PRIOR ART

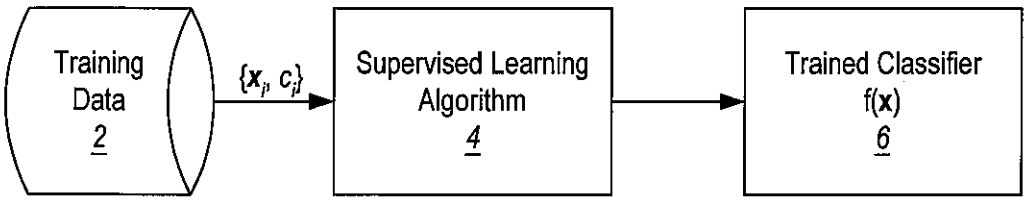
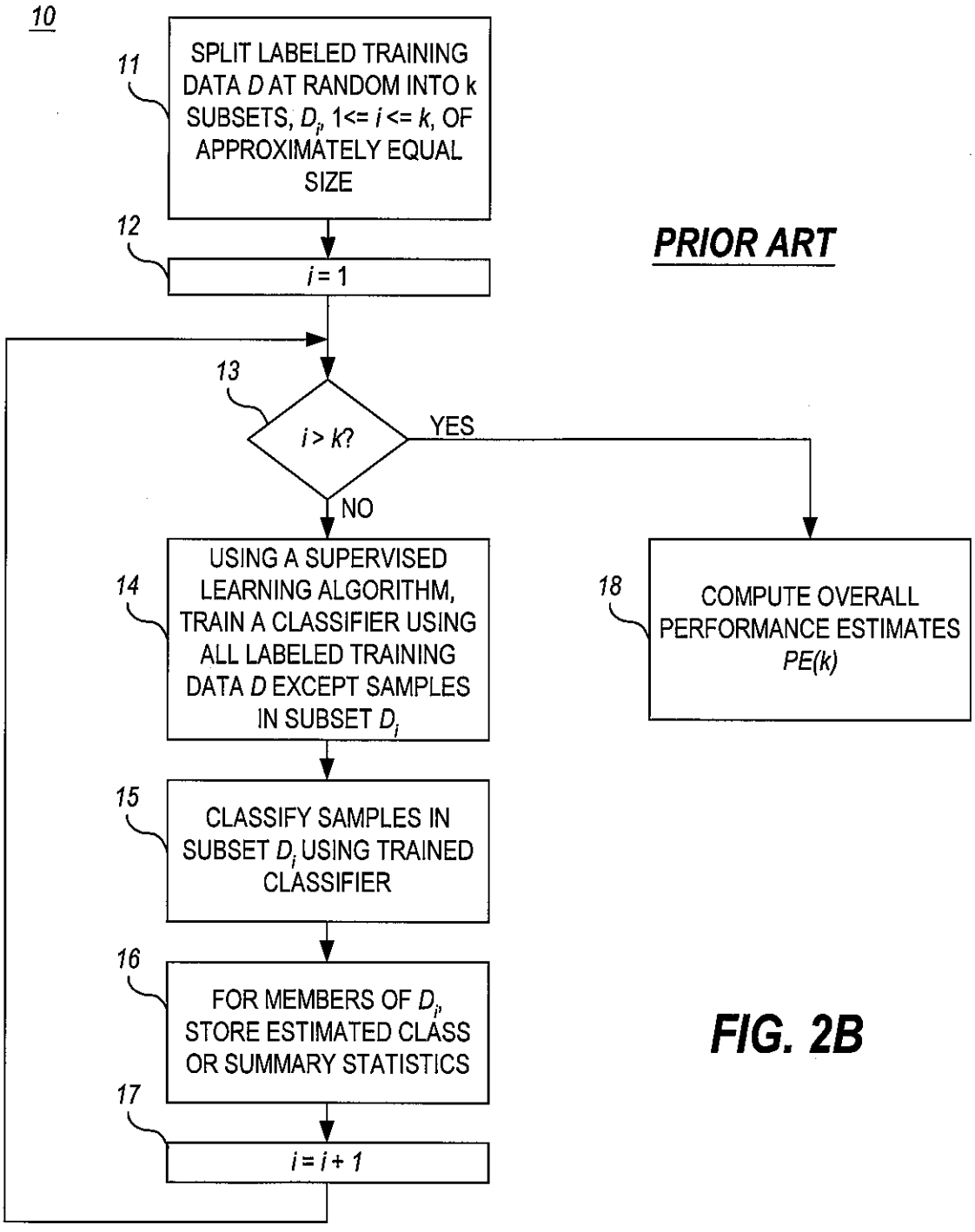


FIG. 1



PRIOR ART

FIG. 2B

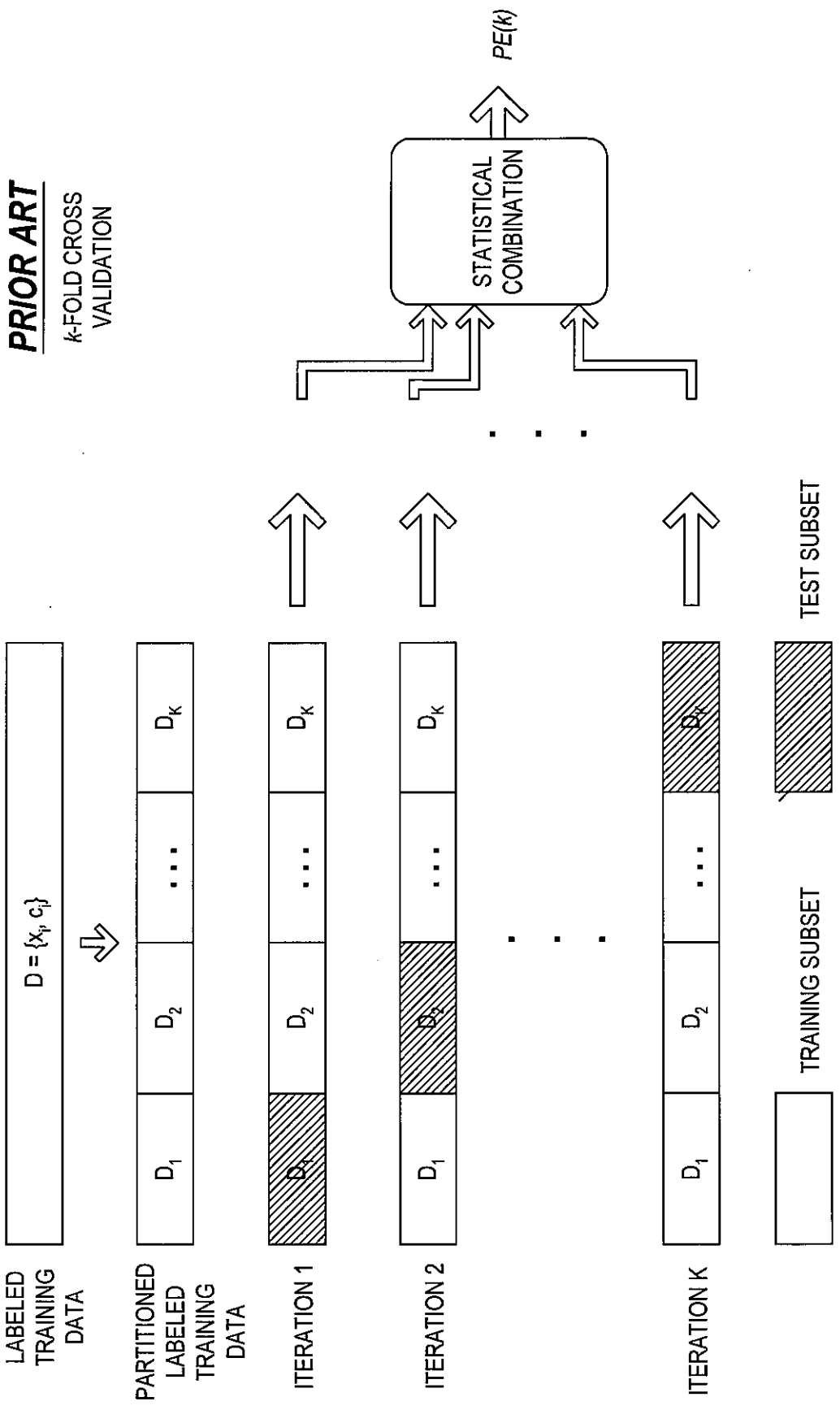

PRIOR ART
k-FOLD CROSS VALIDATION
LABELED TRAINING DATA
D = {x_i, c_i}
PARTITIONED LABELED TRAINING DATA
D1
D2
...
Dk
ITERATION 1
ITERATION 2
ITERATION K
STATISTICAL COMBINATION
PE(k)
TRAINING SUBSET
TEST SUBSET

FIG. 2A

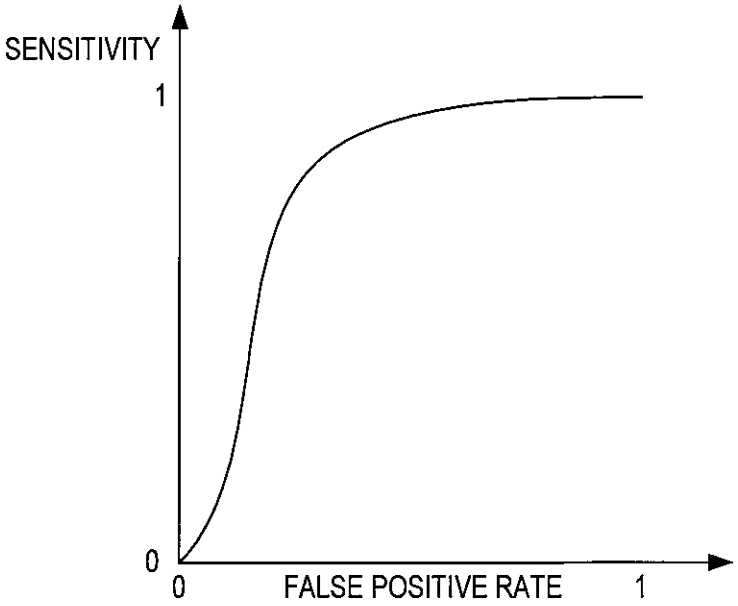


FIG. 3

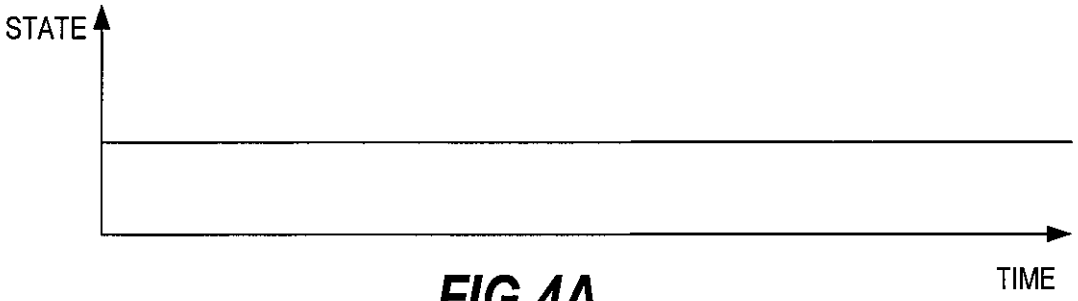


FIG.4A

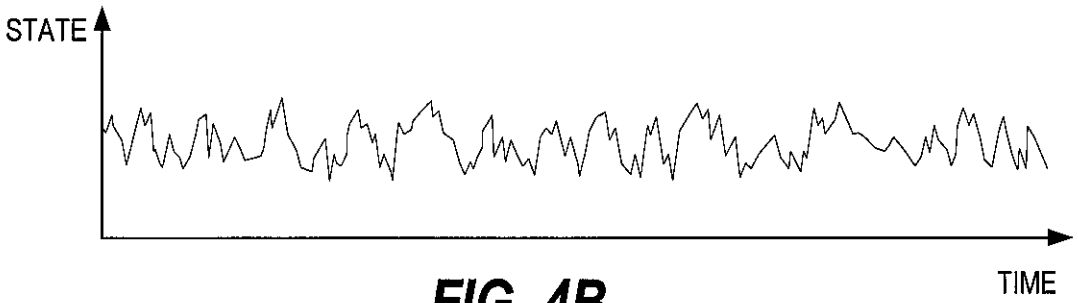


FIG. 4B

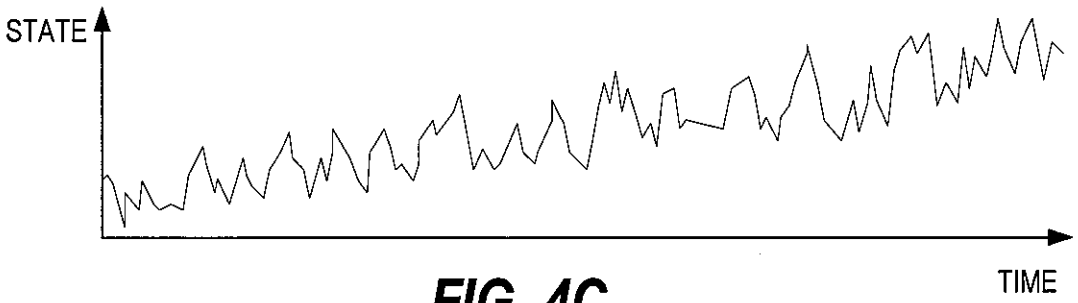


FIG. 4C

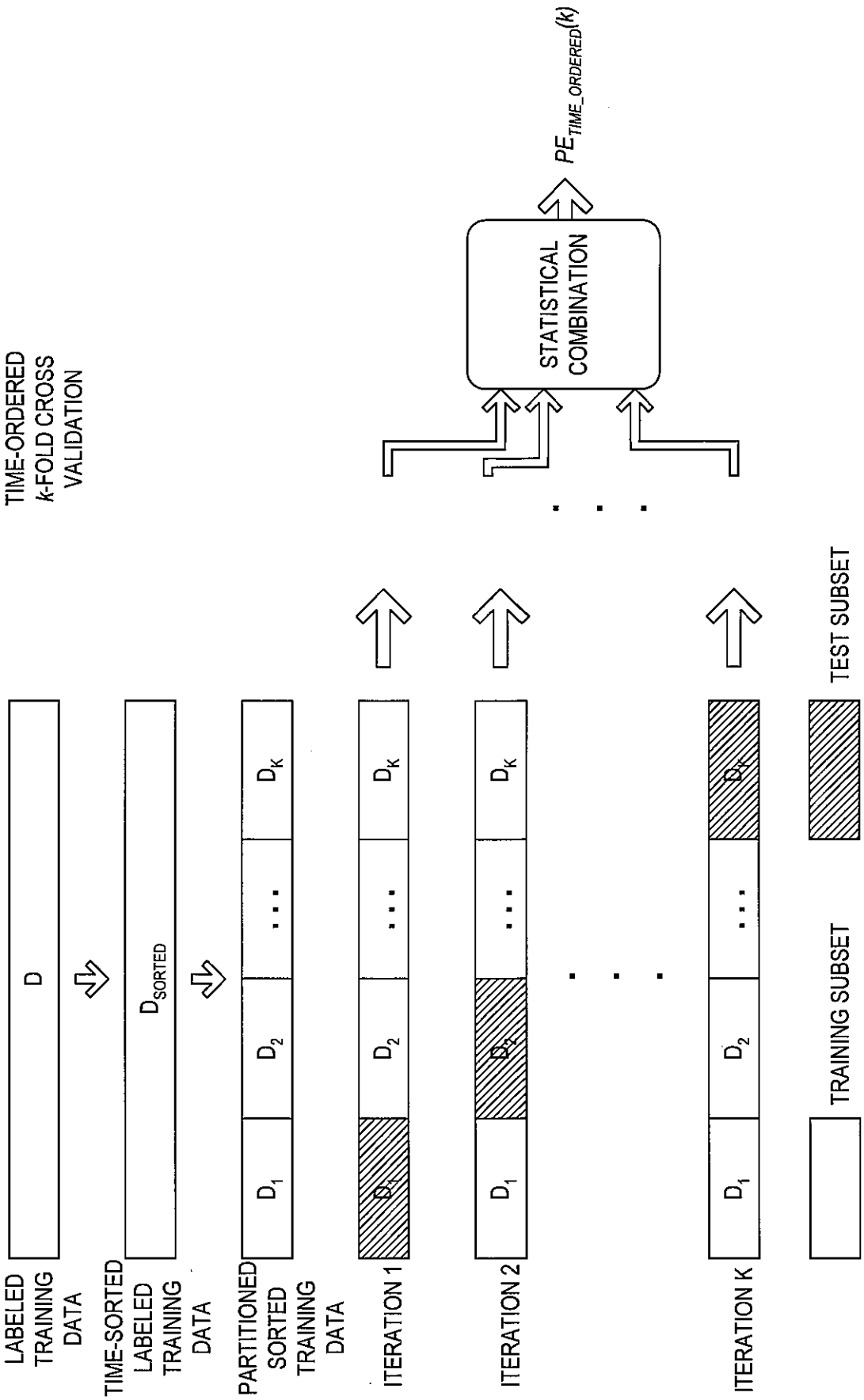

TIME-ORDERED
k-FOLD CROSS
VALIDATION
LABELED
TRAINING
DATA
D
TIME-SORTED
LABELED
TRAINING
DATA
D_{SORTED}
PARTITIONED
SORTED
TRAINING
DATA
D_1
D_2
...
D_K
ITERATION 1
ITERATION 2
ITERATION K
STATISTICAL
COMBINATION
$PE_{TIME_ORDERED}(k)$
TRAINING SUBSET
TEST SUBSET

FIG. 5A

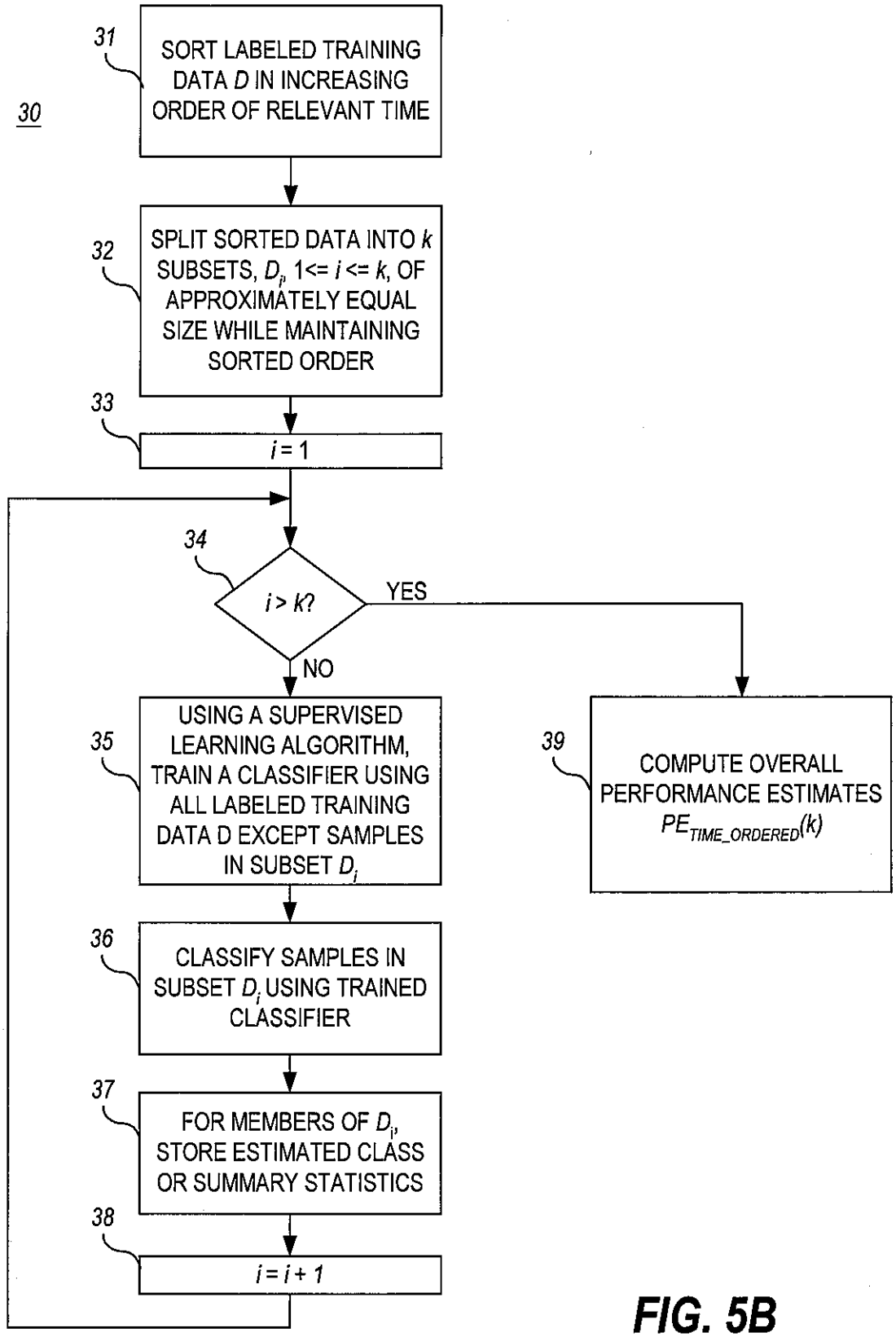

30
31
SORT LABELED TRAINING DATA D IN INCREASING ORDER OF RELEVANT TIME
32
SPLIT SORTED DATA INTO k SUBSETS, D_i, 1<= i <= k, OF APPROXIMATELY EQUAL SIZE WHILE MAINTAINING SORTED ORDER
33
i = 1
34
i > k?
YES
NO
35
USING A SUPERVISED LEARNING ALGORITHM, TRAIN A CLASSIFIER USING ALL LABELED TRAINING DATA D EXCEPT SAMPLES IN SUBSET D_i
36
CLASSIFY SAMPLES IN SUBSET D_i USING TRAINED CLASSIFIER
37
FOR MEMBERS OF D_i, STORE ESTIMATED CLASS OR SUMMARY STATISTICS
38
i = i + 1
39
COMPUTE OVERALL PERFORMANCE ESTIMATES $PE_{TIME_ORDERED}(k)$

FIG. 5B

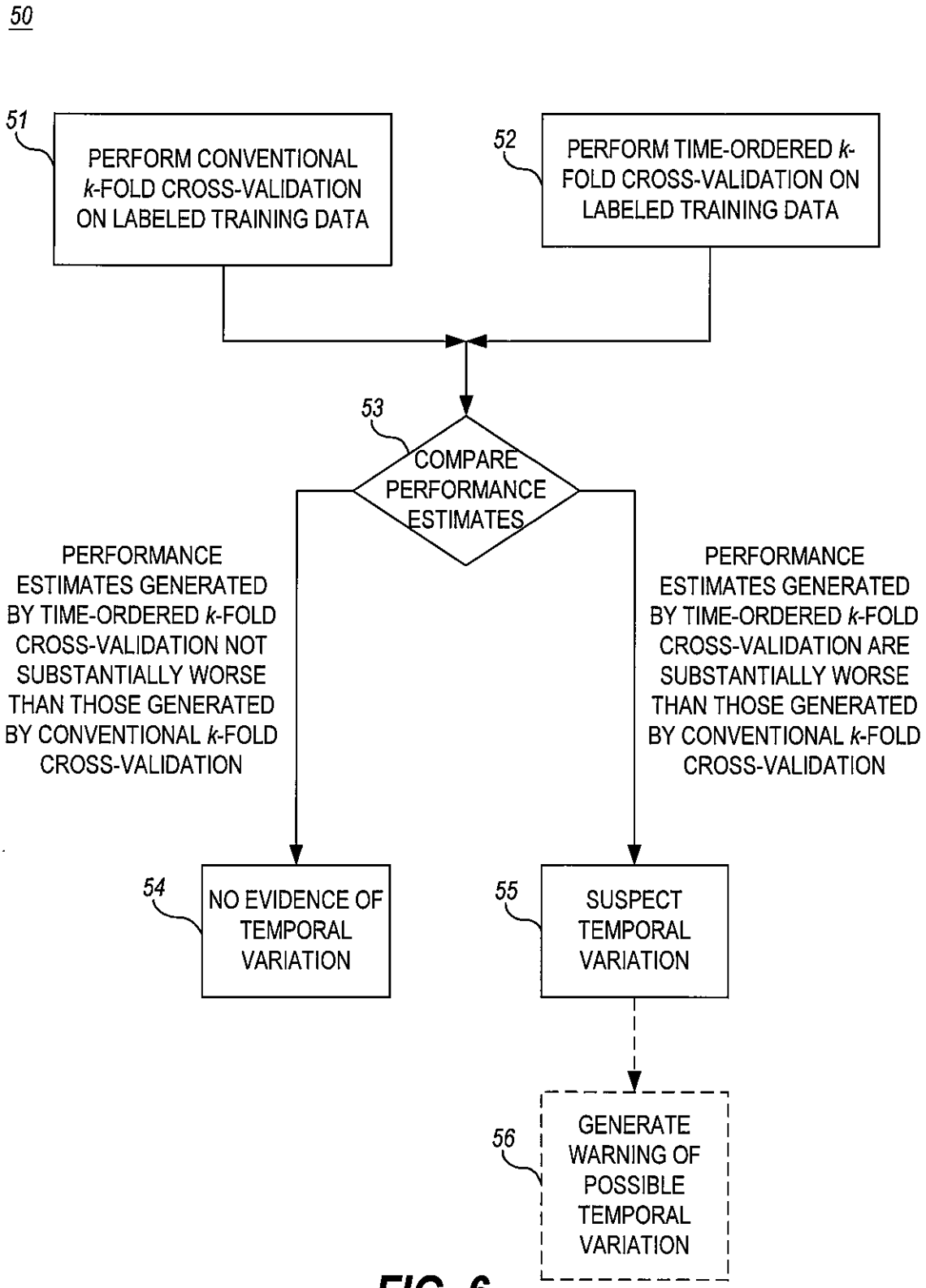

50
51
PERFORM CONVENTIONAL k-FOLD CROSS-VALIDATION ON LABELED TRAINING DATA
52
PERFORM TIME-ORDERED k-FOLD CROSS-VALIDATION ON LABELED TRAINING DATA
53
COMPARE PERFORMANCE ESTIMATES
PERFORMANCE ESTIMATES GENERATED BY TIME-ORDERED k-FOLD CROSS-VALIDATION NOT SUBSTANTIALLY WORSE THAN THOSE GENERATED BY CONVENTIONAL k-FOLD CROSS-VALIDATION
PERFORMANCE ESTIMATES GENERATED BY TIME-ORDERED k-FOLD CROSS-VALIDATION ARE SUBSTANTIALLY WORSE THAN THOSE GENERATED BY CONVENTIONAL k-FOLD CROSS-VALIDATION
54
NO EVIDENCE OF TEMPORAL VARIATION
55
SUSPECT TEMPORAL VARIATION
56
GENERATE WARNING OF POSSIBLE TEMPORAL VARIATION

FIG. 6

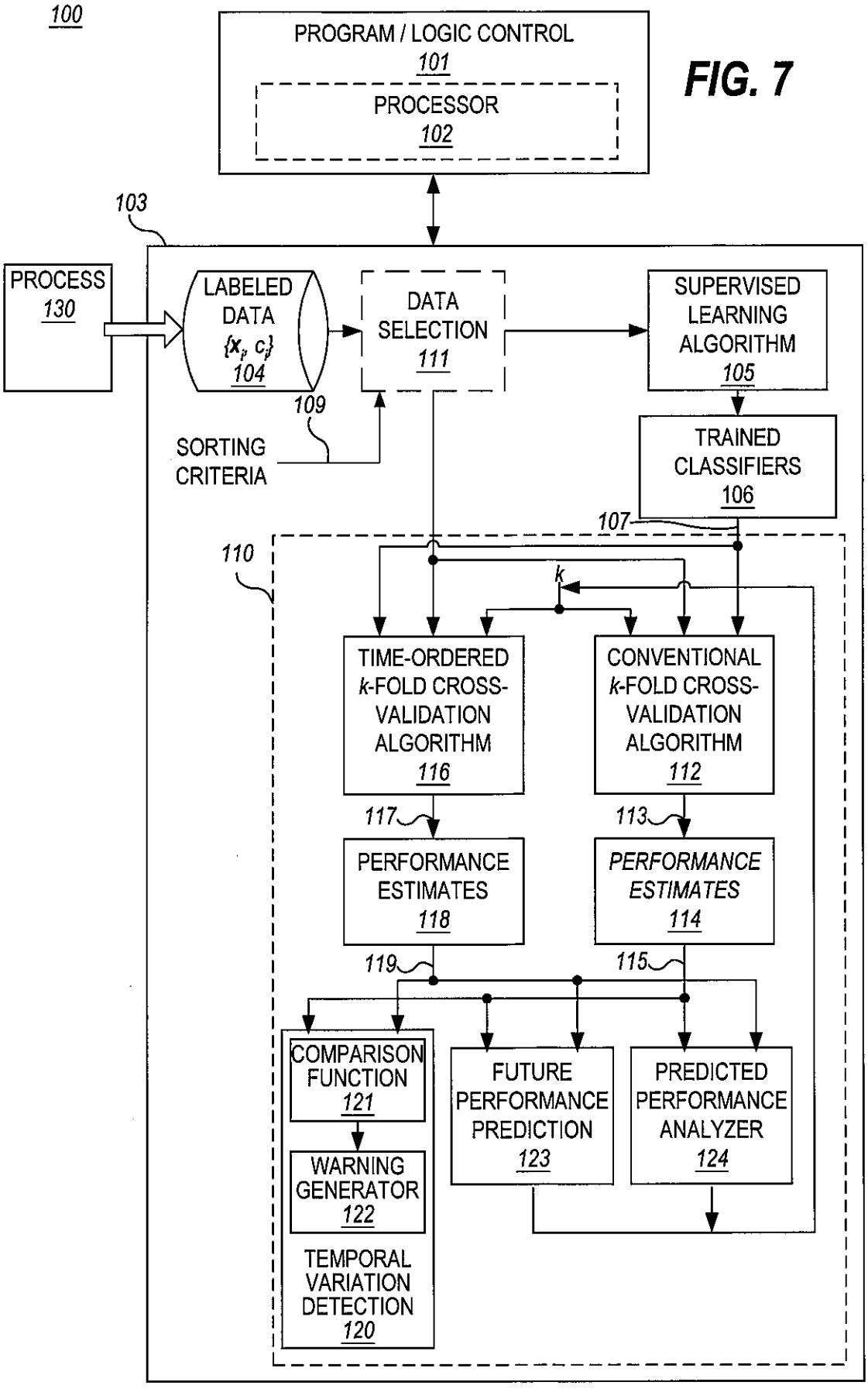

100
PROGRAM / LOGIC CONTROL
101
PROCESSOR
102
103
PROCESS
130
LABELED DATA
{x_i, c_i}
104
DATA SELECTION
111
SUPERVISED LEARNING ALGORITHM
105
109
SORTING CRITERIA
TRAINED CLASSIFIERS
106
107
110
k
TIME-ORDERED k-FOLD CROSS-VALIDATION ALGORITHM
116
CONVENTIONAL k-FOLD CROSS-VALIDATION ALGORITHM
112
117
113
PERFORMANCE ESTIMATES
118
PERFORMANCE ESTIMATES
114
119
115
COMPARISON FUNCTION
121
WARNING GENERATOR
122
TEMPORAL VARIATION DETECTION
120
FUTURE PERFORMANCE PREDICTION
123
PREDICTED PERFORMANCE ANALYZER
124

FIG. 7

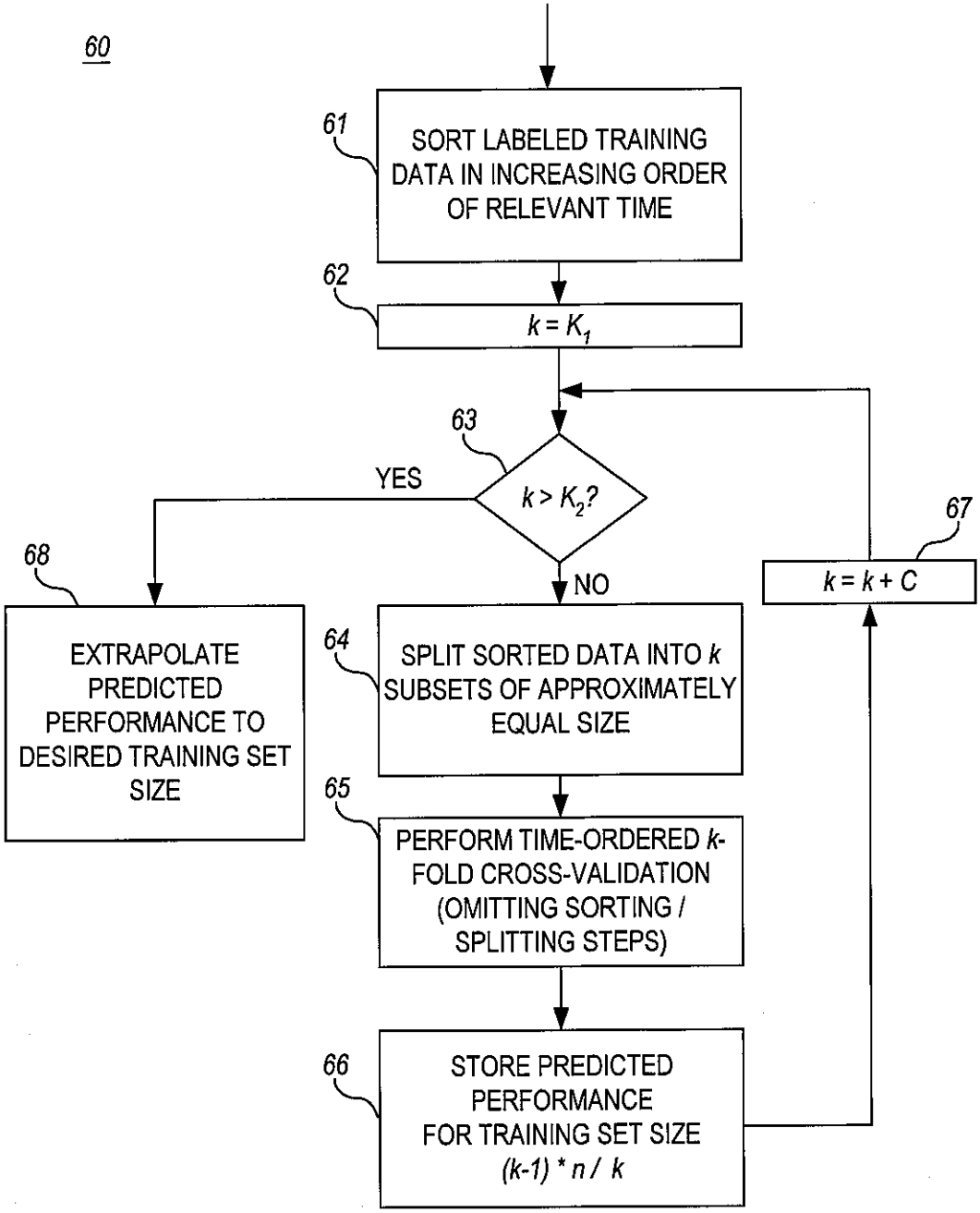

60
61
SORT LABELED TRAINING DATA IN INCREASING ORDER OF RELEVANT TIME
62
$k = K_1$
63
$k > K_2$?
YES
NO
67
$k = k + C$
68
EXTRAPOLATE PREDICTED PERFORMANCE TO DESIRED TRAINING SET SIZE
64
SPLIT SORTED DATA INTO k SUBSETS OF APPROXIMATELY EQUAL SIZE
65
PERFORM TIME-ORDERED k-FOLD CROSS-VALIDATION (OMITTING SORTING / SPLITTING STEPS)
66
STORE PREDICTED PERFORMANCE FOR TRAINING SET SIZE $(k-1) * n / k$

FIG. 8

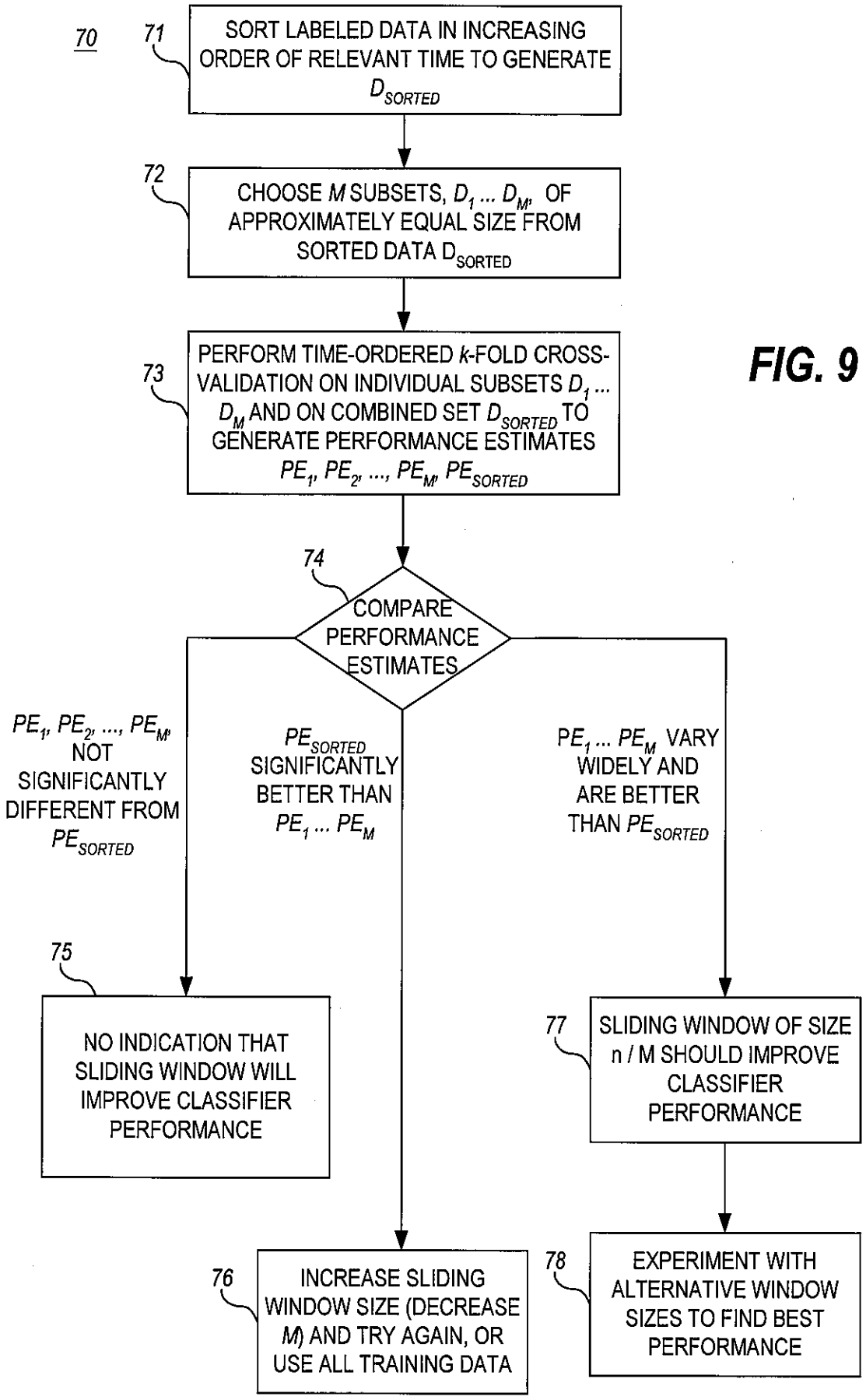

70
71
SORT LABELED DATA IN INCREASING ORDER OF RELEVANT TIME TO GENERATE D_{SORTED}
72
CHOOSE M SUBSETS, $D_1 \ldots D_M$, OF APPROXIMATELY EQUAL SIZE FROM SORTED DATA D_{SORTED}
73
PERFORM TIME-ORDERED k-FOLD CROSS-VALIDATION ON INDIVIDUAL SUBSETS $D_1 \ldots D_M$ AND ON COMBINED SET D_{SORTED} TO GENERATE PERFORMANCE ESTIMATES $PE_1, PE_2, \ldots, PE_M, PE_{SORTED}$
74
COMPARE PERFORMANCE ESTIMATES
$PE_1, PE_2, \ldots, PE_M$ NOT SIGNIFICANTLY DIFFERENT FROM PE_{SORTED}
PE_{SORTED} SIGNIFICANTLY BETTER THAN $PE_1 \ldots PE_M$
$PE_1 \ldots PE_M$ VARY WIDELY AND ARE BETTER THAN PE_{SORTED}
75
NO INDICATION THAT SLIDING WINDOW WILL IMPROVE CLASSIFIER PERFORMANCE
77
SLIDING WINDOW OF SIZE n / M SHOULD IMPROVE CLASSIFIER PERFORMANCE
76
INCREASE SLIDING WINDOW SIZE (DECREASE M) AND TRY AGAIN, OR USE ALL TRAINING DATA
78
EXPERIMENT WITH ALTERNATIVE WINDOW SIZES TO FIND BEST PERFORMANCE

FIG. 9

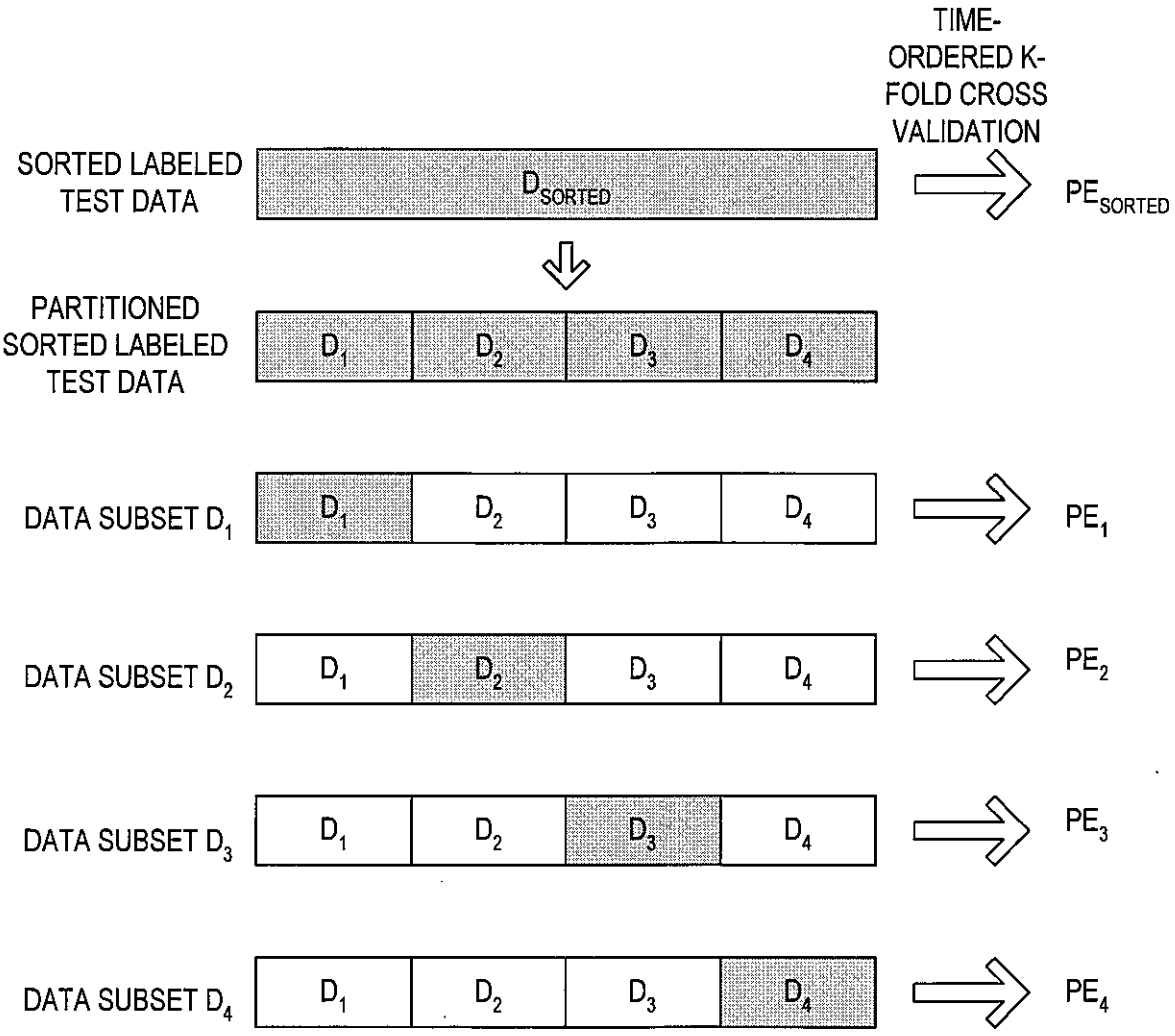

TIME-ORDERED K-FOLD CROSS VALIDATION
SORTED LABELED TEST DATA
D_{SORTED}
PE_{SORTED}
PARTITIONED SORTED LABELED TEST DATA
D_1 D_2 D_3 D_4
DATA SUBSET D_1
DATA SUBSET D_2
DATA SUBSET D_3
DATA SUBSET D_4
PE_1
PE_2
PE_3
PE_4

FIG. 10

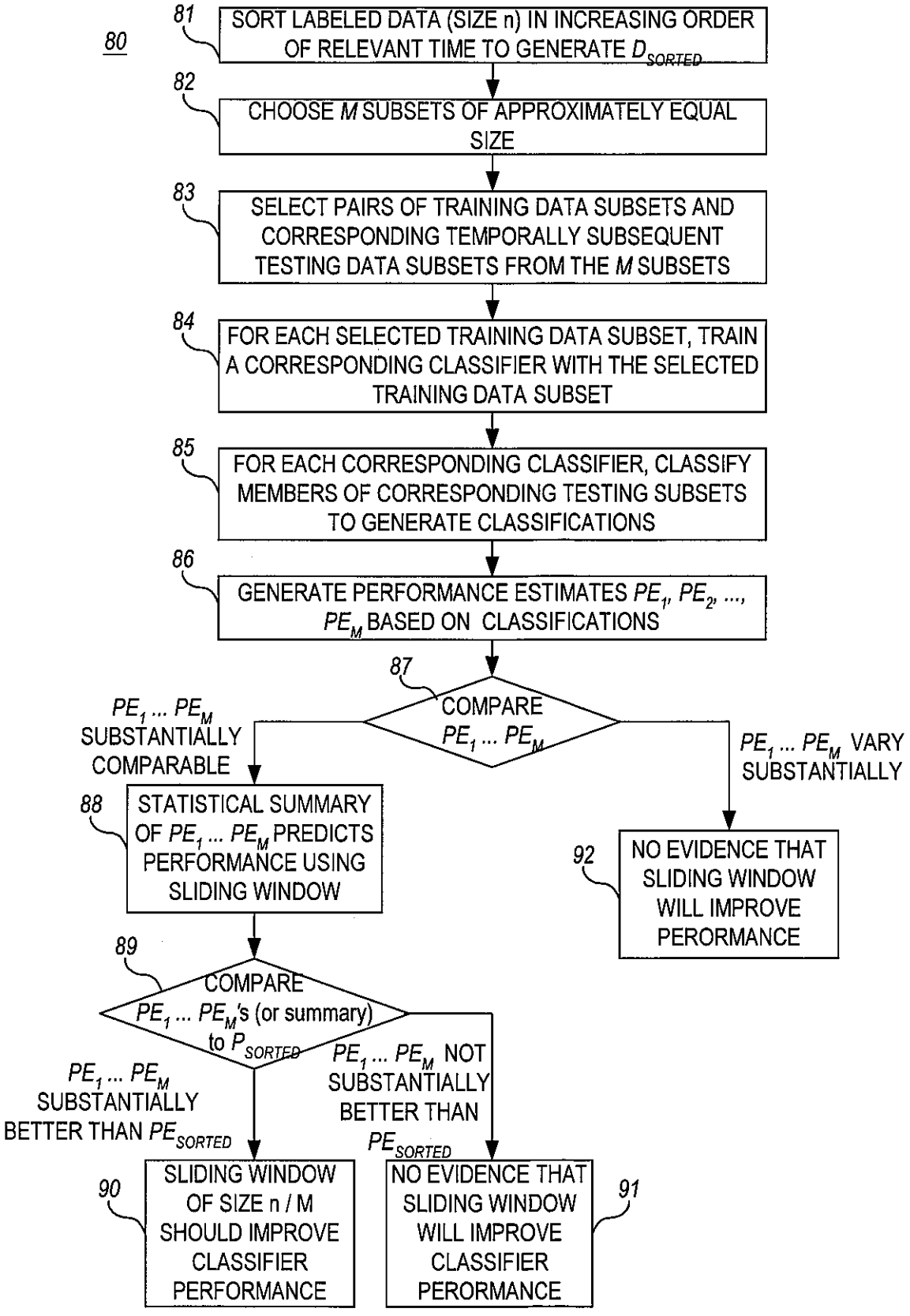

80
81
SORT LABELED DATA (SIZE n) IN INCREASING ORDER OF RELEVANT TIME TO GENERATE D_{SORTED}
82
CHOOSE M SUBSETS OF APPROXIMATELY EQUAL SIZE
83
SELECT PAIRS OF TRAINING DATA SUBSETS AND CORRESPONDING TEMPORALLY SUBSEQUENT TESTING DATA SUBSETS FROM THE M SUBSETS
84
FOR EACH SELECTED TRAINING DATA SUBSET, TRAIN A CORRESPONDING CLASSIFIER WITH THE SELECTED TRAINING DATA SUBSET
85
FOR EACH CORRESPONDING CLASSIFIER, CLASSIFY MEMBERS OF CORRESPONDING TESTING SUBSETS TO GENERATE CLASSIFICATIONS
86
GENERATE PERFORMANCE ESTIMATES PE_1, PE_2, ..., PE_M BASED ON CLASSIFICATIONS
87
COMPARE PE_1 ... PE_M
PE_1 ... PE_M SUBSTANTIALLY COMPARABLE
PE_1 ... PE_M VARY SUBSTANTIALLY
88
STATISTICAL SUMMARY OF PE_1 ... PE_M PREDICTS PERFORMANCE USING SLIDING WINDOW
92
NO EVIDENCE THAT SLIDING WINDOW WILL IMPROVE PERORMANCE
89
COMPARE PE_1 ... PE_M's (or summary) to P_{SORTED}
PE_1 ... PE_M SUBSTANTIALLY BETTER THAN PE_{SORTED}
PE_1 ... PE_M NOT SUBSTANTIALLY BETTER THAN PE_{SORTED}
90
SLIDING WINDOW OF SIZE n / M SHOULD IMPROVE CLASSIFIER PERFORMANCE
91
NO EVIDENCE THAT SLIDING WINDOW WILL IMPROVE CLASSIFIER PERORMANCE

FIG. 11

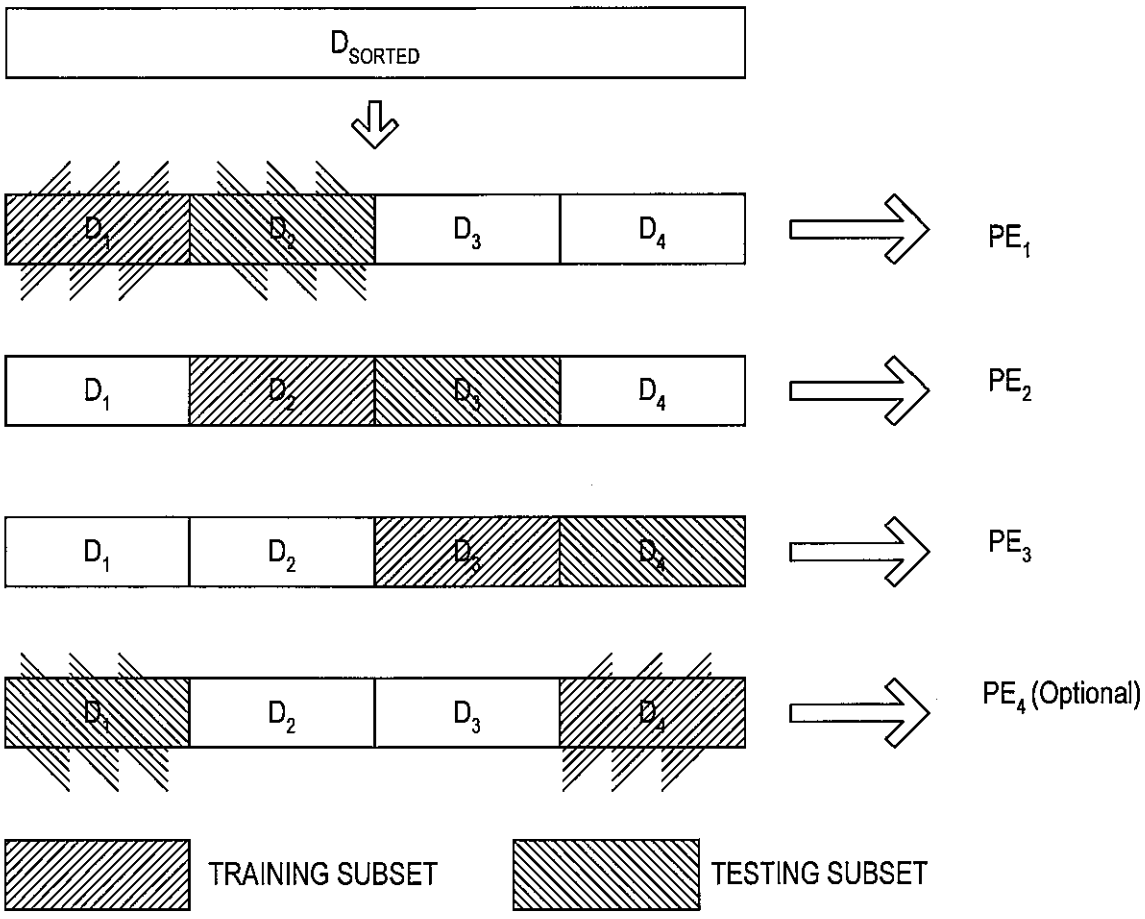

D_{SORTED}
D_1
D_2
D_3
D_4
PE_1
PE_2
PE_3
PE_4 (Optional)
TRAINING SUBSET
TESTING SUBSET

FIG. 12