonic mean $(F1)$ and G-mean [6][49]. The accuracy, recall, precision, specificity, sensitivity and G-mean matric can analyse the optimal selection for MMMS, GMAS, MMMMS, and MGMAS approaches. In this case, the confusion matrix (CM) can evaluate the performance of these algorithms [44].

Accuracy [17] is an intuitive performance metric defined as the ratio of changes correctly detected in HAR to the total observations. The following gives accuracy:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$

We define true negatives (TN) as the false changes that are correctly identified as false. True positives (TP) is the actual changes that are correctly detected, while false positives (FP) are incorrect changes that are identified as true. False negatives (FN) are the actual changes that are identified as incorrect. Therefore, precision, recall (also known as sensitivity), F1 score and specificity are computed as follows:

$$Precision = \frac{TP}{TP + FP}.$$

$$Recall(sensitivity) = \frac{TP}{TP + FN}.$$

$$F1score = \frac{2 * Recall * Precision}{Recall + Precision}.$$

$$Specificity = \frac{TN}{TN + FP}.$$

G-mean [2] [44] is the measure of the correlation and the overall efficiency of the activities. G-mean combines the recall and the specificity. A low G-Mean indicates a poor performance in categorising positive cases regardless of whether the negative possibilities are precisely classified. The G-mean metric is vital to prevent overfitting the false-negative and underfitting the false-positive class.
G-mean is given as:

$$G - mean = \sqrt{Recall * Specificity}.$$

The following Section discusses ways by which we can optimise the parameter of the proposed algorithm for enhanced performance.

A. *Time series pre-processing*

We use a labelled EM data set to test the effectiveness of our algorithm. The ESA Swarm satellite generates the EM time series [13]. The EM is a univariate dataset that contains 3751 data points. The data set include anomalies that seismic experts have labelled. In Figure 2, it is possible to see a plot of the data set and its abnormal fluctuation.

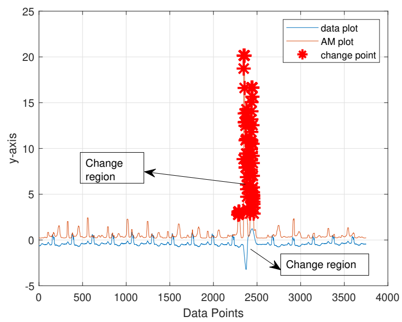

Figure. 2: EM data set

The HAR data set was acquired using the accelerometry data points from the shimmer wireless sensor platform (SWSP) [9] attached to healthy participants. The SWSP have 3-axis MEMs, which are integrated into the device used to capture the acceleration of the individual. The Shimmer devices were attached to the participant right arm, left arm and right leg to enable lateral and anterior-posterior movement of the volunteer to be captured efficiently [65]. The participant performs different scenarios within a home environment [65]. The first set of scenarios (scenario 1) involves the participant following activities which include ascending stairs, walk and sit down, respectively. For the multivariate accelerometer data, a finite impulse response (FIR) filter, is used as a low-pass filter to introduce a set of acceleration values for every specific activity and also to reduce the noise accumulated from such activities [48] [65]. The labelled multivariate HAR data (marked by experts in the field) for the stand to sit scenario are illustrated in Fig 3 to Fig 5. The anomalies are identified in red dotted lines, and it occurs within the region of 3185 to 3342 data points.

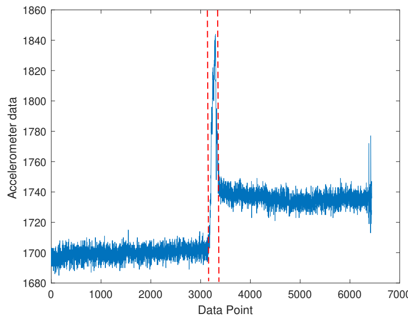

Figure. 3: Accelerometer- X data set

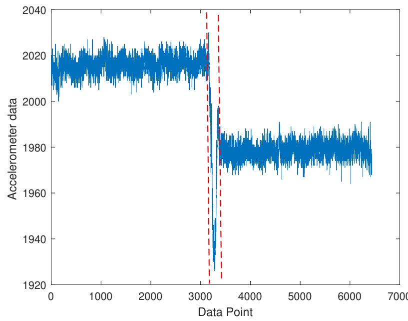

Figure. 4: Accelerometer- Y data set

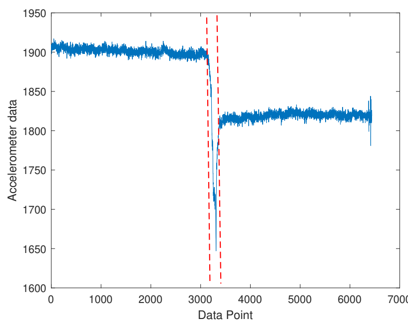

Figure. 5: Accelerometer- Z data set

B. *G-mean enumeration method (GEM)*

The GEM is the process of locating the maximal G-mean value for different parameter ranges [62]. Note that the family of martingale we are working with depends on $\epsilon \in [0, 1]$, window size (WS) [18]. We will use GEM to look at each tuple of parameters (ϵ, WS) for any epsilon value in the set $\{0, 0.1, 0.2, ..., 1\}$ and any window size value in the set $\{1.0, 2.0, 3.0, ..., 20\}$ for both the MMMS and GMAS algorithms using univariate EM series. We first find the optimal parameters; secondly, we use these parameters to check another data set. The implementation of the GEM using the MMMS method is represented as MMMS(GEM). Also, the performance of the GEM using the GMAS method is described as GMAS(GEM); this process is further discussed in the following paragraphs. We apply the GEM approach using MMMS methods on an EM data set of 3750 data points to identify the optimal parameters. We obtain a maximum G-mean value of 0.9182 and the optimal parameters are $\epsilon = 0.52$ and $window\ size = 19$ using the GEM technique. These results are illustrated in Figure 6, which shows the optimal G-mean point. The implementation of the GEM using the MMMS method is represented as MMMS(GEM). Later, we applied these parameter values on another EM data set consisting of 3751 points and obtained a G-mean value of 0.7482. The confusion matrix of results is shown in Table 1. The Table shows that the MMMS(GEM) method can acquire a TP of 33.4% and TN of 97.8% independently.

An analogous procedure is used with GMAS; we identify the optimal parameters using the GEM for EM data set consisting of 3750 data points. The maximum G-mean value of 0.8896 is reached when the parameters are $\epsilon = 0.72$ and $window\ size = 19$. The implementation of GEM using the MMMS method is called MMMS(GEM). Also, when we apply GEM using GMAS approach, we represent it as GMAS(GEM). These results are illustrated in Figure 7, which shows the optimal G-mean point. In a later stage, we apply these optimal parameters using a similar EM data set of 3751 data points and obtain a G-mean value of 0.7682. The CM metrics of the GMAS(GEM) method using different EM data are described in Table 1. The Table shows that the GMAS(GEM) method can acquire a TP of 33.4% and TN of 97.8% independently.

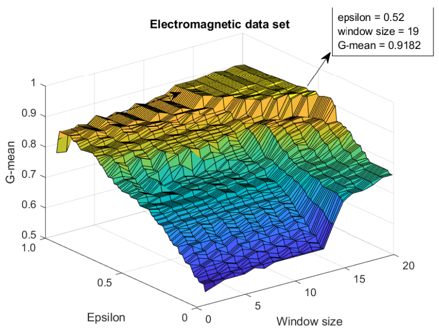

Figure. 6: GEM using MMMS

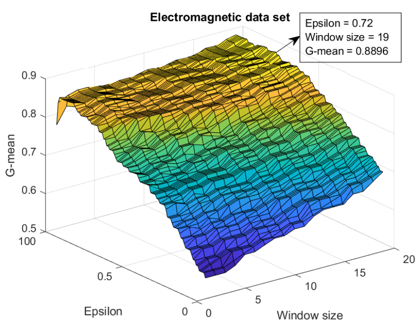

Figure. 7: GEM using GMAS

We use the same technique applying the GEM on both MMMS and MGMAS using the multivariate HAR data set. The results are illustrated in Figure 8 and 9, showing the optimal G-mean point, respectively. The process is referred to as MMMMS(GEM) and MGMAS(GEM), respectively. In

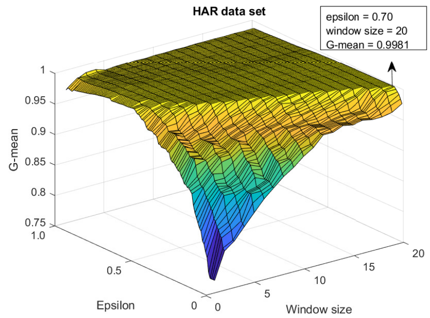

Figure. 8: GEM using MMMMS

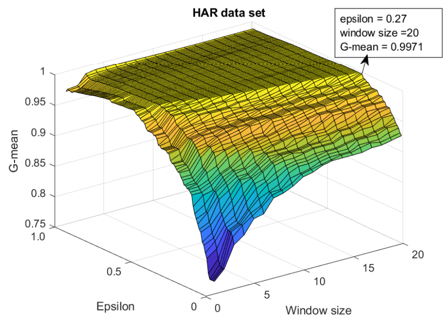

Figure. 9: GEM using MGMAS

Table 2, we observe that the TN identified is 100.0%, and TP discovered is 97.8% using the MMMMS(GEM) approach. While for the MGMAS(GEM) method, the TN detected is 100%, and TP discovered is 91.0%.

Table 1: Confusion metrics of the EM time series

Approach	TN	TP	FN	FP
MMMS(GEM)	3460(97.9%)	110(53.4%)	81(2.3%)	100(47.5%)
GMAS(GEM)	3322(98.0%)	123(34.1%)	68(2.0%)	238(65.9%)

Table 2: Confusion matrix of the HAR time series

Approach	TN	TP	FN	FP
MMMMS(GEM)	2923(100.0%)	91(97.8%)	0(0.0%)	2(2.2%)
MGMAS(GEM)	2916(100.0%)	91(91.0%)	0(0.0%)	9(9.0%)

The GEM optimisation technique has limitations as it cannot optimise complex problems [50][59]. A more efficient way of optimising complex algorithm parameters is the use of the genetic algorithm (GA) or particle swarm optimisation method [40] [14].

C. *Genetic algorithm (GA)*

GA can discover the optimal parameter of an algorithm [30][22]. GA can be used to examine the fittest value over the sequential generation of output. GA is achieved by incorporating results (genetic crossover) using neighbouring search (mutations) and replaces the population with fitter results [57].

GAs use the fitness function to obtain the optimal parameter values of an algorithm [15]. In this case, the fitness function measures the closeness of a given result to the leading solution of a chosen problem. Our proposed method makes use of the following parameters (ϵ, WS) as discussed in Section III. Therefore, we initialise the population of the vectors containing the mentioned input parameters to give the G-mean values [29][63]. In this case, the fitness function is expressed as the maximum G-mean value with a specified range of parameters. The G-mean is the estimate, that locates the ratio and the overall performance of the physical activity by combining specificity and sensitivity [2][29]. The fitness function is represented as follows:

$$G - Mean_{max} = max_{(\epsilon,\beta)}(G - Mean_{(Q)}), \quad (16)$$

where ϵ and β ranges from 0 to 1 for each activity. Q represents the method to be implemented [29]. Our proposed approaches use equation (16) by initialising the upper and lower bounds of the two parameters to identify the maximum G-mean value. The various GA parameters implemented to maximise the fitness function are shown in Table 3. The GA flowchart is illustrated in Figure 14.

Table 3: GA Parameter values

Parameters	value
Population Size	**200**
Selection	**stochastic uniform**
Crossover rate	**0.8000**
Mutation	**Gaussian**
Crossover	**Heuristic**
MaxGenerations:	**100*numberOfVariables**

We implement GAs using our algorithm on both the EM and the HAR data set. GA inspects each tuple of parameters for any epsilon value in the range $\{0, 0.1, 0.2, ..., 1\}$ and any window size value in the range $\{1, 2, 3, ..., 20\}$ for both the MMMS and GMAS algorithms using univariate EM series and for MMMMS, MGMAS algorithms using a multivariate HAR data set. We apply the GA approach using MMMS methods on an EM data set of 3750 data points to find the optimal parameter values. We obtain a maximal G-mean value of 0.9117, and the optimal parameters are $\epsilon = 0.5469$ and $window\ size = 19.2754$ using the GA. Figure 10 shows the best G-mean value identified. In a later stage, the optimised parameter values are applied on another EM data set consisting of 3751 points and obtained a G-mean value of 0.9999. This procedure is repeated using the GMAS method. We discover the maximum G-mean value of 0.8909, and the parameters that produce this result are $\epsilon = 0.7831$ and $window\ size = 19.7267$. Figure 11 show the best G-mean value discovered. We then apply these optimal parameters using a similar EM data set of 3751 data points and obtain a G-mean value of 0.7673. GA using MMMS and GMAS methods can be expressed as MMMS(GA) and GMAS(GA), respectively. Table 4 shows that the TN identified for MMMS(GA) method is 97.7% and TP is 48.5%. Also, for GMAS(GA) method, TN and TP discovered are 98.0% and 26.3% independently.

The MMMMS(GA) and MGMAS(GA) are used to discover the optimal parameter (ϵ, WS) on a HAR data set consisting of 6441 data point. The optimal parameter is used on a new data set consisting of 3017 data point. The best G-mean values obtained using the algorithms are illustrated in Figure 12 and 13, respectively. Table 5 shows that TN identified is 100%, and TP is 100% using MMMMS(GA) method. Also, for MGMAS(GA) method, the TN and TP discovered are 100% and 92.9% independently.

GA usage expands to many fields such as computing and engineering. The limitation of the GA is that it is computationally costly [54][27]. The following Section discusses the PSO approach, which is not computationally expensive and takes less iterations.

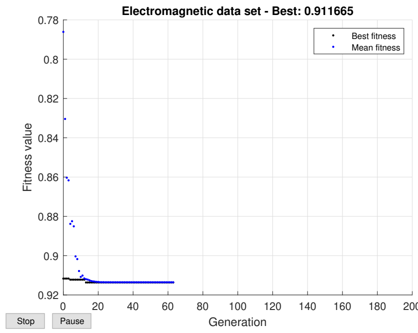

Figure. 10: MMMS(GA) iteration

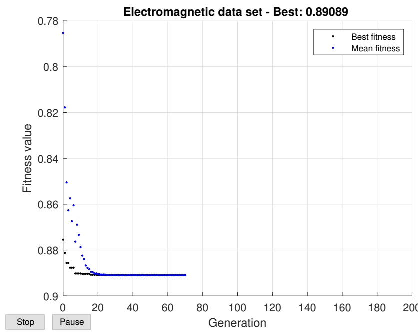

Figure. 11: GMAS(GA) iteration

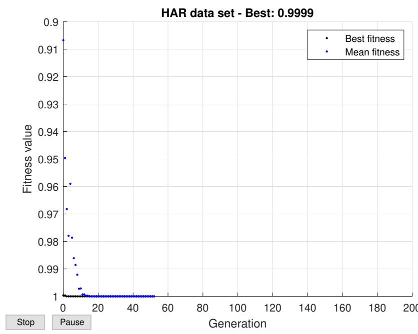

Figure. 12: MMMMS(GA) iteration

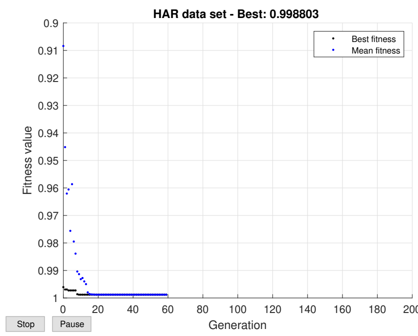

Figure. 13: MGMAS(GA) iteration

Table 4: Confusion metrics of the EM time series

Approach	TN	TP	FN	FP
MMMS(GA)	3442(97.7%)	111(48.5%)	80(2.3%)	118(51.5%)
GMAS(GA)	3263(98.0%)	123(29.3%)	68(2.0%)	297(70.7%)

Table 5: Confusion matrix of the HAR time series

Approach	TN	TP	FN	FP
MMMMS(GA)	2923(100.0%)	91(100.0%)	0(0.0%)	0(0.0%)
MGMAS(GA)	2918(100.0%)	91(92.9%)	0(0.0%)	7(7.1%)

D. *Particle swarm optimisation (PSO)*

PSO [38] [46] is motivated by social characteristics and behaviour, such as a swarm of insects, schools of fish and flocks of birds. In these circumstances, an individual such as a fish or insect is known as a particle. Each individual in the swarm depicts their reasoning and the collective reasoning of the population. In PSO, a particle of the swarm constitutes a prospective solution. Therefore, PSO is a computational technique that optimises a problem by iteratively enhancing the candidate solution (particles) concerning a given estimate of quality [27]
PSO involves the following:

- Several particles placed in the search order of the problem with each particle estimating the problem at its present position
- Each particle then determines its motion through the search order by integrating some features of the past and present positions using one or more constituents of the swarm with some arbitrary changes
- The next iteration occurs when the movement of all the particles is completed
- Consecutively, the swarms are likely to move near the maximal of the fitness function

The main difference between GA and PS0 is that the GA evolves the candidate solution. Like the GA, PSO examines the common values over a problem's possible region. Each individual in the particle swarm consists of 3D dimensional vectors. D represents the search region's dimension, which is the present location s_i, the past best location n_i, and the velocity is represented by v_i [46]. The MMMS and GMAS approaches [18] makes use of the following parameters ($\epsilon, window\ size$) as discussed in Section III. Therefore, PSO optimises our fitness function in equation (16) to produce the optimal parameters that maximise G-mean [29][63]. The various PSO parameters implemented to maximise the fitness function are shown in Table 6. The PSO method flowchart is illustrated in Figure 15.

Table 6: PSO Parameter values

Parameters	value
InertiaRange	**[0.10000, 1.1000]**
InitialSwarmSpan	**200**
MaxIterations	**200 * NumberOfVariables**
MaxStallIterations	**20**
MinNeighboursFraction	**0.250**
SwarmSize:	**100**
SelfAdjustmentWeight	**1.4900**
SocialAdjustmentWeights	**1.4900**

We perform the PSO approach using MMMS on EM data set to obtain the maximal G-mean output of 0.9188 and optimised parameters are $\epsilon = 0.5138$ and $window\ size = 20$. Figure 16 show the best G-mean outcome identified by the algorithm. Subsequently, we use the optimal parameters on an EM data set of 3750 data points to obtain a G-mean value

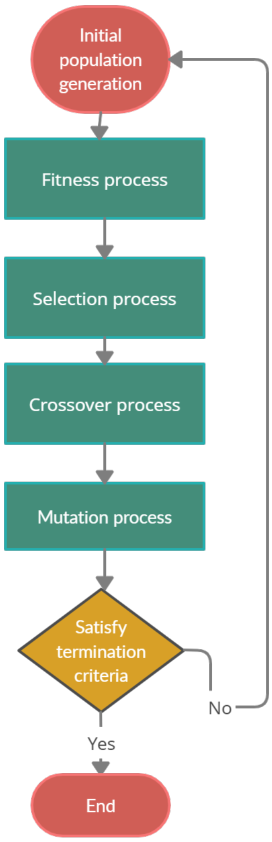

Figure. 14: GA flowchart

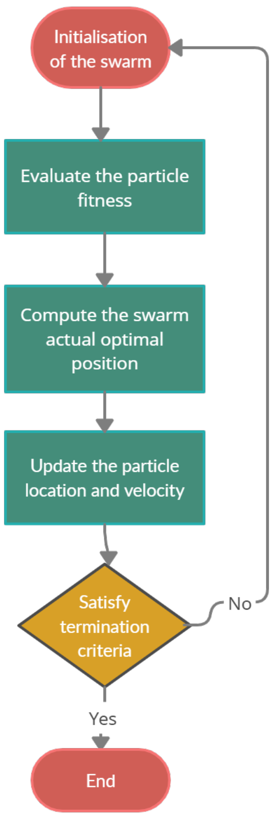

Figure. 15: PSO flowchart

of 0.9188. After that, the optimised parameters are applied on another EM data set consisting of 3751 data points, and we acquire a G-mean value of 0.7548. Table 5 shows that the TN identified for the MMMS(PSO) method is 97.8%, and TP is 47.4%. We repeat the procedure using GMAS method. We discover the maximum G-mean value as 0.8909, and the parameters are $\epsilon = 0.7826$ and $window\ size = 19.7738$. Figure 17 shows the maximal G-mean value obtained by the algorithm. We apply these optimal parameters using a similar EM data set of 3751 data points and get a G-mean value of 0.7684. Table 4 shows that TN identified for the GMAS(PSO) method is 97.8% and TP is 47.4%. Also, for GMAS(PSO) method, TN and TP discovered are 98.0% and 29.4% independently.

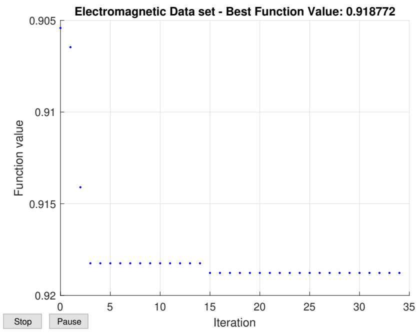

Figure. 16: MMMS(PSO) iterations

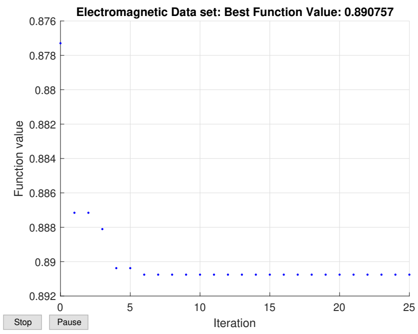

Figure. 17: GMAS(PSO) iterations

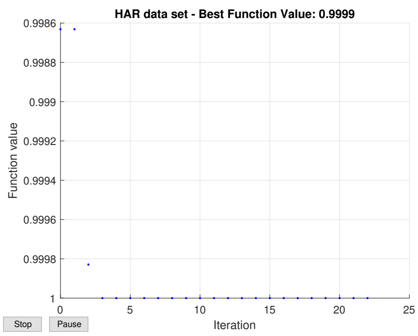

Figure. 18: MMMMS(PSO) iterations

PSO is also used to optimise the parameters of MMMMS and MGMAS using multivariate HAR data set, and the results are shown in Figure 18 and 19. These figures show the best G-mean generated for both techniques. The application of the PSO method using MMMMS and MGMAS will be named as MMMMS(PSO) and MGMAS(GMAS) independently. MMMMS(PSO) and MGMAS(PSO) are used to locate the optimal parameter (ϵ, WS) on a HAR data set consisting of 6441 data point. The optimal parameter is used on a new data set consisting of 3017 data point. Table8 shows that MMMMS(PSO) can identify 100% of TN and 100% of TP while MGMAS(PSO) detected 100% of TN and 94.8% of TP.

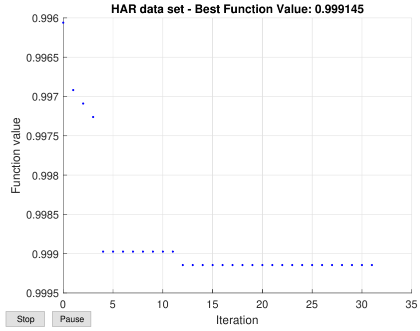

Figure. 19: MGMAS(PSO) iteration

Table 7: Confusion metrics of the EM time series

Approach	TN	TP	FN	FP
MMMS(PSO)	3459(97.8%)	112(47.4%)	79(2.2%)	101(47.4%)
GMAS(PSO)	3264(98.0%)	123(29.4%)	68(2.0%)	296(70.6%)

Table 8: Confusion matrix of the HAR time series

Approach	TN	TP	FN	FP
MMMMS(PSO)	2925(100.0%)	91(100.0%)	0(0.0%)	0(0.0%)
MGMAS(PSO)	2920(100.0%)	91(94.8%)	0(0.0%)	5(5.2%)

We also follow the procedure to optimise the previous MRPM for the HAR data set using GEM, GA and PSO techniques. MRPM(GEM), MRPM(GA) and MRPM(PSO) are used to locate the optimal parameter (ϵ, WS) on a HAR data set consisting of 6441 data point. The optimal parameter is used on a new data set consisting of 3017 data point. These processes are illustrated in Figure 20 to 22. For HAR data analysis, the confusion matrix of the optimal MRPM parameter is shown in Table 9. In the next Section, we will summarise the experimental results using the various optimisation techniques.

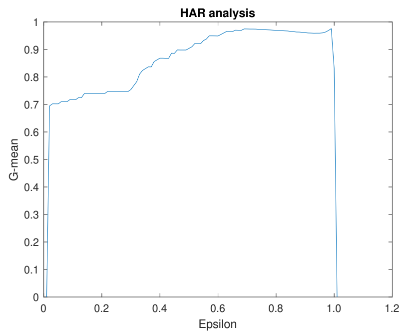

Figure. 20: MRPM(GEM) iteration

Table 9: Confusion matrix of the HAR time series

Approach	TN	TP	FN	FP
MRPM(GEM)	2918(99.0%)	63(90.0%)	28(1.0%)	7(10.0%)
MRPM(GA)	2875(99.9%)	88(63.8%)	03(0.1%)	50(36.2%)
MRPM(PSO)	2875(99.9%)	88(63.8%)	03(0.0%)	50(36.2%)

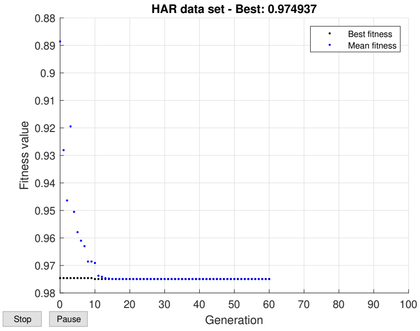

Figure. 21: MRPM(GA) iteration

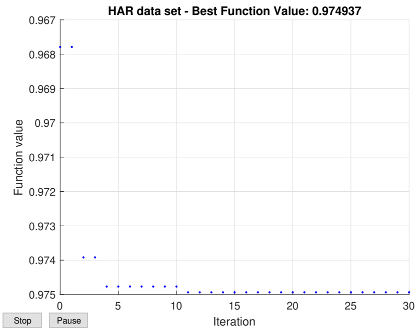

Figure. 22: MRPM(PSO) iteration

E. Summary of results

In this Section, we shall analyse the performance evaluation metrics presented in Table 10 and 11 based on the experimentation carried out.

Table 10: Evaluation metrics using the EM time series

Approach	ϵ	Window size	Accuracy	Recall	Precision	F1	G-mean
MMMS(GEM)	0.52	19	0.9517	0.5759	0.5238	0.5486	0.7482 (14.19%)
GMAS(GEM)	0.72	19	0.9183	0.6335	0.3315	0.4353	0.7682 (14.56%)
MMMS(GA)	0.5469	19.2754	0.9472	0.5812	0.4847	0.5286	0.7496 (14.22%)
GMAS(GA)	0.7831	19.7267	0.9027	0.6440	0.2929	0.4026	0.7683 (14.57%)
MMMS(PSO)	0.5138	20	0.9520	0.5864	0.5258	0.5545	0.7548 (14.31%)
GMAS(PSO)	0.7826	19.7738	0.9030	0.6440	0.2936	0.4033	0.7684 (14.57%)
RPM	0.66	-	0.9200	0.5445	0.3281	0.4090	0.7156 (13.57%)

Table 11: Evaluation metrics using HAR time series

Approach	ϵ	Window size	Accuracy	Recall	Precision	F1	G-mean
MMMMS(GEM)	0.52	19.0	0.9993	0.9999	0.9785	0.9891	0.9997(11.39%)
MGMAS(GEM)	0.59	19.0	0.9970	0.9999	0.9100	0.9529	0.9985(11.37%)
MMMMS(GA)	0.5469	11.1072	0.9999	0.9999	0.9999	0.9999	0.9999(11.39%)
MGMAS(GA)	0.5777	18.2672	0.9977	0.9999	0.9286	0.9630	0.9988(11.38%)
MMMMS(PSO)	0.5073	20	0.9999	0.9999	0.9999	0.9999	0.9999(11.39%)
MGMAS(PSO)	0.5708	19.9492	0.9894	0.9999	0.9479	0.9733	0.9991(11.38%)
MRPM(GEM)	0.99	-	0.9884	0.6923	0.900	0.7826	0.8311(09.47%)
MRPM(GA)	0.6748	-	0.9824	0.9670	0.6377	0.7686	0.9749(11.10%)
MRPM(PSO)	0.6737	-	0.9824	0.9670	0.6377	0.7686	0.9749(11.10%)

Experimentation analysis shows that GMAS(PSO) slightly outperforms GMAS(GA) and GMAS(GEM) approaches in terms of G-mean output when analysing the EM data set. However, the GMAS(PSO) produces a low precision. This situation arises as we aim to capture anomalies in EM data set, leading to a high FP rate. We intend to address this issue in future work. MMMS(PSO) also gives a better G-mean result compared to the MMMS(GA) and MMMS(GEM). These results are illustrated in Table 10. We can assert that the PSO technique is efficient for optimising the abnormality detection algorithm in the analysis of EM data set. We also compare these results with the previous RPM method for EM analysis, which was indicated in our earlier paper [18]. The comparison shows that our proposed optimised methods outperform the RPM approach.

Experimentation results show that MMMMS(PSO) and MMMMS(GA) gives a slightly better performance in terms of G-mean output than the MMMMS(GEM) methods. Also, MGMAS(PSO) produces a marginally better G-mean value compared to MGMAS(GA) and MGMAS(GEM). Furthermore, MRPM(PSO) approach can yield a G-mean output that is insignificantly better than that of the MRPM(GA) and MRPM(GEM) techniques, respectively. These results are shown in Table 11. We can affirm that PSO can effectively study HAR time series.

We can further measure the performance of the optimised algorithms by analysing the evaluation metrics in Table 10 and 11 using histogram charts. It is then possible to determine which optimisation technique is more effective when analysing EM data set. The histogram plots for the algorithms are illustrated in Figure 23 to 24.

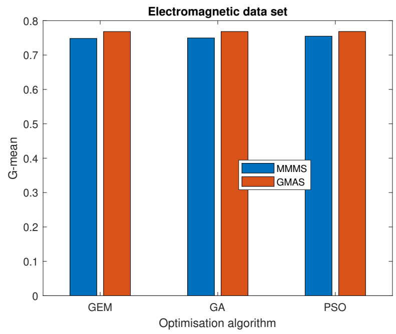

Figure. 23: Optimisation output

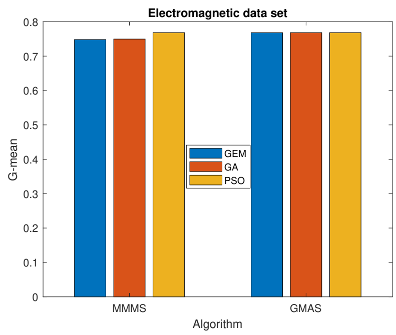

Figure. 24: Algorithms output

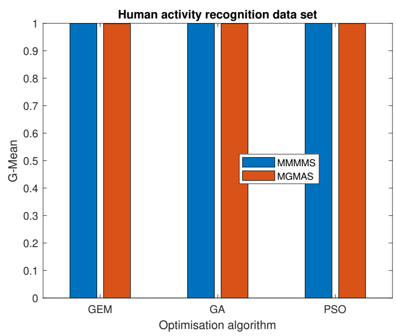

Figure. 25: Optimisation output

Looking at Figure 23 to 24, we can notice that PSO is an effective optimisation algorithm compared to GEM and GA; however, the difference is modest. GA can be effective for handling complex problems but it is limited as the number of elements encountering mutation becomes large, and thus

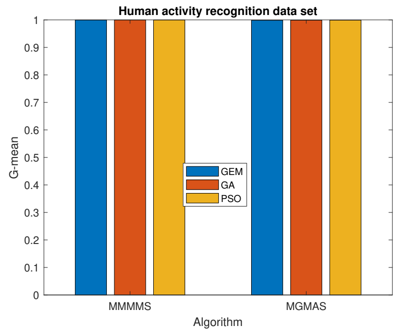

Figure. 26: Algorithm output

cause an increase in the search space [54][27]. This situation makes it longer for GA to identify the optimal parameter value. PSO becomes the better alternative as it needs minimum parameters and a lower iteration process. The variables in PSO can obtain any values depending on the location of the value in the particle space and their reciprocal velocity vector. While the GA merge towards a confine optimum rather than the global one for the problem, the PSO intuitively locate the global optima [54][27].

V. Conclusion and future work

This paper briefly discussed the previous RPM, MMMS and GMAS method. These approaches can identify an anomaly in the data stream. The methods use the martingale framework to achieve change detection. We propose a system that can improve the performance of these methods by optimising their parameters. The approaches are used to analyse the EM and HAR data sets. The output shows that the proposed techniques using the optimisation methods give a higher G-mean value of over (0.20%) than the previous RPM and MRPM methods. We also observe that the PSO optimisation technique performs better than other optimisation methods with a slight difference of over 0.001%.
Future work is needed to validate the approach using a wide range of time series that extend to different areas.

VI. Acknowledgement

This work was supported by the University of Ulster Vice-Chancellor's Research Studentship. The authors would like to thank anonymous reviewers for their constructive suggestions.

References

[1] Abd-Alsabour, N.: On tackling real-life optimization problems. International journal on advanced science, engineering, and information technology, 9l (2) pp. 640–647 (2019)

[2] Akosa, J.: Predictive accuracy: A misleading performance measure for highly imbalanced data. In: Proceedings of the SAS Global Forum. vol. 12 (2017)

[3] Auger, S.D., Jacobs, B.M., Dobson, R., Marshall, C.R., Noyce, A.J.: Big data, machine learning and artificial intelligence: a neurologist's guide. Practical Neurology 21(1), 4–11 (2021)

[4] Babanezhad, M., Behroyan, I., Nakhjiri, A.T., Marjani, A., Rezakazemi, M., Heydarinasab, A., Shirazian, S.: Investigation on performance of particle swarm optimization (pso) algorithm based fuzzy inference system (psofis) in a combination of cfd modeling for prediction of fluid flow. Scientific Reports 11(1), 1–14 (2021)

[5] Beeram, S.R., Kuchibhotla, S.: Time series analysis on univariate and multivariate variables: a comprehensive survey. Communication Software and Networks pp. 119–126 (2021)

[6] Bekkar, M., Djemaa, H.K., Alitouche, T.A.: Evaluation measures for models assessment over imbalanced data sets. J Inf Eng Appl 3(10) (2013)

[7] Biradar, K.M., Gupta, A., Mandal, M., Vipparthi, S.K.: Challenges in time-stamp aware anomaly detection in traffic videos. arXiv preprint arXiv:1906.04574 (2019)

[8] Bouyeddou, B., Harrou, F., Kadri, B., Sun, Y.: Detecting network cyber-attacks using an integrated statistical approach. Cluster Computing 24(2), 1435–1453 (2021)

[9] Burns, A., Greene, B.R., McGrath, M.J., O'Shea, T.J., Kuris, B., Ayer, S.M., Stroiescu, F., Cionca, V.: Shimmer™–a wireless sensor platform for noninvasive biomedical research. IEEE Sensors Journal 10(9), 1527–1534 (2010)

[10] Capio, C.M., Sit, C.H., Eguia, K.F., Abernethy, B., Masters, R.S.: Fundamental movement skills training to promote physical activity in children with and without disability: A pilot study. Journal of Sport and Health Science 4(3), 235–243 (2015)

[11] Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM computing surveys (CSUR) 41(3), 1–58 (2009)

[12] Chen, K., Zhang, D., Yao, L., Guo, B., Yu, Z., Liu, Y.: Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities. ACM Computing Surveys (CSUR) 54(4), 1–40 (2021)

[13] Christodoulou, V., Bi, Y., Zhao, G.: A fuzzy inspired approach to seismic anomaly detection. In: International Conference on Knowledge Science, Engineering and Management. pp. 575–587. Springer, Chongqing, China (2015)

[14] Cus, F., Zuperl, U., Gecevska, V.: High speed end-milling optimisation using particle swarm intelligence. Journal of Achievements in Materials and Manufacturing Engineering 22(2), 75–78 (2007)

[15] Ding, N., Prasad, K., Lie, T.: Design of a hybrid energy management system using designed rule-based control strategy and genetic algorithm for the series-parallel plug-in hybrid electric vehicle. International Journal of Energy Research 45(2), 1627–1644 (2021)

[16] Dua, N., Singh, S.N., Semwal, V.B.: Multi-input cnn-gru based human activity recognition using wearable sensors. Computing pp. 1–18 (2021)

[17] Espíndola, R.P., Ebecken, N.F.: On extending f-measure and g-mean metrics to multi-class problems. WIT Transactions on Information and Communication Technologies 35 (2005)

[18] Etumusei, J., Carracedo, J.M., McClean, S.: Novel martingale approaches for change point detection. In: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020). pp. 2113–27. Springer (2020)

[19] Fahad, L.G., Tahir, S.F.: Activity recognition and anomaly detection in smart homes. Neurocomputing 423, 362–372 (2021)

[20] Fernández, P.M., Sanchís, I.V., Yepes, V., Franco, R.I.: A review of modelling and optimisation methods applied to railways energy consumption. Journal of Cleaner Production 222, 153–162 (2019)

[21] Giambelluca, D., Cannella, R., Vernuccio, F., Comelli, A., Pavone, A., Salvaggio, L., Galia, M., Midiri, M., Lagalla, R., Salvaggio, G.: Pi-rads 3 lesions: role of prostate mri texture analysis in the identification of prostate cancer. Current problems in diagnostic radiology 50(2), 175–185 (2021)

[22] Hamdia, K.M., Zhuang, X., Rabczuk, T.: An efficient optimization approach for designing machine learning models based on genetic algorithm. Neural Computing and Applications 33(6), 1923–1933 (2021)

[23] Harush, S., Meidan, Y., Shabtai, A.: Deepstream: Autoencoder-based stream temporal clustering and anomaly detection. Computers & Security 106, 102276 (2021)

[24] Ho, S.S., Wechsler, H.: A martingale framework for detecting changes in data streams by testing exchangeability. IEEE transactions on pattern analysis and machine intelligence 32(12), 2113–2127 (2010)

[25] Hosseini, S.S., Noorossana, R.: Performance evaluation of ewma and cusum control charts to detect anomalies in social networks using average and standard deviation of degree measures. Quality and Reliability Engineering International 34(4), 477–500 (2018)

[26] Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. IEEE transactions on pattern analysis and machine intelligence 24(7), 881–892 (2002)

[27] Katiyar, S.: A comparative study of genetic algorithm and the particle swarm optimization. International Journal of Technology 2(2), 21–24 (2010)

[28] Kawala-Sterniuk, A., Podpora, M., Pelc, M., Blaszczyszyn, M., Gorzelanczyk, E.J., Martinek, R., Ozana, S.: Comparison of smoothing filters in analysis of eeg data for the medical diagnostics purposes. Sensors 20(3), 807 (2020)

[29] Khan, N., McClean, S., Zhang, S., Nugent, C.: Optimal parameter exploration for online change-point detection in activity monitoring using genetic algorithms. Sensors 16(11), 1784 (2016)

[30] Kim, C., Batra, R., Chen, L., Tran, H., Ramprasad, R.: Polymer design using genetic algorithm and machine learning. Computational Materials Science 186, 110067 (2021)

[31] Larsen, J.J., Pedersen, S.S., Foged, N., Auken, E.: Suppression of very low frequency radio noise in transient electromagnetic data with semi-tapered gates. Geoscientific Instrumentation, Methods and Data Systems 10(1), 81–90 (2021)

[32] Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49(4), 764–766 (2013)

[33] Li, J., Izakian, H., Pedrycz, W., Jamal, I.: Clustering-based anomaly detection in multivariate time series data. Applied Soft Computing 100, 106919 (2021)

[34] Lin, H.C., Wang, L.L., Yang, S.N.: Automatic determination of the spread parameter in gaussian smoothing. Pattern Recognition Letters 17(12), 1247–1252 (1996)

[35] Liu, Y., Ding, S., Jia, W.: A novel prediction method of complex univariate time series based on k-means clustering. Soft Computing 24(21), 16425–16437 (2020)

[36] Livieris, I.E., Stavroyiannis, S., Pintelas, E., Pintelas, P.: A novel validation framework to enhance deep learning models in time-series forecasting. Neural Computing and Applications 32(23), 17149–17167 (2020)

[37] Majhi, S.K.: Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection. Evolutionary intelligence 14(1), 35–46 (2021)

[38] Malczewski, J.: GIS and multicriteria decision analysis. John Wiley & Sons (1999)

[39] Maleki, N., Zeinali, Y., Niaki, S.T.A.: A k-nn method for lung cancer prognosis with the use of a genetic algorithm for feature selection. Expert Systems with Applications 164, 113981 (2021)

[40] Mirjalili, S.: Genetic algorithm. In: Evolutionary algorithms and neural networks, pp. 43–55. Springer (2019)

[41] Mishra, S., Sagban, R., Yakoob, A., Gandhi, N.: Swarm intelligence in anomaly detection systems: an overview. International Journal of Computers and Applications 43(2), 109–118 (2021)

[42] Mo, B., Ma, Z., Koutsopoulos, H., Zhao, J.: Calibrating route choice for urban rail system: A comparative analysis using simulation-based optimization methods. In: Transportation Research Board 99th Annual Meeting (2020)

[43] Mokhtari, S., Abbaspour, A., Yen, K.K., Sargolzaei, A.: A machine learning approach for anomaly detection in industrial control systems based on measurement data. Electronics 10(4), 407 (2021)

[44] Nguyen, G.H., Bouzerdoum, A., Phung, S.L.: Learning pattern classification tasks with imbalanced data sets. Pattern recognition pp. 193–208 (2009)

[45] Noto, K., Brodley, C., Slonim, D.: Frac: a feature-modeling approach for semi-supervised and unsupervised anomaly detection. Data mining and knowledge discovery 25(1), 109–133 (2012)

[46] Poli, R., Kennedy, J., Blackwell, T.: Particle swarm optimization. Swarm intelligence 1(1), 33–57 (2007)

[47] Ray, S.N., Chattopadhyay, S.: Analyzing surface air temperature and rainfall in univariate framework, quantifying uncertainty through shannon entropy and prediction through artificial neural network. Earth Science Informatics 14(1), 485–503 (2021)

[48] Raya, M.A.D., Sison, L.G.: Adaptive noise cancelling of motion artifact in stress ecg signals using accelerometer. In: Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society][Engineering in Medicine and Biology. vol. 2, pp. 1756–1757. IEEE (2002)

[49] Romero, M., Miranda, J.M., Montes, H.A., Acosta, J.C.: A statistical measuring system for rainbow trout. In: Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). p. 1. The Steering Committee of The World Congress in Computer Science, Computer … (2014)

[50] Roy, B.: The optimisation problem formulation: criticism and overstepping. Journal of the Operational Research Society 32(6), 427–436 (1981)

[51] Safaldin, M., Otair, M., Abualigah, L.: Improved binary gray wolf optimizer and svm for intrusion detection system in wireless sensor networks. Journal of ambient intelligence and humanized computing 12(2), 1559–1576 (2021)

[52] Sarker, I.H.: Machine learning: Algorithms, real-world applications and research directions. SN Computer Science 2(3), 1–21 (2021)

[53] Scholz, M., Haase, R., Schriefer, D., Voigt, I., Ziemssen, T.: Electronic health interventions in the case of multiple sclerosis: From theory to practice. Brain Sciences 11(2), 180 (2021)

[54] Shabir, S., Singla, R.: A comparative study of genetic algorithm and the particle swarm optimization. International Journal of Electrical Engineering 9(2016), 215–223 (2016)

[55] Shobande, O.A.: Is climate change a monetary phenomenon? evidence from time series analysis. International Journal of Sustainable Development & World Ecology pp. 1–13 (2021)

[56] Siino, M., Scudero, S., D'Alessandro, A.: Stochastic models for radon daily time series: seasonality, stationarity, and long-range memory detection. Frontiers in Earth Science (2020)

[57] Simoncini, D., Zhang, K.Y.: Population-based sampling and fragment-based de novo protein structure prediction (2019)

[58] Stokes, M.: Physical management in neurological rehabilitation. Elsevier Health Sciences (2004)

[59] Sun, S., Cao, Z., Zhu, H., Zhao, J.: A survey of optimization methods from a machine learning perspective. IEEE transactions on cybernetics 50(8), 3668–3681 (2019)

[60] Vovk, V., Nouretdinov, I., Gammerman, A.: Testing exchangeability on-line. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03). pp. 768–775 (2003)

[61] Wang, F., Zhang, H., Zhou, A.: A particle swarm optimization algorithm for mixed-variable optimization problems. Swarm and Evolutionary Computation 60, 100808 (2021)

[62] Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., Deng, S.H.: Hyperparameter optimization for machine learning models based on bayesian optimization. Journal of Electronic Science and Technology 17(1), 26–40 (2019)

[63] Xuan, X., Lo, D., Xia, X., Tian, Y.: Evaluating defect prediction approaches using a massive set of metrics: An empirical study. In: Proceedings of the 30th Annual ACM Symposium on Applied Computing. pp. 1644–1647 (2015)

[64] Zhang, H., Nian, K., Coleman, T.F., Li, Y.: Spectral ranking and unsupervised feature selection for point, collective, and contextual anomaly detection. International Journal of Data Science and Analytics 9(1), 57–75 (2020)

[65] Zhang, S., Galway, L., McClean, S., Scotney, B., Finlay, D., Nugent, C.D.: Deriving relationships between physiological change and activities of daily living using wearable sensors. In: International Conference on Sensor Systems and Software. pp. 235–250. Springer (2010)

[66] Zhou, B., Yang, Y., Zhang, Y., Gou, X., Cheng, B., Wang, J., Li, L.: Magnetic field data processing methods of the china seismo-electromagnetic satellite. Earth and Planetary Physics 2(6), 455–461 (2018)

[67] Zhou, Y., Ren, H., Li, Z., Wu, N., Al-Ahmari, A.M.: Anomaly detection via a combination model in time series data. Applied Intelligence pp. 1–14 (2021)

Author Biographies

Mr Jonathan Etumusei received a M.Sc.in Software development from Ulster University, Magee, United Kingdom. He is currently working toward a PhD degree at the School of Computing, University of Ulster, Jordanstown. His research interests include data mining, optimisation and anomaly detection in time series.

Dr Jorge Martínez Carracedo is a Lecturer in Computing Science at Ulster University. His main background is in mathematics and he has specialized in statistics and algebra, the field in which he developed his main research during his PhD. Dr Martinez is a co-investigator in BTIIC (BT Ireland Innovation Center), a major partnership between BT and Ulster University that is trying to bring together Telecommunication, IoT and Artificial Intelligence. Before his academic appointment, he worked as a Data Programmer for the pharmaceutical industry and later held a Research Associate Fellowship at Ulster University.

Prof. Sally McClean is Professor of Mathematics at Ulster University, and has previously obtained a M.A. in Mathematics from Oxford University, a M.Sc. in Mathematical Statistics and Operational Research from Cardiff University and a Ph.D. in Mathematics from Ulster University. She has over 400 publications, with research interests in Data Analytics, Applied Probability, Multivariate Statistical Analysis, Process Mining and applications to Healthcare and Computer Science. She is currently working in the BT Ireland Innovation Centre (BTIIC) which is a partnership between BT and Ulster University working at the intersection of Artificial Intelligence, and the Internet of Things.