IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 14, 2021, accepted August 25, 2021, date of publication August 27, 2021, date of current version September 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3108393

Extracting Most Discriminative Features on Transient Multivariate Time Series by Bi-Mode Hybrid Feature Selection Scheme for Transient Stability Prediction

SEYED ALIREZA BASHIRI MOSAVI
Imam Khomeini International University-Buin Zahra Higher Education Center of Engineering and Technology, Buin Zahra, Qazvin 3414896818, Iran
e-mail: a.bashiri@bzeng.ikiu.ac.ir

ABSTRACT Real-time transient stability assessment (TSA) of power systems based on mining system dynamic response has been widely considered by scholars. In this regard, extracting the most discriminative transient features (MDTFs) to achieve high-performance transient stability prediction (TSP) should be regarded as a fundamental issue in the transient learning strategy. In fact, MDTFs extraction is raised to make a trade-off between paradoxically intertwined indices, namely the accuracy and processing time of TSP. To this end, we offer a bi-mode hybrid feature selection scheme called BMHFSS for extracting MDTFs in high dimensional transient multivariate time series (TMTS). First, we used the TMTS, which are effective features on TSA. Next, the trajectory-based filter-wrapper mode (TFWM) is applied on TMTS to surmount the curse of dimensionality in two phases. In the filter phase, statistical and intrinsic characteristics of the TMTS in the form of agglomerative hierarchical clustering (AHC) are measured, and relevant TMTS (RTMTS) is selected according to obtained weight. In the wrapper phase, the RTMTS is entered into the trihedral kernel-based approach, including both fuzzy imperialist competitive algorithm (FICA) and incremental wrapper subset selection (IWSS) to find the intersected most RTMTS (IMRTMTS). As a complementary step, the filter-wrapper scenario in point-based mode (PFWM) is conducted for selecting MDTFs per time series in IMRTMTS. Finally, the aggregated MDTFs (AMDTFs) are tested to verify their efficacy for TSP based on cross-validation. The results show that the proposed framework has prediction accuracy greater than 98 % and a processing time of 52.94 milliseconds for TSA.

INDEX TERMS Fuzzy imperialist competitive algorithm (FICA), most discriminative transient features (MDTFs), support vector machine (SVM), transient stability assessment (TSA).

I. INTRODUCTION

The public vital utilities like water, telecom, natural gas, transportation, and oil reached conjunction with electric power. In fact, electricity supply securely and adequately called grid reliability plays the pivot role to keep the normal operation in interdependent infrastructure. However, the power outage raised due to the problem of quality and resiliency is the main concern for the senior manager at power energy developments. Hence, changing the power energy landscape via electricity restructuring is emerged as a necessary paradigm to achieve a reliable power supply. One of the important aspects of the restructured system is to be equipped with the real-time advanced statistic dashboard mounted on a wide-area monitoring system (WAMS) like phasor measurement units (PMUs). Such a PMU-assisted metering framework makes possible the awareness-based actions by power system operators (PSOs) via joint synchronization status-reported (JSSr) of power system components toward power system stability [1]. One of the significant issues in power system stability studies is addressing the JSSr of power system components under large and sudden disturbance, namely transient stability assessment (TSA) [2]. Since fast-occurred dynamic responses at high sampling frequency are the dominant property in transient space, conducting timely ancillary services accompanied with fast detection of instability is an essential task [3]. However, such time-oriented transient analysis (TA) can affect the accuracy of TSA. In fact, finding an

The associate editor coordinating the review of this manuscript and approving it for publication was Nagesh Prabhu.

approach that makes a trade-off between paradoxically intertwined indices, namely accuracy and the processing time, is the most significant concern on TSA. To this end, data mining (DM) technology has been considered by scholars as a way out of this challenge. The DM offers the feature subset selection (FSS) process for compacting the curse of dimensionality in transient space. In fact, the FSS process by selecting optimal features against high dimensional transient space shrouded the mentioned challenge and balanced the inconsistent indices toward high performance on TSA. Hence, the feature selection problem (FSP) as a joint study has been considered by scholars in transient processes in power systems and machine learning scope. Making the literature review on the FSP-based TSA studies shows this fact that the FSS applied by scholars on transient space including two approaches: 1) information theory/filter-based approach; for example, in Reference [4], the extended Relief-based feature selection algorithm called ReliefF finds the most sensitive features via relevance index for monitoring rotor fault on induction motors. Also, the minimum-redundancy and maximum-relevance (mRMR) applied on features related to power and angle for large-scale power systems TSA have been considered in [5], [6]. In Reference [7], the fast correlation-based filter method (FCBF) to eliminate irrelevant features is applied for the total transfer capability calculation considering static security, static voltage stability, and transient stability. Also, measuring correlations between variables via partial mutual information (PMI) and Pearson correlation coefficient (PCC) has been considered for selecting key features on TSA in [8], and 2) filter-wrapper approach; In Reference [9], the optimal features called global trajectory clusters feature subset (GTCFS) are selected from the large observations of rotor angle and voltage magnitude swing curves by the filter-wrapper approach in the form of the Relief-support vector machine (SVM) scenario for TSP. Also, a two-stage feature selection method including normalized mutual information (NMI) and binary particle swarm optimization (BPSO) has been considered to select the final optimal feature subset for TSP in [10].

Regardless of proposed well-suited solutions against the FSP in high dimensional transient space; in this paper, designing an inclusive FSS scheme is on the agenda regarding the following two aspects:

a) In previous studies on FSP-based TSA, selecting optimal features is triggered via point-based data type characteristic of dynamic responses. Regardless of the proper efficacy of point-oriented FSS, the exploration of the intrinsic traits into excursions of transient multivariate time series (TMTS) is an unvetted issue in the feature selection mechanism. In fact, ignoring the dynamic behavior per univariate trajectory-face of TMTS caused to prevent surviving optimal-blurred features. Hence, the necessity of regarding the cross-effects of point and trajectory-oriented behavior of TMTS for selecting the most discriminative transient features (MDTFs) motivated us to follow this point.

b) Designing the novel FSS scheme compatible with the matter raised in terms of (a) aspect is addressed as a challenge that did not occur in FSP-based TA previous research. In fact, lack of consideration to the intrinsic characteristics in point and trajectory-face of transient responses in previous FSP-based TSA negatively affects the training and testing procedure of classification techniques, which leads to low accuracy on TSP. In other words, selecting the best-laid feature set is the necessary concern to achieve high-performance (time and accuracy) on TSP. This challenge can be solved via conducting the FSS process in the form of a bi-mode scenario. In fact, presenting the framework that can consider the point and trajectory-face of transient response behavior in the feature selection process cohesively is intended in this paper.

According to what was mentioned above, one of the greatest challenges faced in TSA is large amounts of instances, and there are time series features, which consist of the long sequence with members from i^{th} to j^{th}. Therefore, the FSS is one of the useful ways of reducing the dimensionality of TMTS space. In fact, improvement in accuracy rate and reduction in processing time for TSA by exploiting optimal features are considered in this paper. As can be seen in Fig. 1, our proposed framework for TSA including three steps. In the first step, database construction by using power system component features on the New England-New York interconnection (NETS-NYPS) grid case is considered in the form of transient 28-variate time series. Next, we offer a bi-mode hybrid feature selection scheme (BMHFSS) as the cornerstone of our integrated research program on TSA, which is consists of two modes: (1) the trajectory-based filter-wrapper mode (TFWM) and (2) the point-based filter-wrapper mode (PFWM). The TFWM reducing the high dimensional space of TMTS in two phases: a) filter phase; information theory criteria are applied on TMTS via agglomerative hierarchical clustering (AHC) for selecting relevant TMTS (RTMTS), and b) in the wrapper phase, the RTMTS is entered into the trihedral kernel-based approach including both fuzzy imperialist competitive algorithm (FICA) and incremental wrapper subset selection (IWSS) to find the intersected most RTMTS (IMRTMTS). As the second mode of BMHFSS, the PFWM is conducted to select MDTFs per time series in IMRTMTS called aggregated MDTFs (AMDTFs). In the last step of the proposed framework, evaluating the efficacy of AMDTFs for TSA is considered by conducting a cross-validation technique.

The rest of the paper is organized as follows: we elaborate the BMHFSS strategy based on the dual-mode in Section 2. Experimental results of the proposed framework are presented in Section 3. Finally, the conclusion is presented in Section 4.

II. BI-MODE HYBRID FEATURE SELECTION SCHEME (BMHFSS)

A. OVERALL THE TFWM STERATEGY

The TFWM was embedded in BMHFSS for addressing the behavioral content of transient samples in a trajectory

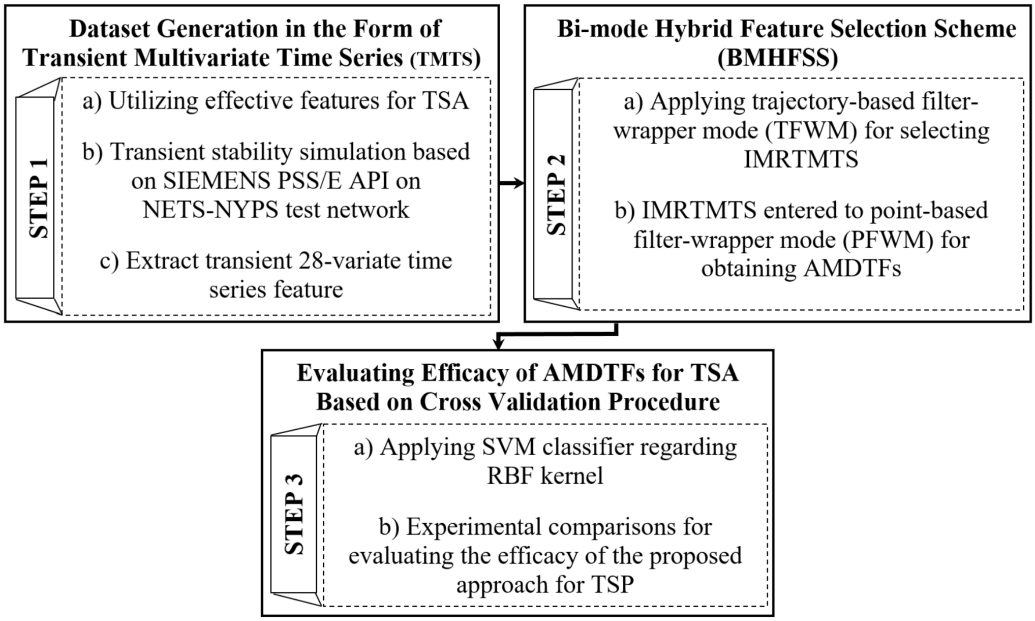


FIGURE 1. The proposed framework for TSA based on BMHFSS.

manner. To this end, information theory criteria and hyperplane-based learning strategy regarding this point were applied to TMTS for selecting RTMTS. The visual summary of the TFWM, including filter and wrapper phase, is shown in Fig. 2. The detailed descriptions of these phases are offered in the following paragraphs:

1) FILTER PHASE OF TFWM

In this phase, which is independent of the learning algorithm and according to statistical and intrinsic characteristics of the features, the dependency of each feature on the target class is measured, and the features are sorted according to their weight. Then, features with weights more than a threshold is selected and placed in the selected subset. Before proceeding to this phase, transient samples are clustered per univariate trajectory of TMTS (UToTMTS) using agglomerative hierarchical clustering (AHC) [11] to provide trajectory-based FSS implementation conditions. This section is considered a preprocessing step for the filter phase. One of the important issues in AHC clustering is the selection of suitable distance measurement for the computation of distances between trajectory-type objects. Historical distance measurement (e.g., Euclidean, Manhattan) that aligns the i^{th} point on one trajectory with the i^{th} point on another will produce a poor similarity in score distance and affect the learning model in terms of TSA. Dynamic time warping (DTW) [12] is a method for finding an optimal match between two given sequences. The DTW has advantages over the Euclidean distance in the aspect of its elastic and robust matching; therefore, it is used in forming the distance matrix of AHC. After the application of the preprocessing step, relevance, interdependence, and redundancy analysis based on information theory criteria are considered. In terms of relevance, symmetric uncertainty (SU) [13] based on mutual information (MI) and entropy is used to measure the amount of information shared by two variables as (1):

$$SU_{i,c}(F_i, C) = 2\frac{MI(F_i; C)}{H(F_i) + H(C)} \tag{1}$$

where F_i is the feature and C target class. Also, $MI(F_i; C)$ and $H(C)$ is considered as (2) and (3), respectively:

$$MI(X; Y) = H(X) - H(X|Y) \tag{2}$$

Let X be a discrete random variable and probability density function $p(x) = \Pr\{X = x\}$:

$$H(X) = -\sum_{x \in X} p(x) \log p(x) \tag{3}$$

In terms of interdependence and redundancy analysis (IR), redundancy and the interdependent ratio between two features are used, which is given as (4):

$$IR(i, j) = 2\frac{MI(F_i; C|F_j) - MI(F_i; C)}{H(F_i) + H(C)} \tag{4}$$

where $-1 \leq IR(i, j) \leq 1$.

The detailed description of the filter phase of the TFWM is considered as the following steps:

Step 1) filter phase of the TFWM contains a preliminary step for clustering time series data per UToTMTS. To this end, three different techniques of AHC, namely single- linkage [14], complete linkage [14], and average linkage [15] are applied to the TMTS dataset. First, the proximity matrix (PM)

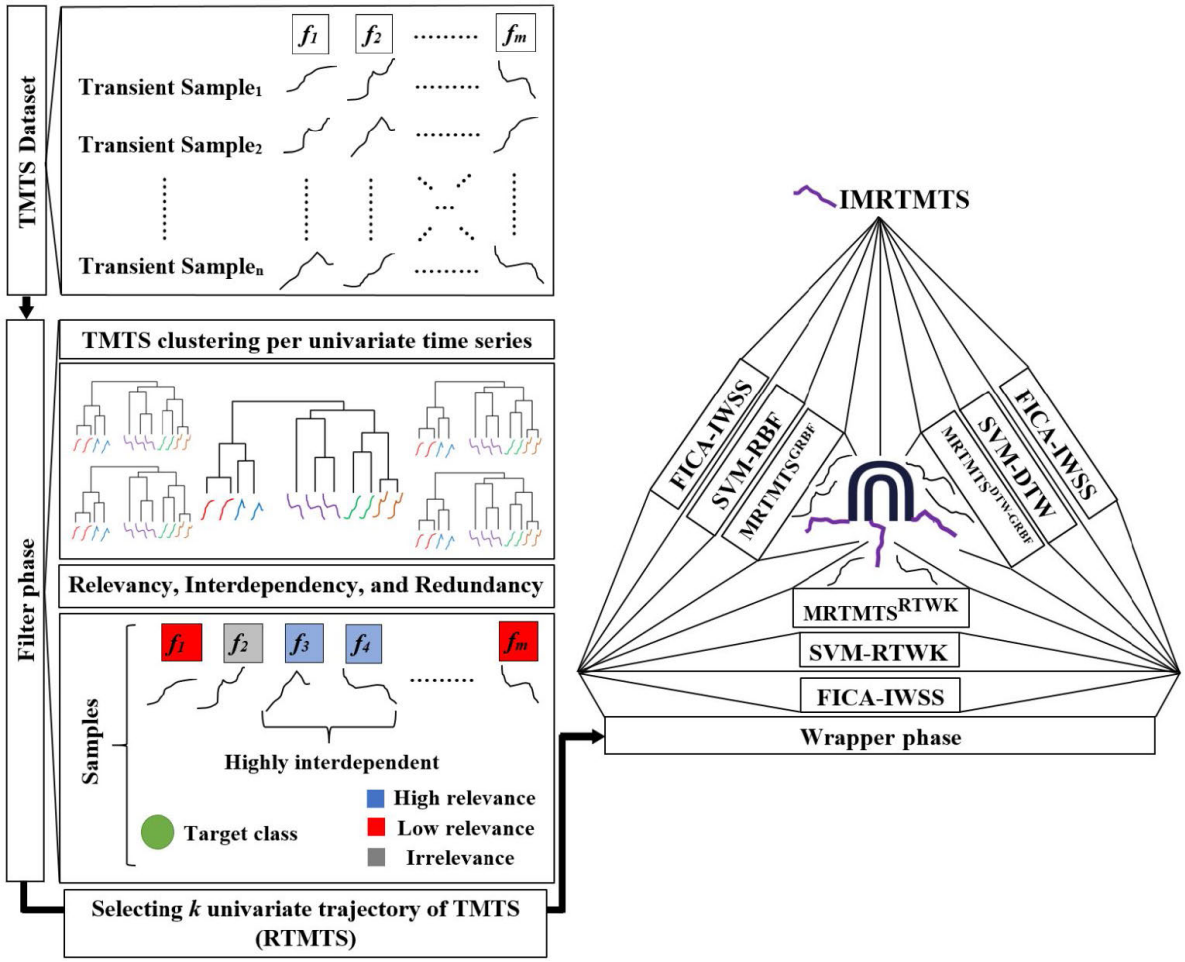


FIGURE 2. Visual summary of TFWM strategy.

is calculated by DTW:

$$PM(i,j) = \text{distance}^{DTW}(UToTMTS_i, UToTMTS_j) \quad (5)$$

where $1 \le i, j \le 28$ and measure is defined as (6):

$$\text{distance}^{DTW}(A_1^p, B_1^q) = d(a(p), b(q)) + Min \begin{pmatrix} \text{distance}^{DTW}(A_1^{p-1}, B_1^q) \\ \text{distance}^{DTW}(A_1^{p-1}, B_1^{q-1}) \\ \text{distance}^{DTW}(A_1^p, B_1^{q-1}) \end{pmatrix} \quad (6)$$

$d(\cdot,\cdot)$ is the L^p norm, and A_1^p is a sequence with a discrete index varying between 1 and p. Given the distance matrix PM, the linkage functions generate a hierarchical cluster tree returning the linkage information based on method type in a matrix $R^{[S|C|A]}$ as (7):

$$[R^{(S|C|A)}]_{n-1\times 3} = Linkage^{functions:S|C|A}(PM) \quad (7)$$

where $[R^{(S|C|A)}]$ is the matrix output by the linkage function which its size is $(n-1)$-by-3 matrix, n the numbers of observations in the TMTS, and $(S|C|A)$ are single, complete and average functions as clustering approach respectively.

To investigate natural cluster divisions per UToTMTS based on triple approach, cross-correlation coefficient or cophenetic correlation (CP) [16] is used as (8):

$$CP = \frac{\sum_{i<j}(x(i,j) - \bar{x}) \cdot (t(i,j) - \bar{t})}{\sqrt{\sum_{i<j}(x(i,j) - \bar{x})^2 \cdot \sum_{i<j}(t(i,j) - \bar{t})^2}} \quad (8)$$

where $x(i,j) = |X_i - X_j|$, the DTW distance between the i^{th} and j^{th} observations and $t(i,j)$ is the dendrogrammatic distance between the model points T_i and T_j. This distance is the height of the node at which these two points are first joined together. $\bar{x}$ is the $x(i,j)$ average, and $\bar{t}$ is the average of the $t(i,j)$. In fact, in natural cluster divisions, the linking of objects in the hierarchical cluster tree has a proper correlation with the distances between objects in the distance vector. The cophenetic correlation compares these two sets of values and calculates their correlation. The closer value to 1 shows the clustering is quite fit. In fact, after clustering all UToTMTS based on the triple linkage function, m CPs are obtained for each AHC technique. Thereafter, the average of CPs is

considered as the accuracy rate of each AHC technique for clustering the trajectory data.

Step 2) After the clustering process, in the first step of the filter phase, the weight per UToTMTS computed using SU measure (See Equation (1)) represents the degree of importance of the UToTMTS. By selecting UToTMTS with the highest SU called $UToTMTS^{h(1)}$, it is situated in the RTMTS subset as the $RTMTS^1$:

$$RTMTS^1 = UToTMTS^{h(1)} = Max(SU_{[UToTMTS^{1:28}]}) \quad (9)$$

Step 3) The injection of *IR* ratio (Equation. (4)) in SU amount is considered in this step. Based on (4), we have:

$$\begin{aligned} &SU^{New}_{[UToTMTS^{1:28}-UToTMTS^{h(1)}]} \\ &= W^{(1)} \cdot * \left(SU_{[UToTMTS^{1:28}-UToTMTS^{h(1)}]}\right); \\ &W^{(1)} \\ &= W^{(0)} \cdot * IR\left(UToTMTS^{1:28} - UToTMTS^{h(1)}, UToTMTS^{h(1)}\right) \end{aligned} \quad (10)$$

where $W^{(0)}$ for initializing the weight per UToTMTS (28-variate trajectory features) to 1 equally.

Now, $UToTMTS^2$ is selected and added to RTMTS subset:

$$\begin{aligned} RTMTS^2 &= UToTMTS^{h(2)} \\ &= Max(SU^{New}_{[UToTMTS^{1:28}-UToTMTS^{h(1)}]}) \end{aligned} \quad (11)$$

Step 4) We repeat this scenario to select the k^{th} member of RTMTS subset as follow:

$$\begin{aligned} &SU^{New}_{[UToTMTS^{1:28}-UToTMTS^{h(1):h(k-1)}]} \\ &= W^{(k-1)} \cdot * \left(SU_{[UToTMTS^{1:28}-UToTMTS^{h(1):h(k-1)}]}\right); \\ &W^{(k-1)} \\ &= W^{(k-2)} . * IR\left(UToTMTS^{1:28} - UToTMTS^{h(1):h(k-1)}, \right. \\ &\quad \left. UToTMTS^{h(k-1)}\right) \\ &RTMTS^k \\ &= UToTMTS^{h(k)} \\ &= Max(SU^{New}_{[UToTMTS^{1:28}-UToTMTS^{h(1):h(k-1)}]}) \end{aligned} \quad (12)\ (13)$$

After applying the filter phase of TFWM on TMTS, $[RTMTS]_{n\times k\times c}$ where n is the number of transient samples, k number of selected features (called $RTMTS^1$ to $RTMTS^k$), and c is the number of observed cycles after fault clearing; entered into the wrapper phase of TFWM equipped with hyperplane-based learning strategies.

2) WRAPPER PHASE OF TFWM

After completing the two-stage filter phase, selected RTMTS is entered as input into the wrapper phase. The wrapper phase is used to cover the shortcoming of the filter phase in ignoring the accuracy of the classifier. In this phase, both FICA and IWSS are applied to find the most relevant TMTS (MRTMTS) subset in the wrapper phase. The IWSS is one of the hybrid feature selection algorithms for the selection of the most relevant features [17]. Generally, in hybrid methods like IWSS, the weighting process is conducted on the features in the filter phase, and features are sorted according to their weight. Then, an incremental mechanism is used to select a subset of features. At first, the selected subset is empty, and in the first iteration, the feature that has the highest weight is added to the selected subset. Then, a classifier is trained based on the selected subset, and classification accuracy is saved as the best result. In subsequent iterations, the new features with high weight are added to the selected subset, and a classifier is trained based on it. If by adding a new feature, accuracy increased against the preceding subset accuracy, the feature has remained in the selected subset; otherwise, this attribute is removed, and the next feature is added (See Fig. 3).

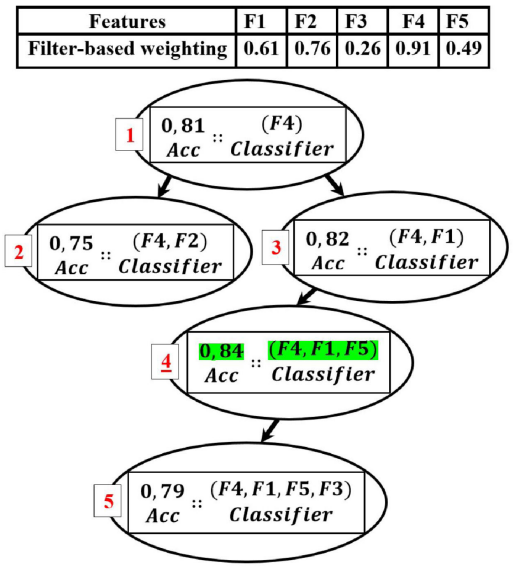


FIGURE 3. The IWSS algorithm.

The imperialist competitive algorithm (ICA) is one of the evolutionary techniques for solving optimization problems. ICA shows convergence and high speeds in comparison with other evolutionary algorithms (e.g., genetic algorithm (GA), simulated annealing (SA), etc.) in the presence of a large number of independent variables [18]. In recent years, more researchers working on ways to improve ICA and its efficiency in solving optimization problems, but weaknesses like falling into local optimum points are the main challenges of this algorithm. The FICA is an improved version of ICA [19] with higher convergence and speed rate than ICA and solves the mentioned challenges against optimization problems. To solve a given optimization problem using FICA, it can be assumed that there are N countries (or features), and each country represents the trajectory-face features in the RTMTS space. Among them, countries with the lowest cost according to optimization problems are considered imperialist, and the rest are considered colonies. FICA considers imperialist as a fuzzy set and colonies as a member of

imperialist communities that are associated with the specific membership function for each colony [19].

In the following example, the typical iteration of the wrapper phase according to Fig. 4 is offered for readers at a glance for a better understanding of how each step is formalized.

Example: Let us assume an iteration of wrapper phase:

E.1. WRAPPER PHASE; FICA Initialization

E.1.1. Formation of countries (See Fig. 5 (a)): N countries are created based on selected RTMTS of filter phase by weighted random selection.

E.1.2. Cost calculation: Cost per country as (14) is calculated using IWSS:

$$\begin{aligned}&\text{CountryCost}^i\\&= 1 - {}^{SVM}\text{Classification Accuracy}^{IWSS}_{Kerneltype}(country^i)\end{aligned} \quad (14)$$

where SVM in (14) is a robust classification technique that maximizes predictive accuracy without overfitting training data. SVM is originally designed to be a binary classification, and it follows the principle of structural risk minimization. SVM employs a separating hyperplane with low structural risk in the classification of data and is not linearly separable in feature space. The optimization problem and the constraints of SVM are defined according to (15):

$$\begin{aligned}a^* &= \arg\min_{\alpha} \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{k=1}^{l}\alpha_k;\\ 0 \le \alpha_i \le C, &\sum_{j=1}^{l}\alpha_i y_i = 0, \quad i, j = 1, \ldots, l\end{aligned} \quad (15)$$

where $K(x_i, x_j)$ for projecting TMTS space to a higher dimension space. The optimal separating hyperplane in feature space is determined by solving (16):

$$\begin{aligned}f(x) &= \text{sgn}\left(\sum_{i\in s}\alpha_i y_i K(x_i, x) + b\right);\\ b &= \frac{1}{s}\sum_{i\in s}\left[y_i - \sum_{j}\alpha_j y_j K(x_j, x_i)\right]\end{aligned} \quad (16)$$

For calculating CountryCost^i in (14), the accuracy index is used for the evaluation of classifier performance as (17):

$$Acc = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \quad (17)$$

where *Positive*: identified stable sample, *Negative*: identified unstable sample, *TP*: correctly identified stable sample, *TN*: correctly identified unstable sample, *FP*: incorrectly identified stable sample, and *FN*: incorrectly identified unstable sample.

E.2. WRAPPER PHASE; FICA main body (e.g., i^{th} iteration)

E.2.1. Determination of imperialists and colonies ((See Fig. 5 (b)): Countries are sorted based on country cost, then by defining threshold parameter countries with low cost are selected as imperialist (green-face), and the rest are considered as a colony (blue-face).

E.2.2. Construction of membership matrix ((See Fig. 5 (c)): In FICA, each colony belongs to each imperialist according to a membership function.

E.2.3. Absorption phase ((See Fig. 5 (d)): After calculating the membership degree of each colony per imperialist, colonies must move towards the position of imperialists. The higher degree of colony membership caused the j^{th} imperialist to have more impact on the colony movement. The number of features that should be transferred from all imperialists to colony i are computed according to the absorption phase equations. Fig. 5 (e) shows features per colony after implementation of the absorption phase.

E.2.4. Revolution phase: To implement the revolution phase, all countries (colony and imperialist countries) are sorted according to their cost. Then, a percentage of high-cost countries are discarded, and new countries are created (See E.1.1). Then, new countries are substituted with the removed countries.

According to the designed strategy in the wrapper phase, MRTMTS is selected based on the FICA-IWSS mechanism accompanied by the SVM classifier. An important point to note is that the SVM in FICA-IWSS used for cost calculation per country has an important embedded component, namely kernel type (See Equation (14)) that impacts the performance of SVM, which itself affect the TFWM capacity for selecting the MRTMTS (transitive relation). In fact, the choice of the efficient kernel function caused to constructs an optimal separating hyperplane against the TMTS space. Hence, we consider three efficient kernels plugged into SVM to feed the FICA-IWSS for selecting MRTMTS as follows:

(a) Standard Gaussian radial basis function (GRBF) [20]: GRBF kernel as $K(x, x')$ in (15) is defined as (18):

$$K(x, x') = \exp\left(-\frac{||x - x'||^2}{2\sigma^2}\right) \quad (18)$$

where $||x - x'||^2$ is squared Euclidean distance between the two time series feature.

(b) DTW in Gaussian RBF kernel (DTW-GRBF kernel) [21]: Since the DTW outperforms Euclidean distance in most cases because its matching is elastic and robust, it is tempting to substitute DTW distance for Euclidean distance in the Gaussian RBF kernel and plug it into SVM for sequence classification. Hence, the DTW-GRBF kernel as $K(x, x')$ in (15) is defined as (19):

$$K(x, x') = \exp\left(-\frac{\left[\text{distance}^{DTW}(A_1^p, B_1^q)\right]^2}{2\sigma^2}\right) \quad (19)$$

where;

$$\begin{aligned}&\text{distance}^{DTW}(A_1^p, B_1^q)\\&= d(a(p), b(q)) + Min\begin{pmatrix}\text{distance}^{DTW}(A_1^{p-1}, B_1^q)\\ \text{distance}^{DTW}(A_1^{p-1}, B_1^{q-1})\\ \text{distance}^{DTW}(A_1^p, B_1^{q-1})\end{pmatrix}\end{aligned}$$

(WRAPPER PHASE: PRELIMINARY STEP)

$RTMTS^k = \{UToTMTS^1, UToTMTS^2 \ldots, UToTMTS^k\}$; *% selected RTMTS based on two-stage filter phase entered to the preliminary step of wrapper phase.*

(1) for i=1 to k

(2) $ProbScoreUToTMTS[i] = SU(UToTMTS^i)/\sum_{j=1}^{k} SU(UToTMTS^j)$; *% The probability score per UToTMTS is calculated via retrieved SU value of filter phase. This score is used for weighted random choice in country creation.*

(3) end

(WRAPPER PHASE: FICA INITIALIZATION)

(4) $N_{SubRTMTS} = x$; *% Number of UToTMTS (SubRTMTS) assigned per country.*

(5) $N_{SubRTMTS}^{couns} = y$; *% Number of countries including SubRTMTS.*

(6) $N_{SubRTMTS}^{imp} = N_{SubRTMTS}^{couns} \times l$; *% Number of imperialist countries including SubRTMTS. e.g., l=0.2*

(7) $N_{SubRTMTS}^{col} = N_{SubRTMTS}^{couns} - N_{SubRTMTS}^{imp}$; *% Number of colony countries including SubRTMTS.*

(8) for i=1 to $N_{SubRTMTS}^{couns}$

(9) $Country_{SubRTMTS}[i] = datasample(RTMTS, N_{SubRTMTS}, without\ Replacement, ProbScoreUToTMTS[1:k])$; *% Selecting for typical country based on probability scores of RTMTS. The RTMT with a high probability score has more chance for selecting as a member of the country. The subset of RTMTS is selected per country.*

(10) $CountryCost_{SubRTMTS}[i] = 1 - {}^{SVM}Classification\ Accuracy_{Kernel\ type}^{IWSS}(Country_{SubRTMTS}[i])$; *% Cost calculation per country based on its SubRTMTS using IWSS and kernel type: GRBF || DTW-GRBF || RTWK.*

(11) end

(WRAPPER PHASE: FICA MAIN BODY)

(12) for iteration=1 to z

(13) $SortedCountriesCost[1:N_{SubRTMTS}^{couns}] = Ascending\ Sorting\ (CountryCost_{SubRTMTS}[1:N_{SubRTMTS}^{couns}])$ *% Countries sorted based on CountryCost in an ascending manner*

(14) for i=1 to $N_{SubRTMTS}^{imp}$

(15) $ImperialistCountry_{SubRTMTS}[i] = SortedCountriesCost[i]$; *% Imperialist countries array*

(16) end

(17) for i=1 to $N_{SubRTMTS}^{col}$

(18) $ColonyCountry_{SubRTMTS}[i] = SortedCountriesCost[i + Num_{imperialist}]$; *% Colony countries array*

(19) end

$\forall\, i \in \{1:Num_{colony}\}$ & $\forall\, j \in \{1:Num_{imperialist}\}$; we have:

(20) $NEoNC(i,j) = NumEle^{non-common}(ColonyCountry_{SubRTMTS}[i], ImperialistCountry_{SubRTMTS}[j])$; *% Number elements of non-common (NEoNC) SubRTMTS between imperialist countries and colony countries.*

(21) $NC(j) = Max(ImperialistCountry_{SubRTMTS}^{cost}[1:N_{SubRTMTS}^{imp}]) - ImperialistCountry_{SubRTMTS}^{cost}[j]$; *% Normalize the cost of imperialist countries.*

(22)
$$\begin{matrix} ColonyCountry[i] \\ ImperialistCountry\ [j] \end{matrix}\left(P; a, b\right) = \begin{cases} 0 & , x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2 & , a \le x \le \frac{a+b}{2} \\ 1 - 2\left(\frac{x-b}{b-a}\right)^2 & , \frac{a+b}{2} \le x \le b \\ 1 & , x \ge b \end{cases}$$ *% Formation of membership*

$\vdots$

FIGURE 4. Pseudocode of wrapper phase (FICA-IWSS) embedded to TFWM.

matrix; membership degree (MD) of each colony country to each imperialist countries. a and b are the parameters and x is:

$$P = Imp_j(Col_i) = \frac{NC(j)}{\sum NC[1:N^{imp}_{SubRTMTS}]} \times \left(1 - \frac{NEoNC(i,j)}{\sum NEoNC(i,[1:N^{imp}_{SubRTMTS}]}\right)$$

(23) ${}^{\#UToTMTS}t^{ImperialistCountry\,[j]}_{ColonyCountry[i]} = {}^{\#UToTMTS}T^{total}_{ColonyCountry[i]} \times \frac{NC(j)}{\sum NC[1:N^{imp}_{SubRTMTS}]}$ *% Absorption phase; Transfer #UToTMTS of imperialist j to colony i. In line (23)* ${}^{\#UToTMTS}T^{All\,ImperialistCountry}_{ColonyCountry[i]}$ *indicate the #UToTMTS that are transferred from all imperialists to colony i.*

$${}^{\#UToTMTS}T^{All\,ImperialistCountry}_{ColonyCountry[i]} = \frac{\sum_{j=1}^{N^{imp}_{SubRTMTS}} {}^{ColonyCountry[i]}_{ImperialistCountry\,[j]}\left(P;a,b\right) \times NEoNC(i,j)}{\sum_{j=1}^{N^{imp}_{SubRTMTS}} {}^{ColonyCountry[i]}_{ImperialistCountry\,[j]}\left(P;a,b\right)}$$

(24) $ColonyCountry_{SubRTMTS}[i] = ColonyCountry_{SubRTMTS}[i] \cup {}^{\#UToTMTS}t^{ImperialistCountry\,[1:N^{imp}_{SubRTMTS}]}_{ColonyCountry[i]}$ *% After the termination transforming process, the UToTMTS list of colony countries is updated.*

(25) $ColonyCountryCost_{SubRTMTS}[i] = 1 - {}^{SVM}Classification\ Accuracy^{IWSS}_{Kernel\ type}(ColonyCountry_{SubRTMTS}[i])$; *% Cost calculation on colony country based on its SubRTMTS using IWSS and kernel type: GRBF || DTW-GRBF || RTWK.*

(26) $Countries[1:N^{couns}_{SubRTMTS}] = [ColonyCountry_{SubRTMTS} \cup ImperialistCountry_{SubRTMTS}]$ *% Accession of countries for applying revolution phase.*

(27) $Countries[1:N^{couns}_{SubRTMTS}] = Ascending\ Sorting\ (CountriesCost\,[1:N^{couns}_{SubRTMTS}])$ *% Countries sorted based on CountryCost in an ascending manner*

(28) for i=$N^{couns}_{SubRTMTS}$ down to $N^{couns}_{SubRTMTS} \times h$ *% (revolution phase idea) Regardless of whether a country is an imperialist or a colony, the percentage (h factor) of high-cost countries is eliminated and the newly produced countries will replace them.*

(29) $Country_{SubRTMTS}[i] = datasample(RTMTS, N_{SubRTMTS}, without\ Replacement\ , ProbScoreUToTMTS[1:k])$;

(30) $CountryCost_{SubRTMTS}[i] = 1 - {}^{SVM}Classification\ Accuracy^{IWSS}_{Kernel\ type}(Country_{SubRTMTS}[i])$;

(31) end

(32) end of iteration loop of wrapper phase main body

(33) $\text{MRTMTS}^{Kernel\ Type} = {}^{Kernel\ Type}_{FICA-IWSS}CountryCost^{lowest}_{SubRTMTS}$ *% SubRTMTS of Country with the lowest cost is selected as MRTMTS with regarding kernel type plugged into FICA-IWSS.*

****

TRIHEDRAL WRAPPER PHASE

****

$IMRTMTS = \text{MRTMTS}^{GRBF} \cap \text{MRTMTS}^{DTW-GRBF} \cap \text{MRTMTS}^{RTWK}$ *% Generally, the above-mentioned wrapper scenario is exerted three times in the presence of triple kernels (GRBF, DTW-GRBF, and RTWK), and kernel-specific MRTMTS is obtained (namely;* $MRTMTS^{GRBF}, MRTMTS^{DTW-GRBF}$, *and* $MRTMTS^{RTWK}$*). Finally, intersecting these three results called IMRTMTS entered to PFW scenario.*

FIGURE 4. ***(Continued.)*** **Pseudocode of wrapper phase (FICA-IWSS) embedded to TFWM.**

(c) Recursive time warp kernel, called RTWK [22]: The positive definite recursive elastic kernels outperform the distance substituting kernels for the classical elastic distances. A function $\langle\cdot,\cdot\rangle : U \times U \rightarrow R$ is called a recursive time warp kernel (RTWK) if, for any pair of sequences A_1^p, B_1^q there exists a function $f : R \rightarrow R$ such that the following recursive equation is satisfied:

$$\langle A_1^p, B_1^q\rangle = \sum \begin{cases} \left\langle A_1^{p-1}, B_1^q\right\rangle f\,(\Gamma(A(p) \rightarrow \Lambda)) \\ \left\langle A_1^{p-1}, B_1^{q-1}\right\rangle f\,(\Gamma(A(p) \rightarrow B(q))) \\ \left\langle A_1^{p}, B_1^{q-1}\right\rangle f\,(\Lambda \rightarrow B(q)) \end{cases} \quad (20)$$

(a) Formation of countries

Country	RTMTS (F_i)				
1	F1	F4	F7	F9	F14
2	F12	F23	F17	F21	F9
3	F1	F20	F14	F17	F4
4	F1	F4	F17	F20	F7
5	F20	F23	F22	F1	F4

(b) Determination of imperialists and colonies

Country	RTMTs (F_i)			
1 imp	F1	F4	F7	F14
2 col	F12	F23	F17	
3 imp	F1	F20	F14	
4 col	F1	F4	F17	
5 imp	F20	F23	F22	

(c) Membership matrix

	imp^1	imp^2	imp^3
$F_{imp}(col^1)$	0.12	0.32	0.09
$F_{imp}(col^2)$	0.05	0.18	0.45

(d) No. transitional RTMTS (F_i) for each empire

	imp^1	imp^2	imp^3
$Count_1$ (imp ^{j})	1	1	0
$Count_2$ (imp ^{j})	0	1	2

(e) Colony RTMTS (F_i) after implementing absorption phase

	Before absorption :	After absorption
$Colony^1$	F12-F23-F17 :	F12-F23-F17-F1-F20
$Colony^2$	F1-F4-F17 :	F1-F4-F17-F14-F20-F22

FIGURE 5. The wrapper phase (FICA-IWSS) of TFWM in i^{th} iteration.

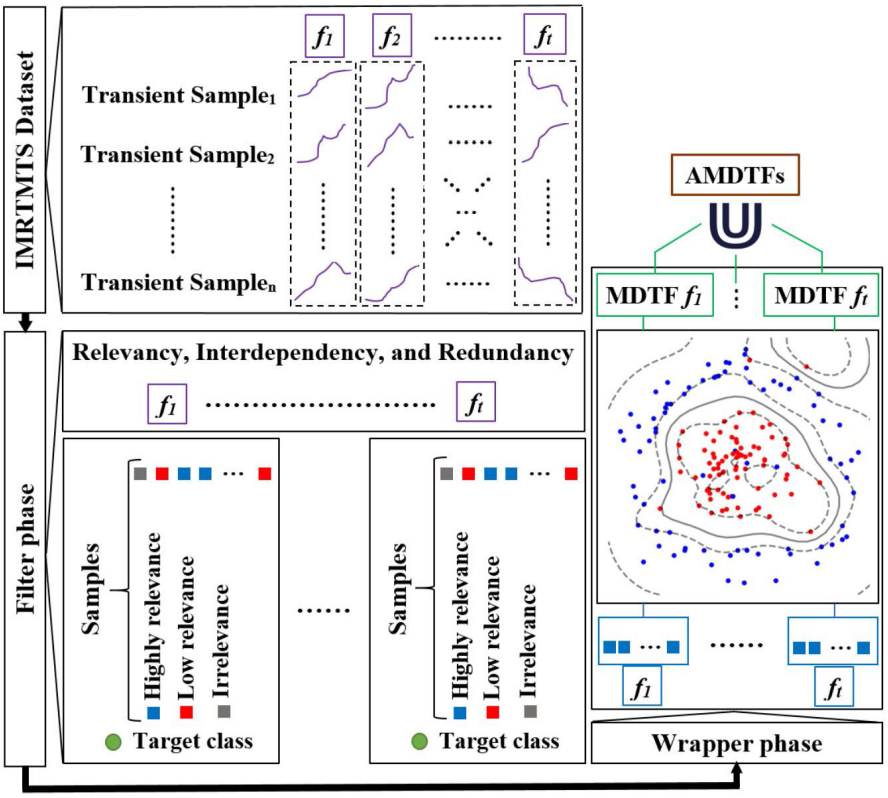


FIGURE 6. Visual summary of PFWM strategy.

Let U be the set of finite sequences (sequences or time series); $U = \{A_1^p | p \in \mathbb{N}\}$. A_1^p is a sequence with a discrete index varying between 1 and p. Also, $\Gamma(h)$ is the cost function for edit operation.

According to what was mentioned above, we confronted the trihedral wrapper phase based on triple kernels for conducting wrapper phase (trihedral FICA-IWSS) of the TFWM. In fact, three categories of MRTMTS based on kernel

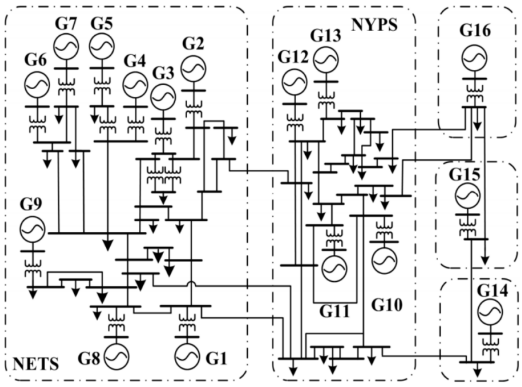


FIGURE 7. Single line diagram of NETS-NYPS grid case.

substitution in FICA-IWSS are obtained which is called $\text{MRTMTS}^{\text{GRBF}}$, $\text{MRTMTS}^{\text{DTW-GRBF}}$, and $\text{MRTMTS}^{\text{RTWK}}$ (See Fig. 2). Hence, the intersection of obtained results in three categories was considered as intersected MRTMTS. Finally, by performing the wrapper phase of the TFWM based on RTMTS, the obtained IMRTMTS are entered into the PFWM.

B. OVERALL THE PFWM STRATEGY

As the complementary mode of the proposed BMHFSS, the obtained IMRTMTS (obtained from the TFWM strategy) as input is fed to the PFWM strategy. In fact, for selecting MDTFs, the point-based filter-wrapper strategy is applied per trajectory of IMRTMTS individually. Finally, after obtaining MDTFs for each trajectory of IMRTMTS, the aggregated MDTFs (AMDTFs) are selected as final optimum features. The graphical abstract of the PFWM strategy is shown in Fig. 6. As can be seen in Fig. 6, the filter and wrapper phases are applied on each trajectory-face member of the IMRTMTS subset, and MDTFs per trajectory are selected. Finally, by aggregating the trajectory-specific MDTFs, the AMDTFs are obtained as final survived features. We elaborated on the PFWM strategy in the following steps:

After conducting TFWM strategy, we have:

$$IMRTMTS^{t} = \{UToTMTS^{1}, UToTMTS^{2}, \ldots, UToTMTS^{t}\} \tag{21}$$

The following steps are conducted for each *UToTMTS* of *IMRTMTS* subset. For i^{th} *UToTMTS*, we have:

Step 1) The weight per point-feature (PF) of $UToTMTS^{i}$ computed using SU measure (See Equation (1)) represents the degree of importance of the PF of $UToTMTS^{i}$. By selecting PF with the highest SU called $PF^{h(1)}$, it is situated in the preliminary MDTFs subset of $UToTMTS^{i}$ as the *pMDTFs*.

$$UToTMTS_{i}^{PF^{h(1)}} = Max\left(SU_{[UToTMTS_{i}^{PF^{1:s}}]}\right);$$
$$(1 \le i \le t),\ s \text{ number of point-feature in} UToTMTS^{i} \tag{22}$$

TABLE 1. Transient multivariate time series features (28-variate).

	Math formula
$F_1^{t_m}$	$= Max([\frac{PELEC_i}{P\max_i}]^{i=1:N_{genbus}})$
$F_2^{t_m}$	$= Var([\frac{PELEC_i}{P\max_i}]^{i=1:N_{genbus}})$
$F_3^{t_m}$	$= Max([\frac{QELEC_i}{Q\max_i}]^{i=1:N_{genbus}})$
$F_4^{t_m}$	$= Min([\frac{QELEC_i}{Q\max_i}]^{i=1:N_{genbus}})$
$F_5^{t_m}$	$= Var([\frac{QELEC_i}{Q\max_i}]^{i=1:N_{genbus}})$
$F_6^{t_m}$	$= Max([VOLT_i]^{i=1:N_{bus}})$
$F_7^{t_m}$	$= Var([VOLT_i]^{i=1:N_{bus}})$
$F_8^{t_m}$	$= Max([VANGLE_i]^{i=1:N_{bus}});\ slack\ bus = 0$
$F_9^{t_m}$	$= Min([VANGLE_i]^{i=1:N_{bus}});\ slack\ bus = 0$
$F_{10}^{t_m}$	$= Var([VANGLE_i]^{i=1:N_{bus}}); slack\ bus = 0$
$F_{11}^{t_m}$	$= Max(abs([VANGLE_i - VANGLE_j]^{i,j=1:N_{bus}}))$
$F_{12}^{t_m}$	$= Mean(abs([VANGLE_i - VANGLE_j]^{i,j=1:N_{bus}}))$
$F_{13}^{t_m}$	$= Var(abs([VANGLE_i - VANGLE_j]^{i,j=1:N_{bus}}))$
$F_{14}^{t_m}$	$= \frac{\sum_{i=1}^{N_{busgen}} QLOAD_i}{\sum_{i=1}^{N_{busgen}} QELEC_i}$
$F_{15:28}^{t_m}$	$= Gradient\ of\ features\ F_1 - F_{14}$

Symbol: t_m= moments in simulation time [1: s], $N_{bus\ gen}$= number of bus generator in test case, PELEC= machine electrical power (pu), P_{max}= maximum amount of machine electrical power, QELEC=machine reactive power, Qmax= maximum amount of machine reactive power, Qload= reactive power consumption, Volt= bus pu voltages, N_{bus}= number of buses in test case, VANGLE= voltage phase angle, Var= variance, Max= maximum, Min= minimum, Mean= average.

So now we have one point-feature in the preliminary MDTFs subset of $UToTMTS^i$;

$$UToTMTS_i^{\{pMDTFs\}^{len:1\#}} \leftarrow \{UToTMTS_i^{PF^{h(1)}}\};\ len = length.$$

Step 2) The injection of IR ratio (4) in SU amount is considered in this step. Based on (4), we have:

$$SU^{New}_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1)}}]} = W^{(1)} .* \left(SU_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1)}}]} \right);$$

$$W^{(1)} = W^{(0)} .* IR \left(UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1)}},\ UToTMTS_i^{PF^{h(1)}} \right) \quad (23)$$

Now, the second point-feature with high value is selected and added to preliminary MDTFs subset of $UToTMTS^i$:

$$UToTMTS_i^{PF^{h(2)}} = Max \left(SU^{New}_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1)}}]} \right) \quad (24)$$

So far we have two point-feature in the preliminary MDTFs subset of $UToTMTS^i$;

$$UToTMTS_i^{\{pMDTFs\}^{len:2\#}} \leftarrow \{UToTMTS_i^{PF^{h(1)}}, UToTMTS_i^{PF^{h(2)}}\}$$

Step 3) We repeat this scenario to select the r^{th} member of preliminary MDTFs subset of $UToTMTS^i$ as follow:

$$SU^{New}_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1):h(r-1)}}]} = W^{(r-1)} .* \left(SU_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1):h(r-1)}}]} \right);$$

$$W^{(r-1)} = W^{(r-2)} .* IR \left(UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1):h(r-1)}},\ UToTMTS_i^{PF^{h(r-1)}} \right) \quad (25)$$

$$UToTMTS_i^{PF^{h(r)}} = Max \left(SU^{New}_{[UToTMTS_i^{PF^{1:s}} - UToTMTS_i^{PF^{h(1):h(r-1)}}]} \right) \quad (26)$$

Here we have r point-feature in the preliminary MDTFs subset of $UToTMTS^i$:

$$UToTMTS_i^{\{pMDTFs\}^{len:r\#}} \leftarrow \{UToTMTS_i^{PF^{h(1)}}, UToTMTS_i^{PF^{h(2)}}, \ldots, UToTMTS_i^{PF^{h(r)}}\}$$

Step 4) After obtaining $UToTMTS_i^{\{pMDTFs\}^{len:1\#}}$, $UToTMTS_i^{\{pMDTFs\}^{len:2\#}}$, $UToTMTS_i^{\{pMDTFs\}^{len:3\#}}$, $\ldots$, $UToTMTS_i^{\{pMDTFs\}^{len:r\#}}$; each point-feature package ({1#}, {2#} to {r#}) belongs to $UToTMTS^i$ are fed to GRBF-SVM for measuring performance of them. Next, the high-performance point-feature (PF) package among packages are selected in the form of MDTFs per $UToTMTS^i$ as (27), shown at the bottom of the next page.

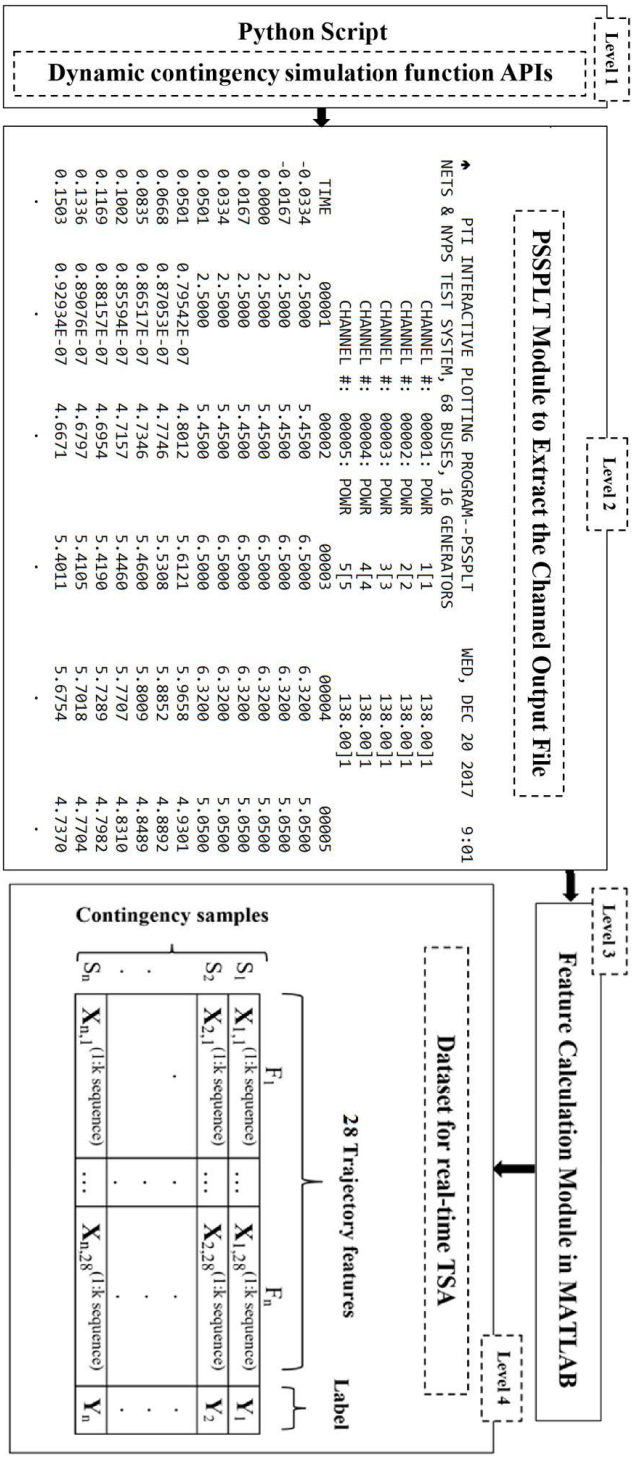


FIGURE 8. Dynamic contingency simulation recording workflow.

In Eq. (27), the matrix Z contains the obtained accuracy grouped by selected p MDTFs with different lengths ({1#}, {2#} to {r#} belongs to $UToTMTS^i$. In fact, by applying pMDTFs with length of {1#}, {2#} to {r#} (selected after conducting filter phase of PFWM) to SVM-GRBF (wrapper phase of PFWM), the prediction accuracy of pMDTFs with

length ({1#}, {2#} to {r#}) are obtained. Finally, according to (27), the pMDTFs with length x which is caused to high prediction accuracy are selected as MDTFs of $UToTMTS^i$.

In the final step of the PFWM strategy, the obtained MDTFs per $UToTMTS^i$ are gathered and AMDTFs is obtained as (28). The size of the 2-dimension AMDTFs matrix is $[AMDTFs]_{n\times PF}$ where n is the number of transient samples, and TPF (total PF) is the selected total MDTFs of $UToTMTS_{1:t}$ [selected cycles of F_1+ selected cycles of $F_2+\ldots+$ selected cycles of F_t].

$$[AMDTFs]_{n\times TPF}=\begin{pmatrix} MDTFs_{\{pMDTFs\}^{len:x\#}}^{UToTMTS_1} \\ \cup \\ MDTFs_{\{pMDTFs\}^{len:x\#}}^{UToTMTS_2} \\ \cup \\ \vdots \\ \cup \\ MDTFs_{\{pMDTFs\}^{len:x\#}}^{UToTMTS_t} \end{pmatrix} \tag{28}$$

III. EXPERIMENTAL DESIGN

A. TRANSIENT DATASET CONSTRCUTION

In this paper, transient data required to exert all scenarios embedded in the proposed framework (See Fig. 1) is generated through coupling Python code and SIEMENS power system simulator for engineering (PSS/E) planning tools. SIEMENS PSS/E is a powerful network analysis tool, and Python technology has powerful data processing modules. Hence, interlacing python into PSS/E leading to results in high reliability for dynamic simulation [23]. In this section, contingencies that are generated using the PSS/E application program interface (API) routine are substation outages, generator outages, line outages, and random combinations of generator and line outages. Also, time parameters of dynamic simulation, including fault duration time (0.23 second (s)) and fault clearing time (after the end of fault duration time) regarded at 0.0167 s time step. For each transient sample, 4 cycles (0.0668 s) per univariate time series are observed after fault clearing time for required analysis in the proposed framework. Furthermore, different load characteristics are considered in generating dynamic responses of the test systems. This parameter is used to convert the constant MVA load for a specified grouping of network loads to a specified mixture of the constant MVA, constant current, and constant admittance load characteristics. The test system used in this paper is the 68-Bus New England-New York Interconnection (NETS-NYPS) system (See Fig. 7) [24]. In all, 800 simulation transient samples were generated based on defined 28-variate trajectory features [25]–[27] (See Table 1) for existing processes in our study. As a concise report of transient sample recording workflow, according to Fig. 8, after implementing the Python code to save the channel activity behavior of the desired attributes (e.g., voltage, active power, and so on), the PSSPLT module is employed to put the data into a common file format (e.g., text file) (See level 2 of Fig. 8). When all contingency samples of the test systems are obtained, the data file is sent to the feature calculation module in MATLAB to extract the desired features (See level 4 of Fig. 8). For example, Fig. 9 shows the $Max[busvoltage]$ variations (F_6 univariate of 28-variate time series in Table 1) for stable and unstable samples of the NETS-NYPS grid case.

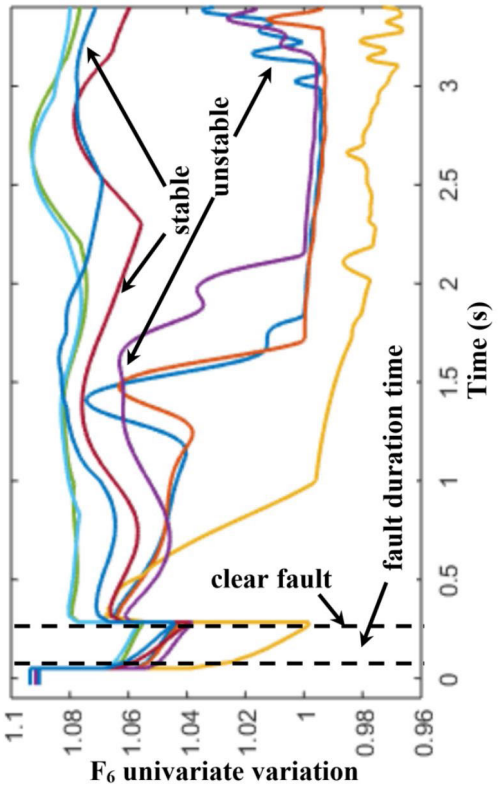


FIGURE 9. The transient samples stability status based on monitoring F_6 excursions: stable and unstable samples.

$$MDTFs_{\{pMDTFS\}^{len:x\#}}^{UToTMTS_i}=\{pMDTFs\}^{len:x\#}\left[Max_{Acc^{x\#}}(Z)\right]$$
$$Z=\begin{pmatrix} Acc^{1\#}=SVMClassifier^{Kernel:GRBF}(UToTMTS_i^{\{pMDTFs\}^{len:1\#}}) \\ Acc^{2\#}=SVMClassifier^{Kernel:GRBF}(UToTMTS_i^{\{pMDTFs\}^{len:2\#}}) \\ \vdots \\ Acc^{r\#}=SVMClassifier^{Kernel:GRBF}(UToTMTS_i^{\{pMDTFs\}^{len:r\#}}) \end{pmatrix} \tag{27}$$

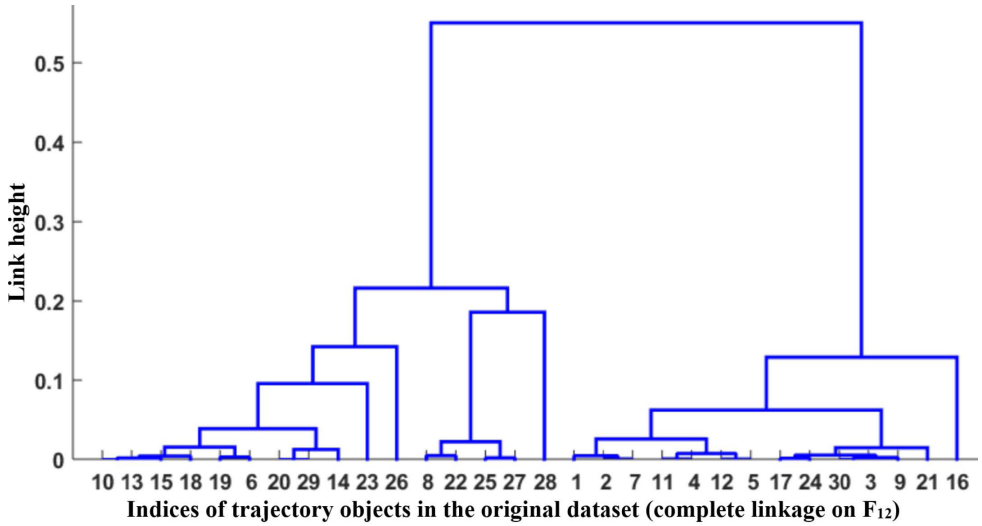


FIGURE 10. The hierarchical cluster tree of F_{12} based on the complete linkage function.

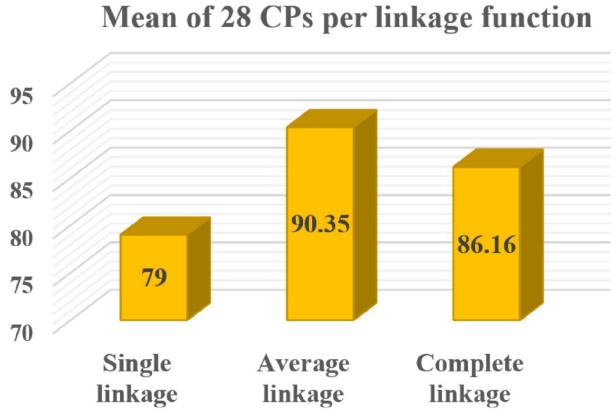


FIGURE 11. The CP rate per linkage function.

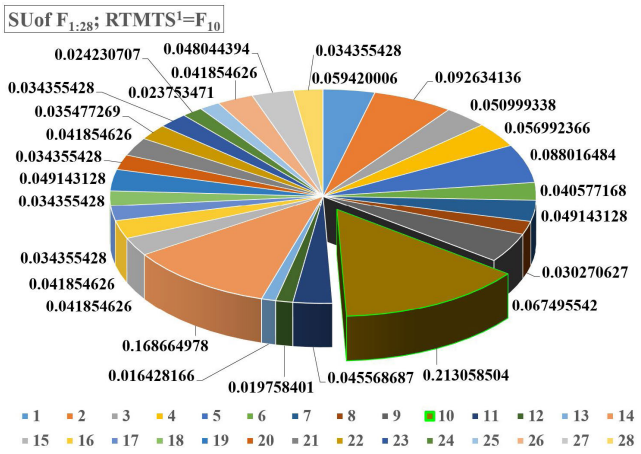


FIGURE 12. Weights of 28 trajectory features obtained by SU measure; $RTMTS^1 = F_{10}$.

B. OBTAINED RESULTS BASED ON BMHFSS

1) TFWM STRATEGY RESULTS

a: FILTER PHASE OF TFWM

According to what was mentioned in Step 1 of Section 2.1, the calculated proximity matrix via DTW (Eqs. (5), (6)) for TMTS is entered into linkage functions to generate a hierarchical cluster tree per UToTMTS (Eq. (7)). For example, the hierarchical cluster tree of F_{12} UToTMTS (See Table 1) based on the complete linkage method is shown in Fig. 10.

After clustering 28 UToTMTS based on triple AHC techniques (single, complete and average linkage), linkage function-specific 28 CPs are obtained based on Eq. (8). Fig. 11 shows the mean of 28 CPs (accuracy rate) in clustering solution per AHC technique which the average linkage outperforms the other two linkage functions in clustering the trajectory data. Hence, In the continuation of the TFWM strategy, we will use the results obtained from this approach. After conducting the preprocessing step based on the average linkage method, we first ranked the 28 UToTMTS based on the SU measure (See Eqs. (1) to (3)). Fig. 12 shows the SU amount of 28 trajectory features. Thereafter, we follow Steps 2 to 4 of TFWM's filter phase by using Eqs. (9) to (13). Based on Step 2, UToTMTS with the highest SU called $UToTMTS^{h(1)}$ is situated in the RTMTS subset as the $RTMTS^1$(See Eq. (9)). As can be seen in Fig. 12, F_{10} has a highest SU among other trajectory features. Hence, F_{10} is considered as the $RTMTS^1$. Next, by considering Eqs. (10) to (13), k-1 other RTMTS are selected according to the obtained results of Figure 13. By performing the filter phase of TFWM, 22 RTMTS of the 28 initial trajectory features including F_1: F_{11}, F_{14}, F_{15}, F_{16}, F_{18}: F_{22}, F_{24}, F_{26}, and F_{27} are selected. Figure 13 shows the weights of the features and selected feature (point explosion in 3-D pie) in each filter phase execution. In fact, each 3-D pie in Fig. 13 shows how k-1 RTMTS is selected by conducting the TFWM filter phase. For example, after selecting UToTMTS with the highest SU, called $RTMTS^1 = F_{10}$ based on Eq. (9), F_{10} was removed from the feature set with 28 members (a green-face legend

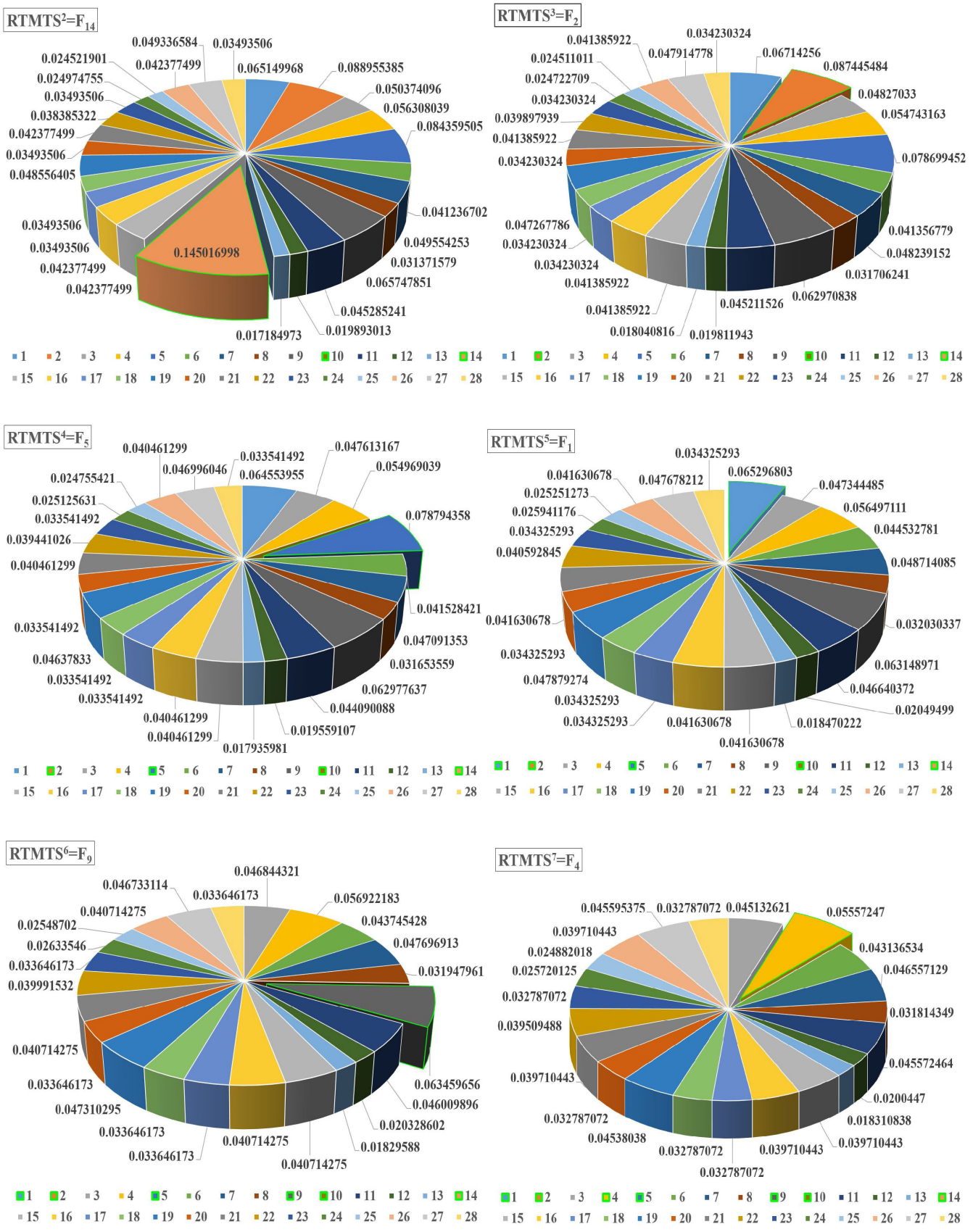


FIGURE 13. Selecting RTMTS subset based on filter phase of TFWM strategy.

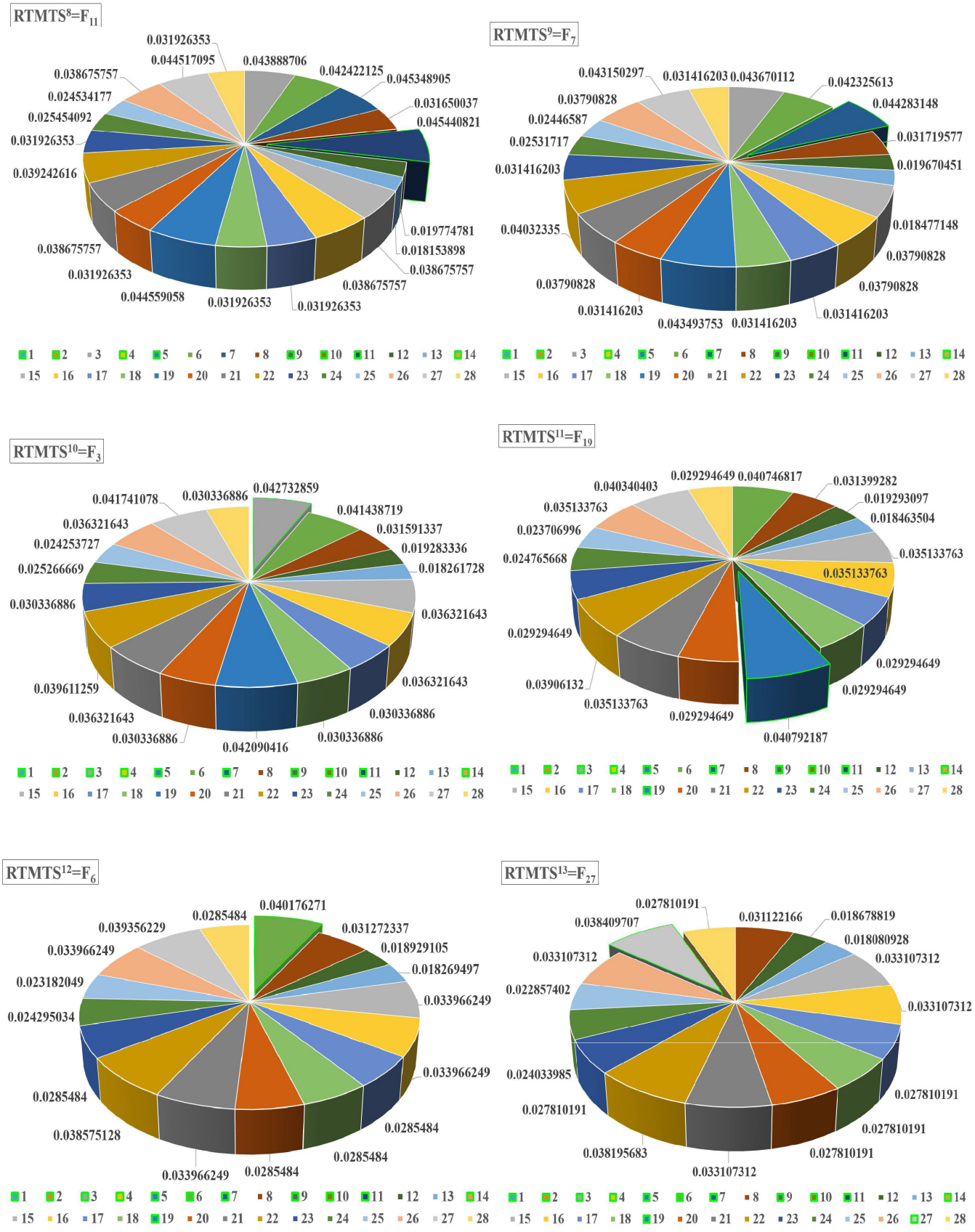


FIGURE 13. *(Continued.)* **Selecting RTMTS subset based on filter phase of TFWM strategy.**

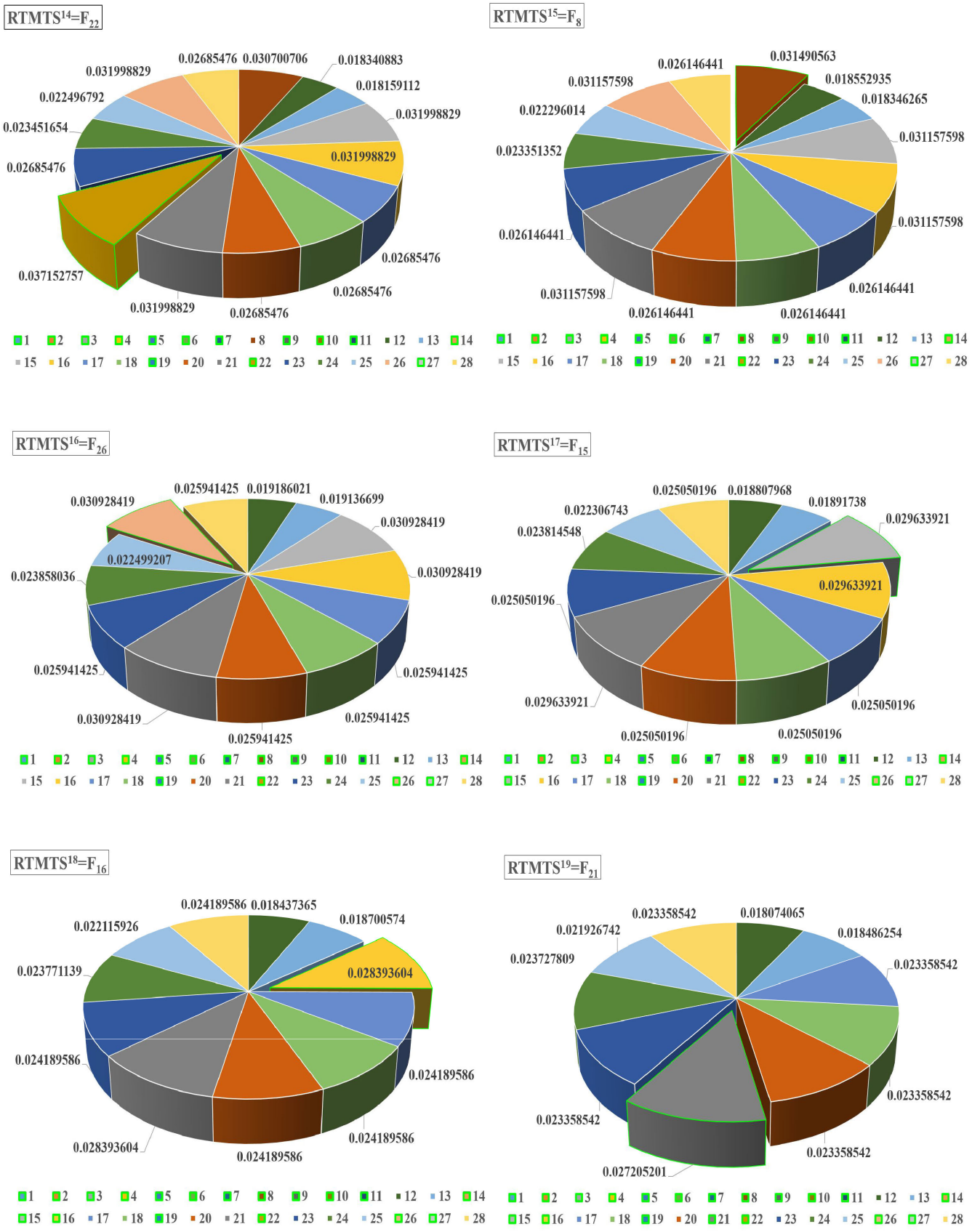


FIGURE 13. *(Continued.)* **Selecting RTMTS subset based on filter phase of TFWM strategy.**

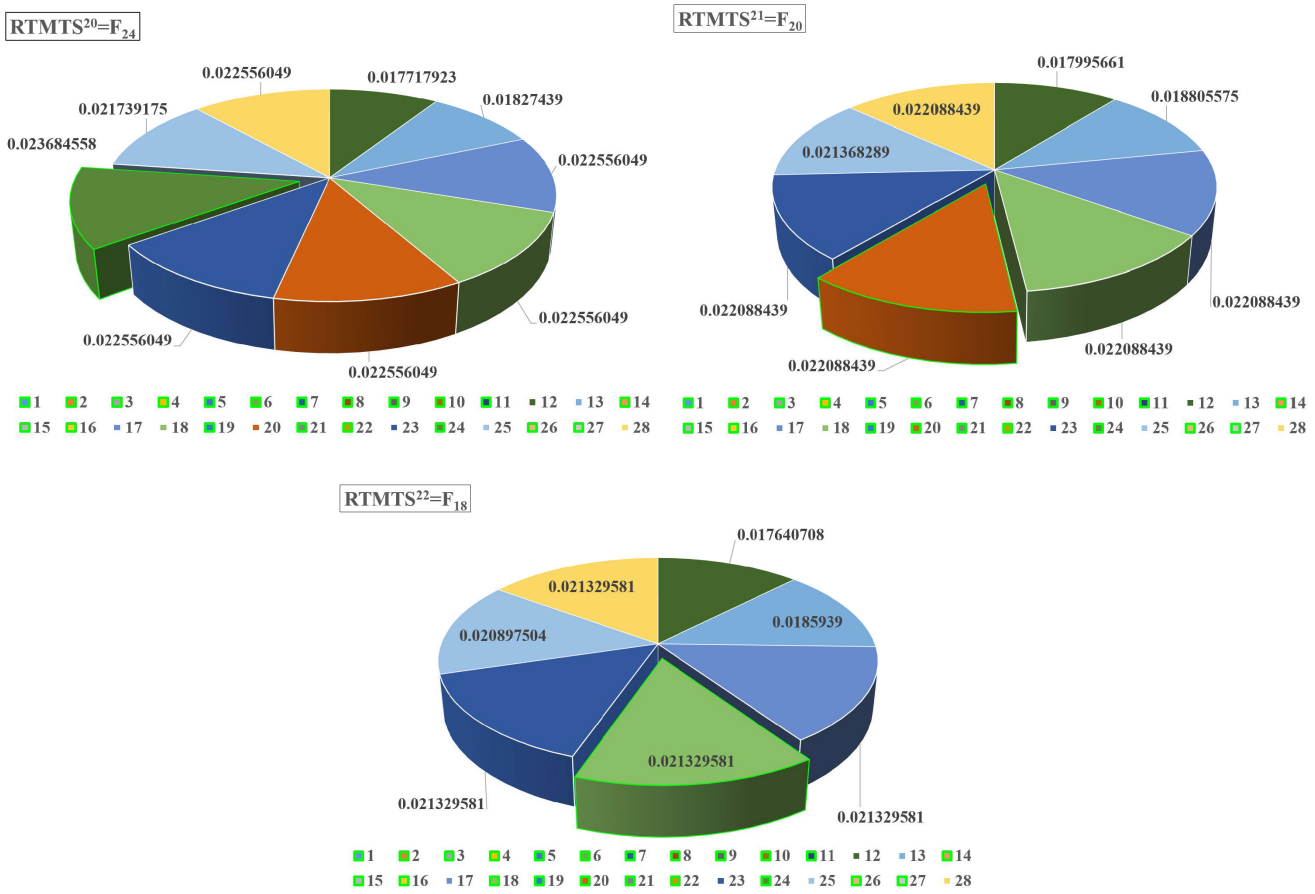


FIGURE 13. *(Continued.)* **Selecting RTMTS subset based on filter phase of TFWM strategy.**

in Fig. 12). Next, by considering the Eq. (10), the *IR* value (refer to Eq. (4)) of the remaining features (27 features minus F_{10}) in the presence of F_{10} was calculated and then multiplied to $W^{(0)}$ ($W^{(0)}$ represents the initial weight of the features which is equal to 1). In this way, the new weight of 27 features called $W^{(1)}$ is obtained. Then, $W^{(1)}$ is multiplied by the SU value of 27 features, and thereafter each feature has the new value called SU^{New}. Now, according to Eq. (11), the feature with the highest SU^{New} is selected as $RTMTS^2$ (namely F_{14}; point explosion in the top-left corner 3-D pie of the first 3-D package of Fig. 13). For selecting k-2 RTMTS, this process continues as Eqs. (12) and (13).

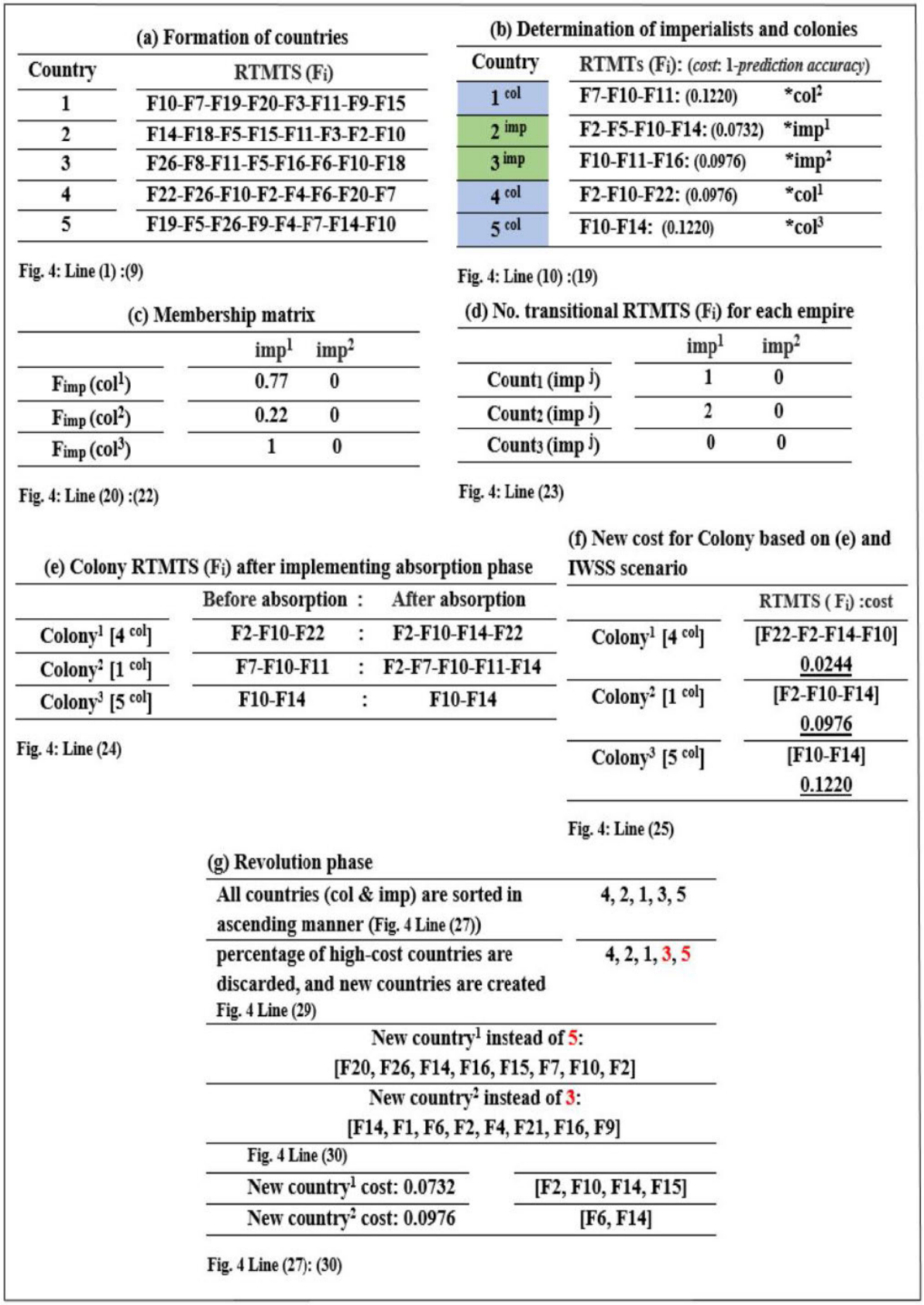
b: WRAPPER PHASE OF TFWM

After conducting the filter phase of TFWM, the RTMTS subset (22-variate transient time series features) is entered into the wrapper phase of TFWM mounted on a trihedral kernel-based approach including, FICA-IWSS scenario to find IMRTMTS (See second subsection of Section 2.1 including Fig. 4 and example of an iteration of the wrapper phase). In fact, the wrapper phase of TFWM based on the FICA-IWSS scenario is exerted three times in the presence of the triple kernel (Eq. (18) to Eq. (20): GRBF, DTW-GRBF, and RTWK) for learning

TABLE 2. **Selecting IMRTMTS based on FICA-IWSS accompanied with triple kernel.**

FICA-IWSS: kernel	$MRTMTS^{kernel}$
FICA-IWSS: GRBF	$MRTMTS^{GRBF}$: [F_2, F_{10}, F_{14}, F_{22}, F_{24}]
FICA-IWSS: DTW-GRBF	$MRTMTS^{DTW\text{-}GRBF}$: [F_2, F_{10}, F_{14}, F_{22}]
FICA-IWSS: RTWK	$MRTMTS^{RTWK}$: [F_2, F_{10}, F_{14}, F_{15}]
IMRTMTS	
[F_2, F_{10}, F_{14}]	

procedures in this phase. Consequently, the kernel-specific MRTMTS involved selected trajectory features in the form of $MRTMTS^{GRBF}$, $MRTMTS^{GRBF-DTW}$, and $MRTMTS^{RTWK}$ are obtained. Thereafter, the intersection of $MRTMTS^{GRBF}$, $MRTMTS^{GRBF-DTW}$, and $MRTMTS^{RTWK}$ called IMRTMTS is selected as the optimal trajectory features subset in the wrapper phase of TFWM. According to what was mentioned in the wrapper phase of TFWM, the results involved MRTMTS grouped by kernels and IMRTMTS are shown in Table 2. For more clarity on details of the obtained results in the wrapper phase, how selecting $MRTMTS^{DTW-GRBF}$ (F_2, F_{10}, F_{14}, F_{22}) via FICA-IWSS based on DTW-GRBF kernel conducted in two iterations are shown in Fig. 14 and Fig. 15.



(a) Formation of countries

Country	RTMTS (F_i)
1	F10-F7-F19-F20-F3-F11-F9-F15
2	F14-F18-F5-F15-F11-F3-F2-F10
3	F26-F8-F11-F5-F16-F6-F10-F18
4	F22-F26-F10-F2-F4-F6-F20-F7
5	F19-F5-F26-F9-F4-F7-F14-F10

Fig. 4: Line (1) :(9)

(b) Determination of imperialists and colonies

Country	RTMTs (F_i): (*cost*: 1-*prediction accuracy*)	
1 col	F7-F10-F11: (0.1220)	$*col^2$
2 imp	F2-F5-F10-F14: (0.0732)	$*imp^1$
3 imp	F10-F11-F16: (0.0976)	$*imp^2$
4 col	F2-F10-F22: (0.0976)	$*col^1$
5 col	F10-F14: (0.1220)	$*col^3$

Fig. 4: Line (10) :(19)

(c) Membership matrix

	imp^1	imp^2
$F_{imp}(col^1)$	0.77	0
$F_{imp}(col^2)$	0.22	0
$F_{imp}(col^3)$	1	0

Fig. 4: Line (20) :(22)

(d) No. transitional RTMTS (F_i) for each empire

	imp^1	imp^2
$Count_1$ (imp^j)	1	0
$Count_2$ (imp^j)	2	0
$Count_3$ (imp^j)	0	0

Fig. 4: Line (23)

(e) Colony RTMTS (F_i) after implementing absorption phase

	Before absorption :	After absorption
$Colony^1$ [4 col]	F2-F10-F22 :	F2-F10-F14-F22
$Colony^2$ [1 col]	F7-F10-F11 :	F2-F7-F10-F11-F14
$Colony^3$ [5 col]	F10-F14 :	F10-F14

Fig. 4: Line (24)

(f) New cost for Colony based on (e) and IWSS scenario

	RTMTS (F_i) :cost
$Colony^1$ [4 col]	[F22-F2-F14-F10] 0.0244
$Colony^2$ [1 col]	[F2-F10-F14] 0.0976
$Colony^3$ [5 col]	[F10-F14] 0.1220

Fig. 4: Line (25)

(g) Revolution phase

All countries (col & imp) are sorted in ascending manner (Fig. 4 Line (27))	4, 2, 1, 3, 5
percentage of high-cost countries are discarded, and new countries are created	4, 2, 1, 3, 5

Fig. 4 Line (29)

New $country^1$ instead of 5:
[F20, F26, F14, F16, F15, F7, F10, F2]

New $country^2$ instead of 3:
[F14, F1, F6, F2, F4, F21, F16, F9]

Fig. 4 Line (30)

New $country^1$ cost: 0.0732	[F2, F10, F14, F15]
New $country^2$ cost: 0.0976	[F6, F14]

Fig. 4 Line (27): (30)

FIGURE 14. The obtained results by applying FICA-IWSS based on DTW-GRBF kernel in the first iteration.

(b) Determination of imperialists and colonies

Country	RTMTs (F_i): (*cost*: 1-*prediction accuracy*)	
1 col	F2-F10-F14: (0.0976)	*col²
2 imp	F2-F5-F10-F14: (0.0732)	*imp²
3 col	F6-F14: (0.0976)	*col³
4 imp	F22-F2-F14-F10: (0.0244)	*imp¹
5 col	F2-F10-F14-F15: (0.0732)	*col¹

Fig. 4: Line (12) :(19)

(c) Membership matrix

	imp¹	imp²
F_{imp} (col¹)	0.5	0
F_{imp} (col²)	0	0
F_{imp} (col³)	0.5	0

Fig. 4: Line (20) :(22)

(d) No. transitional RTMTS (F_i) for each empire

	imp¹	imp²
$Count_1$ (imp ^{j})	1	0
$Count_2$ (imp ^{j})	0	0
$Count_3$ (imp ^{j})	1	0

Fig. 4: Line (23)

(e) Colony RTMTS (F_i) after implementing absorption phase

	Before absorption :	After absorption
Colony¹ [5 col]	F2-F10-F14-F15 :	F2-F10-F14-F15-F22
Colony² [1 col]	F2-F10-F14 :	F2-F10-F14
Colony³ [3 col]	F6-F14 :	F6-F14-F22

Fig. 4: Line (24)

(f) New cost for Colony based on (e) and IWSS scenario

	RTMTS (F_i): cost
Colony¹ [5 col]	[F2-F10-F14-F15] 0.0.732
Colony² [1 col]	[F2-F14-F10] 0.0976
Colony³ [3 col]	[F6-F14] 0.0976

Fig. 4: Line (25)

(g) Revolution phase

All countries (col & imp) are sorted in ascending manner (Fig. 4 Line (27))	4, 2, 5, 1, 3
percentage of high-cost countries are discarded, and new countries are created	4, 2, 5, 1, 3

Fig. 4 Line (29)

New country¹ instead of 3: [F4, F19, F8, F10, F3, F14, F5, F18]
New country² instead of 1: [F10, F27, F2, F11, F3, F26, F1, F19]

Fig. 4 Line (30)

New country¹ cost: 0.0976	[F3, F10, F14]
New country² cost: 0.0976	[F2, F10, F11]

Fig. 4 Line (27): (30)

MRTMTS$^{DTW\text{-}GRBF}$

4^{imp} (cost: 0.0244) = [F2-F10-F14-F22]

Fig 4. Line (33)

FIGURE 15. The obtained results by applying FICA-IWSS based on DTW-GRBF kernel in the second iteration.

To help the reader better understating, the obtained results per step of the wrapper phase are explained in a detailed manner. In the first iteration of FICA-IWSS based on DTW-GRBF kernel (See Fig. 14), five countries (part (a) of Fig. 14) contain 8 features (feature assigned per country based on probability score of 22-variate RTMTS; lines 1-9 of Fig. 4) are obtained. Then, based on the cost factor (Refer to lines 10-19 of Fig. 4), the cost per country was calculated, and the countries were divided into imperialists (low cost) and colonies (high cost) (part (b) of Fig. 14; e.g., two counties with index 2 and 3 is selected as imperialists and rest (index 1, 4, and 5) is considered as colonies). In part (c) of Fig. 14, the membership degree of each colony per imperialists was calculated according to membership function (lines 20-22 of Fig. 4). Thereafter, the number of features that should be transformed from all imperialists to colonies with index 1,

4, and 5 according to absorption phase (See part (d) and (e); e.g., one feature from imperialist with index 2 should be transformed to the colony with index 4; namely F_{14}). After conducting the absorption phase for all colonies, a new cost per colony based on (e) part and IWSS scenario is calculated. Next, according to the revolution phase (See part (g) of Fig. 14), all countries are sorted, and two high-cost countries are discarded (the second row of the part (g) of Fig. 14; countries with index 3 and 5 highlighted by red-face) and new countries are substituted with them (the third and fourth row in part (g) of Fig. 14). In the final step of the first iteration, the cost of new countries was calculated (the fifth row of the part (g) of Fig. 14). Also, these steps are continued for the second iteration of FICAIWSS based on DTW-GRBF which is results are shown in Fig. 15. Furthermore, for selecting $\text{MRTMTS}^{\text{GRBF}}$ and $\text{MRTMTS}^{\text{RTWK}}$, the defined scenario in the wrapper phase regarding the other kernels (GRBF and RTWK) for FICAIWSS was followed according to Fig. 4.

2) PFWM STRATEGY RESULTS

After selecting optimal transient trajectory features based on the TFWM strategy in the form of the IMRTMTS subset, namely F_2, F_{10}, and F_{14}, we conducted the PFWM strategy, including filter and wrapper phase per element of IMRTMTS subset separately for selecting MDTFs. Hence, in this section, we elaborated on the obtained results of exerting the PFWM strategy in the following paragraphs.

a: FILTER PHASE OF PFWM

According to what was mentioned in Section 2.2, first, we conducted the filter phase for selecting preliminary MDTFs-specific per UToTMTS of IMRTMTS subset in the form of the point-feature package $\{UToTMTS_i^{PF^{h(1)}}, \ldots, UToTMTS_i^{PF^{h(r)}}\}$ according to Eq. (22) to Eq. (26) (See Section 2.2; Step 1 to Step 3). We set $r = 3$ as length of point-feature package ({1#}, {2#}, and {3#}) in the preliminary

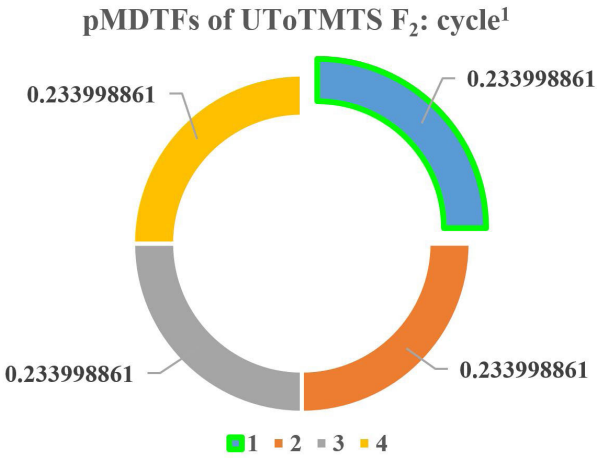


FIGURE 16. Selecting pMDTFs of F_2 (Cycle^1).

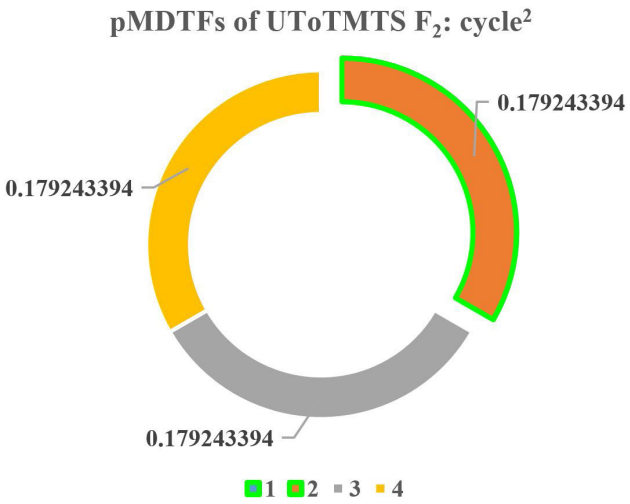


FIGURE 17. Selecting pMDTFs of F_2 (Cycle^2).

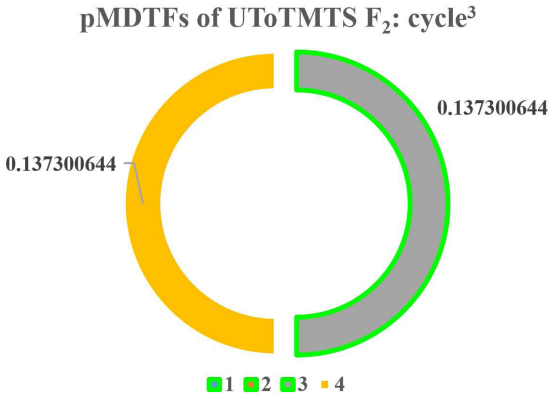


FIGURE 18. Selecting pMDTFs of F_2 (Cycle^3).

MDTFs subset per $UToTMTS^i$. Given that each UToTMTS of IMRTMTS has 4 PF (4 cycles observed after fault clearing time), for example, after conducting filter phase on UToTMTS F_2, we selected preliminary MDTFs (*pMDTFs*) for F_2 as follow (Fig. 16 to Fig. 18):

$$UToTMTS_{F_2}^{PF^{h(1)}} = cycle^1$$
$$UToTMTS_{F_2}^{\{pMDTFs\}^{len:1\#}} \leftarrow UToTMTS_{F_2}^{PF^{h(1)}} \quad (29)$$

$$UToTMTS_{F_2}^{PF^{h(2)}} = cycle^2$$
$$UToTMTS_{F_2}^{\{pMDTFs\}^{len:2\#}} \leftarrow \{UToTMTS_{F_2}^{PF^{h(1)}}, UToTMTS_{F_2}^{PF^{h(2)}}\} \quad (30)$$

$$UToTMTS_{F_2}^{PF^{h(3)}} = cycle^3$$
$$UToTMTS_{F_2}^{\{pMDTFs\}^{len:3\#}} \leftarrow \{UToTMTS_{F_2}^{PF^{h(1)}}, UToTMTS_{F_2}^{PF^{h(2)}}, UToTMTS_{F_2}^{PF^{h(3)}}\} \quad (31)$$

Also, *pMDTFs* for UToTMTS F_{10} and UToTMTS F_{14} were obtained according to Table 3.

TABLE 3. Selecting point-feature package of F_{10} and F_{14}.

F_{10}: pMDTFs	F_{14}: pMDTFs
len: 1# {cycle^1}	len: 1# {cycle^1}
len: 2# {cycle^1, cycle^2}	len: 2# {cycle^1, cycle^2}
len: 3# {cycle^1, cycle^2, cycle^3}	len: 3# {cycle^1, cycle^2, cycle^3}

- Wrapper phase of PFWM

TABLE 4. Selecting high performance point-feature package of F_2, F_{10} and F_{14}.

F_2: MDTFs	F_{10}: MDTFs
len: 1# {cycle^1}; Acc. (79.45 %)	len: 1# {cycle^1}; Acc. (76.59 %)
len: 2# {cycle^1, cycle^2}; Acc. (82.52 %)	len: 2# {cycle^1, cycle^2}; Acc. (83.55 %)
len: 3# {cycle^1, cycle^2, cycle^3}; Acc. (84.50 %)	**len: 3# {cycle^1, cycle^2, cycle^3}; Acc. (84.46 %)**
F_2: MDTFs	
len: 1# {cycle^1}; Acc. (87.50 %)	
len: 2# {cycle^1, cycle^2}; Acc. (88.46 %)	
len: 3# {cycle^1, cycle^2, cycle^3}; Acc. (89.48 %)	

TABLE 5. The performance metrics.

Metrics	Descriptions
Accuracy	Acc=(TP+TN)/(TP+TN+FP+FN)
Sensitivity	TPR=TP/(TP+FN)
Specificity	TNR=TN/(TN+FP)

Symbols; P: stable sample, N: unstable sample,
T: predicted correctly, F: predicted incorrectly

b: WRAPPER PHASE OF PFWM

After finding optimal cycles of F_2, F_{10}, and F_{14} in the form of triple point-feature package, each point-feature package ({1#}, {2#}, and {3#}) belongs to $UToTMTS^i$ are fed to GRBF-SVM for selecting high-performance point-feature package as the MDTFs for $UToTMTS^i$ according to Eq. (27) (See Section 2.2; Step 4). As can be seen in Table 4, $UToTMTS^{\{pMDTFs\}^{len:3\#}}$ for F_2, F_{10}, and F_{14} has a high performance based on the learning model applied in the wrapper phase of the PFWM strategy.

Finally, the obtained MDTFs per $UToTMTS^i$ are gathered and AMDTFS (total point-feature: #9) is obtained according to Eq. (28):

$$AMDTFs^{TPF:\#9} = \begin{pmatrix} MDTFs^{UToTMTS_{F_2}}_{\{cycle^1,cycle^2,cycle^3\}^{len:3\#}} \\ \cup \\ MDTFs^{UToTMTS_{F_{10}}}_{\{cycle^1,cycle^2,cycle^3\}^{len:3\#}} \\ \cup \\ MDTFs^{UToTMTS_{F_{14}}}_{\{cycle^1,cycle^2,cycle^3\}^{len:3\#}} \end{pmatrix} \quad (32)$$

C. TSP BASED ON AMDTFS

After selecting AMDTFs based on BMHFSS, assessing the efficacy of AMDTFs for TSP was considered in this section. To this end, the 10-fold cross-validation technique based on the GRBF-SVM classifier was applied on transient samples of the NETS-NYPS grid case. Also, triple metrics (See Table 5) per folds (10-fold training and testing sets) were considered via fine-tuning on C and σ to find optimal pairs on GRBF-SVM. These parameters in GRBF-SVM are selected from $\{C = 2^i | i = 0, 1, \ldots, 15\}$ and $\{\sigma = 2^j | j = -5, -4, \ldots, 15\}$. Based on the learning scenario, the maximum value of the *Acc* index per fold is shown in Table 6 (Fig. 19 shows the GRBF-SVM accuracy variations based on fine-tuning (C, σ) in some folds.). Also, according to obtained *Acc* per folds, *TPR* and *TNR* are calculated for more analysis on the performance capacity of AMDTFs on TSP. Finally, the average of triple metrics (*Acc*, *TPR*, *TNR*) in folds are calculated (underline-face in Table 6). As can be seen in Table 6, the classification accuracy of GRBF-SVM based on AMDTFs shows the high-performance capacity for TSP (*Acc*: 98.25 %, *TPR*: 97.75 %, and *TNR*: 98.75 %).

Besides the importance of prediction accuracy on TSA, the low processing time for timely corrective control action (<1 s) [3] is the most significant issue. By considering on transient simulation parameters setting discussed in Section 3.1, 4 cycles after fault clearing time are selected per trajectory features for conducting the proposed FSS scheme. By applying BMHFSS, selected AMDTFs via IMRTMTS does not exceed 3 cycles. Hence, only 3 cycles after fault clearing time are regarded as observation window (50.1 milliseconds (ms)). In addition to the observation window time, the prediction time based on the learning scenario should be

TABLE 6. Results of TSP based on AMDTFs.

Classifier	Test case	10-fold cross validation			
		Max(Acc.) per fold based on fine-tuning on C and σ Accuracy [TPR / TNR]			
		fold 1	**fold 2**	**fold 3**	**fold 4**
		100 [100 / 100]	97.5 [100 / 95]	96.25 [92.5 / 100]	95 [92.5 / 97.5]
		fold 5	**fold 6**	**fold 7**	**fold 8**
GRBF- SVM	NETS-NYPS	100 [100 / 100]	100 [100 / 100]	97.5 [97.5 / 97.5]	98.75 [97.5 / 100]
		fold 9		**fold 10**	
		97.5 [97.5 / 97.5]		100 [100 / 100]	
		Mean(measure) of folds: Accuracy [TPR / TNR]			
		98.25 [97.75 / 98.75]			

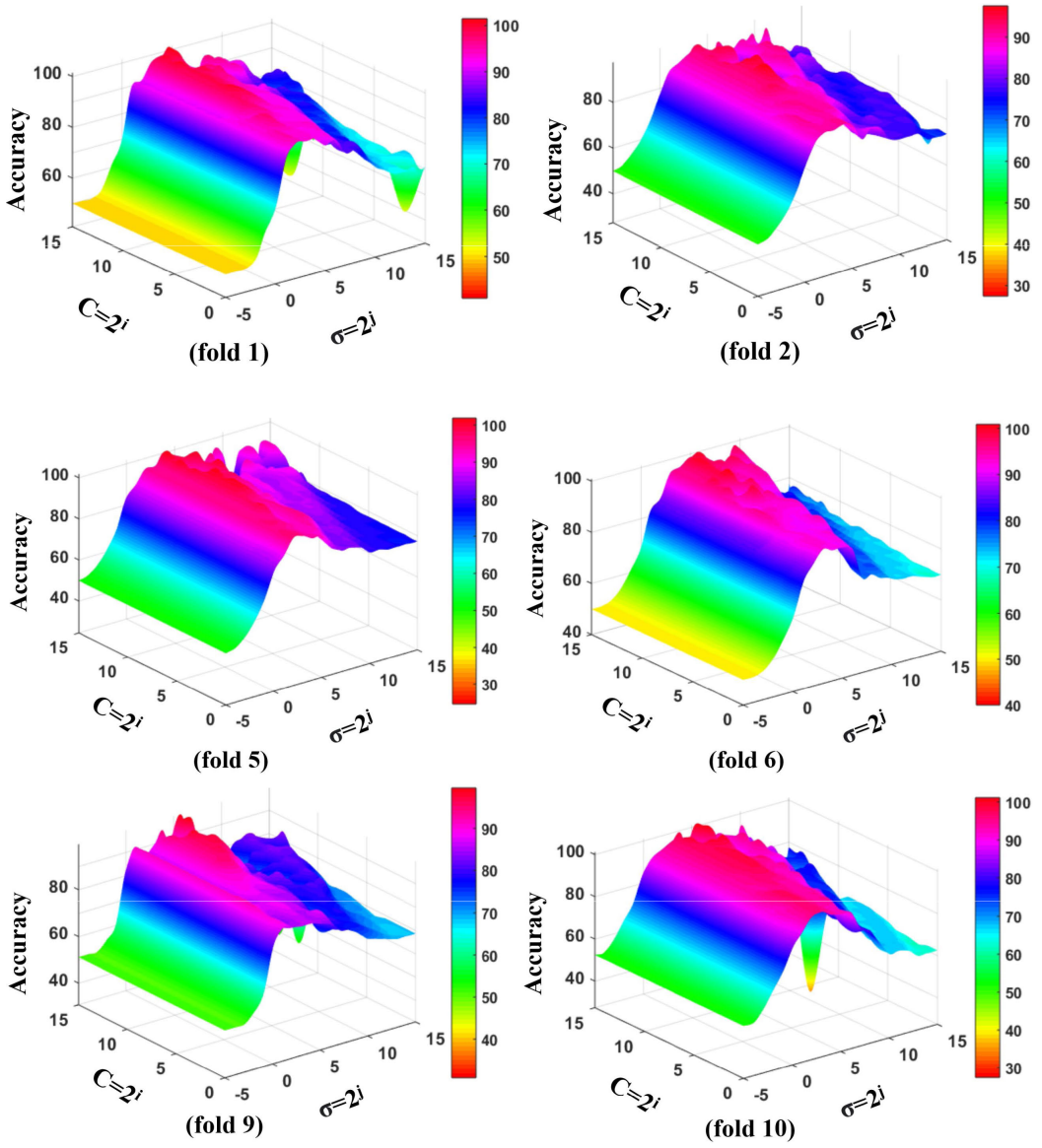


FIGURE 19. Illustration of GRBF-SVM performance variations (Acc metric) for TSP based on learning parameters in some folds.

TABLE 7. Processing time for TSP based on AMDTFs.

Observation window (OW) in cycle / second	Processing time [a]
3 / 0.0501	50.1 ms + 2.848 ms= 52.948 ms

[a] Processing time=OW+prediction time.

calculated. Based on GRBF-SVM and AMDTFs, the prediction time is 2.848 ms. Consequently, the processing time is 52.948 ms which provides a very acceptable time to take corrective action (See Table 7).

D. COMPARISON OF EXPERIMENTAL METHODS: BMHFSS vs. OTHER RELIABLE FSS SCHEMES

Comparing the performance of BMHFSS with other reliable FSS schemes for TSA is considered in this section. Hence, we focused on the three reliable FSS (3RFSS) strategies, including minimum redundancy maximum relevance (mRMR), ReliefF, and fast correlation-based filter (FCBF) methods which are considered by scholars for finding optimal transient features [4], [5], [7]. First, 3RFSS were

TABLE 8. The obtained optimal cycles based on other FSS schemes.

FSS schemes	optimal trajectories	selected optimal cycles of trajectory
mRMR	F_{20}	{cycle^3}
	F_{23}	{cycle^1, cycle^2, cycle^4}
	F_{18}	{cycle^2, cycle^3}
	F_{21}	{cycle^1, cycle^3, cycle^4}
FCBF	F_1	{cycle^1, cycle^2 cycle^3, cycle^4}
	F_2	{cycle^1, cycle^2, cycle^3, cycle^4}
	F_3	{cycle^3}
ReliefF	F_7	{cycle^2, cycle^3 }
	F_{11}	{cycle^1, cycle^2, cycle^3, cycle^4}
	F_{13}	{cycle^2, cycle^3, cycle^4}

TABLE 9. Results of TSP based on selected optimal cycles by 3RFSS schemes.

(FSS Scheme) Classifier	Test case	10-fold cross validation Max(Acc.) per fold based on fine-tuning on C and σ Accuracy [TPR / TNR]			
		fold 1	**fold 2**	**fold 3**	**fold 4**
		93.75 [97.5 / 90]	93.75 [97.5 / 90]	90 [95 / 85]	90 [82.5 / 97.5]
		fold 5	**fold 6**	**fold 7**	**fold 8**
(mRMR) GRBF- SVM	NETS-NYPS	95 [97.5 / 92.5]	95 [95 / 95]	95 [100 / 90]	91.25 [87.5 / 95]
		fold 9		**fold 10**	
		88.75 [95 / 82.5]		93.75 [92.5 / 95]	
		Mean(measure) of folds: Accuracy [TPR / TNR]			
		92.62 [94 / 91.25]			
		fold 1	**fold 2**	**fold 3**	**fold 4**
		98.75 [97.5 / 100]	95 [95 / 95]	96.25 [92.5 / 100]	97.5 [95 / 100]
		fold 5	**fold 6**	**fold 7**	**fold 8**
(FCBF) GRBF- SVM	NETS-NYPS	96.25 [92.5 / 100]	97.5 [97.5 / 97.5]	97.5 [95 / 100]	97.5 [97.5 / 97.5]
		fold 9		**fold 10**	
		97.5 [95 / 100]		100 [100 / 100]	
		Mean(measure) of folds: Accuracy [TPR / TNR]			
		97.37 [95.75 / 99]			
		fold 1	**fold 2**	**fold 3**	**fold 4**
		98.75 [97.5 / 100]	95 [95 / 95]	96.25 [92.5 / 100]	97.5 [95 / 100]
		fold 5	**fold 6**	**fold 7**	**fold 8**
(ReliefF) GRBF- SVM	NETS-NYPS	96.25 [92.5 / 100]	97.5 [97.5 / 97.5]	97.5 [95 / 100]	97.5 [97.5 / 97.5]
		fold 9		**fold 10**	
		97.5 [95 / 100]		100 [100 / 100]	
		Mean(measure) of folds: Accuracy [TPR / TNR]			
		97.37 [95.75 / 99]			

exerted on 28-variate trajectory features (See Table 1) for selecting optimal PFs. Next, the obtained optimal cycles by 3RFSS fed to GRBF-SVM classifier for performance evaluation on TSP. After applying 3RFSS on 28-variate time series features, the obtained optimal cycles grouped by 3RFSS were obtained according to Table 8 (the optimal cycles of the proposed method are given in Table 4). According to obtained results (See Table 9), regarding the same learning scenario, our strategy outperformed 3RFSS approaches based on the triple metric (ignoring only 0.25% less than TNR than ReliefF and FCBF). The obtained results of Table 4 and Table 8 show that BMHFSS in the presence of 9-cycles of 3-variate trajectory has better performance (Acc, TPR, and TNR) than mRMR, which uses 9-cycle of 4-variate trajectory features. In the case of the FCBF and ReliefF methods, which have an equal number of extracted trajectory features with our method, the selected cycles of 3-variate trajectory feature via BMHFSS leading to high *Acc* and *TPR* than FCBF and ReliefF schemes. Also, based on a three-cycle observation window, BMHFSS has low processing time (52.948 ms) than 3RFSS (mRMR: 68.793 ms, FCBF: 68.930 ms, and ReliefF: 68.910 ms), which uses a four-cycle observation window for TSP (See Table 8). For more details on the processing time of BMHFSS and 3RFSS, refer to Table 7 and Table 9.

TABLE 10. Processing time for TSP-based 3RFSS schemes.

FSS schemes	Observation window (OW) in cycle / second	Processing time [a]
mRMR	4 / 0.0668	66.8 ms + 1.993 ms= 68.793 ms
FCBF	4 / 0.0668	66.8 ms + 2.130 ms= 68.930 ms
ReliefF	4 / 0.0668	66.8 ms + 2.110 ms= 68.910 ms

[a] Processing time=OW+prediction time.

IV. CONCLUSION

In this paper, we proposed a feature selection scheme to compact the curse of dimensionality institutionalized over power system transient data. Given the necessity of focusing on transient data characteristics from two aspects (point and trajectory-based transient behaviors), the bi-mode hybrid feature selection scheme (BMHFSS) is considered for addressing this concern in the form of point-specific mode and the trajectory-specific mode. After the data gathering

process, the trajectory-oriented filter-wrapper mode (TFWM) is applied on transient multivariate time series (TMTS) in two phases. In the filter phase, information theory concepts based on triple criteria are considered for weighting the transient trajectories. By selecting high-weight univariate trajectories of TMTS called relevant TMTS (RTMTS), RTMTS entered to wrapper phase, including both fuzzy imperialist competitive algorithm (FICA) and incremental wrapper subset selection (IWSS) in the form of the trihedral kernel-based approach to find the intersected most RTMTS (IMRTMTS). Next, the filter-wrapper scenario in point-based mode (PFWM) is conducted for selecting the most discriminative transient features (MDTFs) per time series in IMRTMTS. Finally, aggregated MDTFs (AMDTFs) are tested to verify their efficacy for TSP based on the cross-validation technique. The obtained results show that the selected AMDTFs based on the proposed framework have a high-performance capacity (Acc greater than 98 %, TPR 97.75 %, TNR 98.75 %), and a processing time of 52.948 ms) for TSP. Also, for more clarity, we justify the effectiveness of the proposed method by comparing it with the other reliable FSS schemes. According to the obtained results based on triple metrics and applying the same learning conditions for all FSS schemes, using AMDTFs survived by BMHFSS compared to selected optimal features by mRMR, ReliefF, and FCBF algorithms lead to high performance on TSP.

REFERENCES

[1] S. Das, S. P. Singh, and B. K. Panigrahi, "Transmission line fault detection and location using wide area measurements," *Electr. power Syst. Res.*, vol. 151, pp. 96–105, Oct. 2017.
[2] P. Kundur *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definition," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387–1401, Aug. 2004.
[3] M. Pavella, M. Ernest, and D. Ruiz-Vega, *Transient Stability of Power Systems: A Unified Approach to Assessment and Control*, 1st ed. New York, NY, USA: Springer, 2000.
[4] A. Stief, J. R. Ottewill, and J. Baranowski, "Relief F-based feature ranking and feature selection for monitoring induction motors," in *Proc. 23rd Int. Conf. Methods Models Autom. Robot. (MMAR)*, Miedzyzdroje, Poland, Aug. 2018, pp. 171–176.
[5] X. Li, Z. Zheng, L. Wu, R. Li, J. Huang, X. Hu, and P. Guo, "A stratified method for large-scale power system transient stability assessment based on maximum relevance minimum redundancy arithmetic," *IEEE Access*, vol. 7, pp. 61414–61432, May 2019.
[6] J. Liu, H. Sun, Y. Li, W. Fang, and S. Niu, "An improved power system transient stability prediction model based on mRMR feature selection and WTA ensemble learning," *Appl. Sci.*, vol. 10, no. 7, p. 2255, Mar. 2020.
[7] J. Yan, C. Li, and Y. Liu, "Deep learning based total transfer capability calculation model," in *Proc. Int. Conf. Power Syst. Technol. (POWERCON)*, Guangzhou, China, Nov. 2018, pp. 952–957.
[8] S. Liu, L. Liu, Y. Fan, L. Zhang, Y. Huang, T. Zhang, J. Cheng, L. Wang, M. Zhang, R. Shi, and D. Mao, "An integrated scheme for online dynamic security assessment based on partial mutual information and iterated random forest," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3606–3619, Jul. 2020.
[9] L. Ji, J. Wu, Y. Zhou, and L. Hao, "Using trajectory clusters to define the most relevant features for transient stability prediction based on machine learning method," *Energies*, vol. 9, no. 11, p. 898, Nov. 2016.
[10] Z. Chen, X. Han, C. Fan, T. Zheng, and S. Mei, "A two-stage feature selection method for power system transient stability status prediction," *Energies*, vol. 12, no. 4, p. 689, Feb. 2019.
[11] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient agglomerative hierarchical clustering," *Expert Syst. with Appl.*, vol. 42, no. 5, pp. 2785–2797, Apr. 2015.
[12] M. Muller, *Information Retrieval for Music and Motion*, 1st ed. Heidelberg, Germany: Springer, 2007.
[13] Y. Piao and K. H. Ryu, "A hybrid feature selection method based on symmetrical uncertainty and support vector machine for high-dimensional data classification," in *Proc. Asian Conf. Intell. Inf. Database Syst.*, Kanazawa, Japan, 2017, pp. 721–727.
[14] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. Hoboken, NJ, USA: Wiley, 2011.
[15] G. J. Szekely and M. L. Rizzo, "Hierarchical clustering via joint between-within distances: Extending ward's minimum variance method," *J. Classification*, vol. 22, pp. 151–183, Sep. 2005.
[16] R. R. Sokal and F. J. Rohlf, "The comparison of dendrograms by objective methods," *Taxon*, vol. 11, no. 2, pp. 33–40, Feb. 1962.
[17] R. Ruiz, J. C. Riquelme, and J. S. Aguilar-Ruiz, "Incremental wrapper-based gene selection from microarray data for cancer classification," *Pattern Recognit.*, vol. 39, no. 12, pp. 2383–2392, Dec. 2006.
[18] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *Proc. IEEE Congr. Evol. Comput.*, Singapore, Sep. 2007, pp. 4661–4667.
[19] S. Arish, A. Amiri, and K. Noori, "FICA: Fuzzy imperialist competitive algorithm," *J. Zhejiang Univ. Sci. C*, vol. 15, no. 5, pp. 363–371, May 2014.
[20] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, Sep. 1995.
[21] H. Shimodaira, K.-I. Noma, M. Nakai, and S. Sagayama, "Support vector machine with dynamic time-alignment kernel for speech recognition," in *Proc. Eur. Conf. Speech Commun. Technol. (EUROSPEECH)*, Aalborg, Denmark, 2001, pp. 1841–1844.
[22] P.-F. Marteau and S. Gibet, "On recursive edit distance kernels with application to time series classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1121–1133, Jun. 2015.
[23] M. U. Khalid, F. Rashid, W. Akhtar, M. H. Ghauri, and O. Lateef, *Python Based Power System Automation in PSS/E*, Abu Dhabi, United Arab Emirates: Lulu, 2014.
[24] C. Canizares, T. Fernandes, E. Geraldi, Jr., L. Gérin-Lajoie, M. Gibbard, I. Hiskens, J. Kersulis, R. Kuiava, L. Lima, F. de Marco, N. Martins, B. Pal, A. Piardi, R. Ramos, J. dos Santos, D. Silva, A. Singh, B. Tamimi, and D. Vowles, "Benchmark systems for small-signal stability analysis and control," IEEE Power Energy Soc., Piscataway, NJ, USA, Tech. Rep. PES-TR18, Aug. 2015.
[25] A. B. Mosavi, A. Amiri, and H. Hosseini, "A learning framework for size and type independent transient stability prediction of power system using twin convolutional support vector machine," *IEEE Access*, vol. 6, pp. 69937–69947, 2018.
[26] S. A. B. Mosavi and S. Y. Banihashem, "Defining geometric cross-relevance multivariate trajectory features for transient stability analysis based on elastic kernel support vector machine," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 5, pp. 369–385, Oct. 2021.
[27] S. A. B. Mosavi, "Extracting discriminative features in reactive power variations by 1-persistence parallel fragmented hybrid feature selection scheme for transient stability prediction," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 4, pp. 500–513, Aug. 2021.

SEYED ALIREZA BASHIRI MOSAVI received the B.Sc. degree in computer engineering-software from Islamic Azad University of Zanjan, Zanjan, Iran, in 2011, the M.Sc. degree in information technology engineering-electronic commerce from the University of Qom, Qom, Iran, in 2013, and the Ph.D. degree in computer engineering-artificial intelligence and robotics from the University of Zanjan, Zanjan, in 2019. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Imam Khomeini International University-Buin Zahra Higher Education Center of Engineering and Technology, Qazvin, Iran. His master's thesis on customer value analysis won the Tejarat Bank Award as the best thesis related to quality. His doctoral thesis on designing a novel learning framework for size and type independent transient stability prediction of the power systems. He has published several papers and conferences in the field of customer relationship management (CRM), transient processes in power systems, and machine learning scope. His main research interests include CRM, data mining, pattern recognition, power system transient analysis, and power system dynamic and stability.