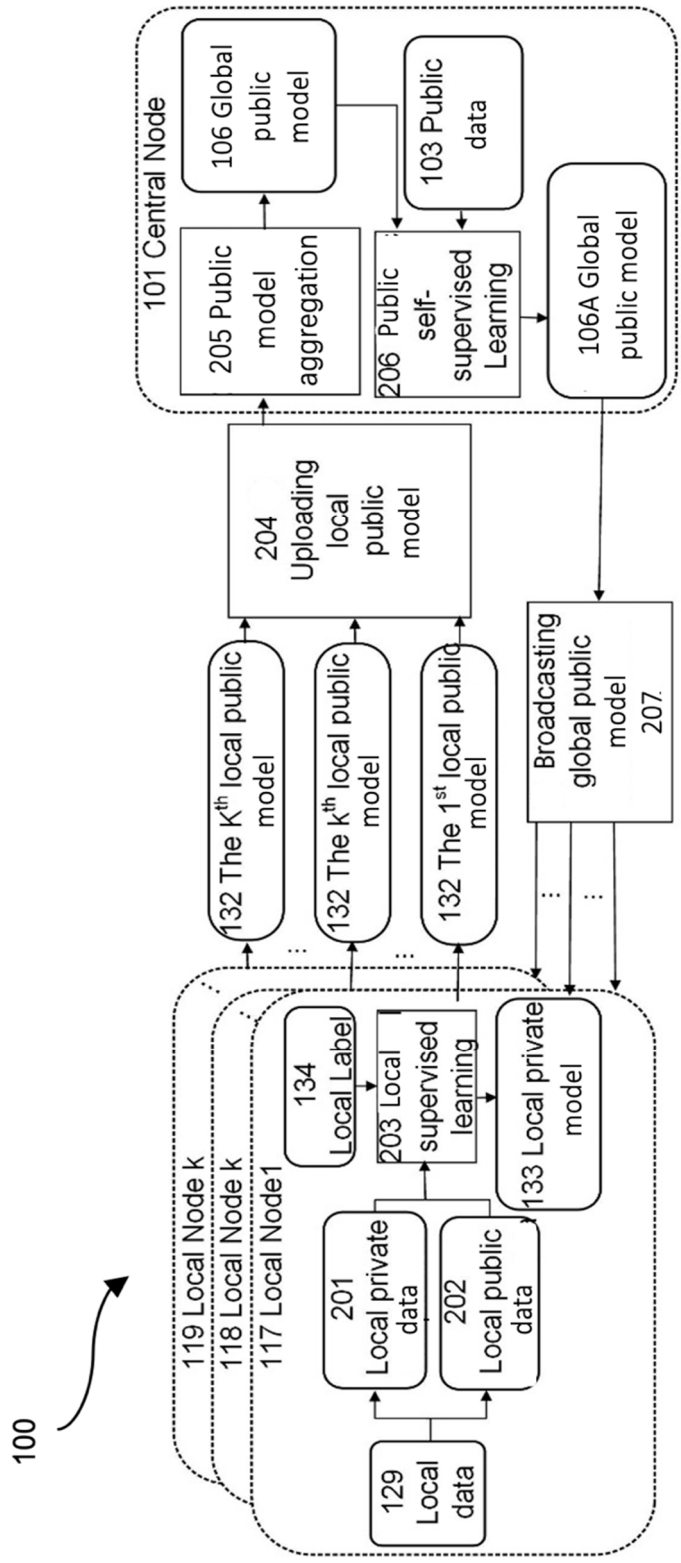
100
101 Central Node
106 Global public model
103 Public data
205 Public model aggregation
206 Public self-supervised Learning
106A Global public model
204 Uploading local public model
132 The K^th local public model
132 The k^th local public model
132 The 1^st local public model
Broadcasting global public model 207
119 Local Node k
118 Local Node k
117 Local Node1
134 Local Label
203 Local supervised learning
133 Local private model
201 Local private data
202 Local public data
129 Local data

FIG. 1

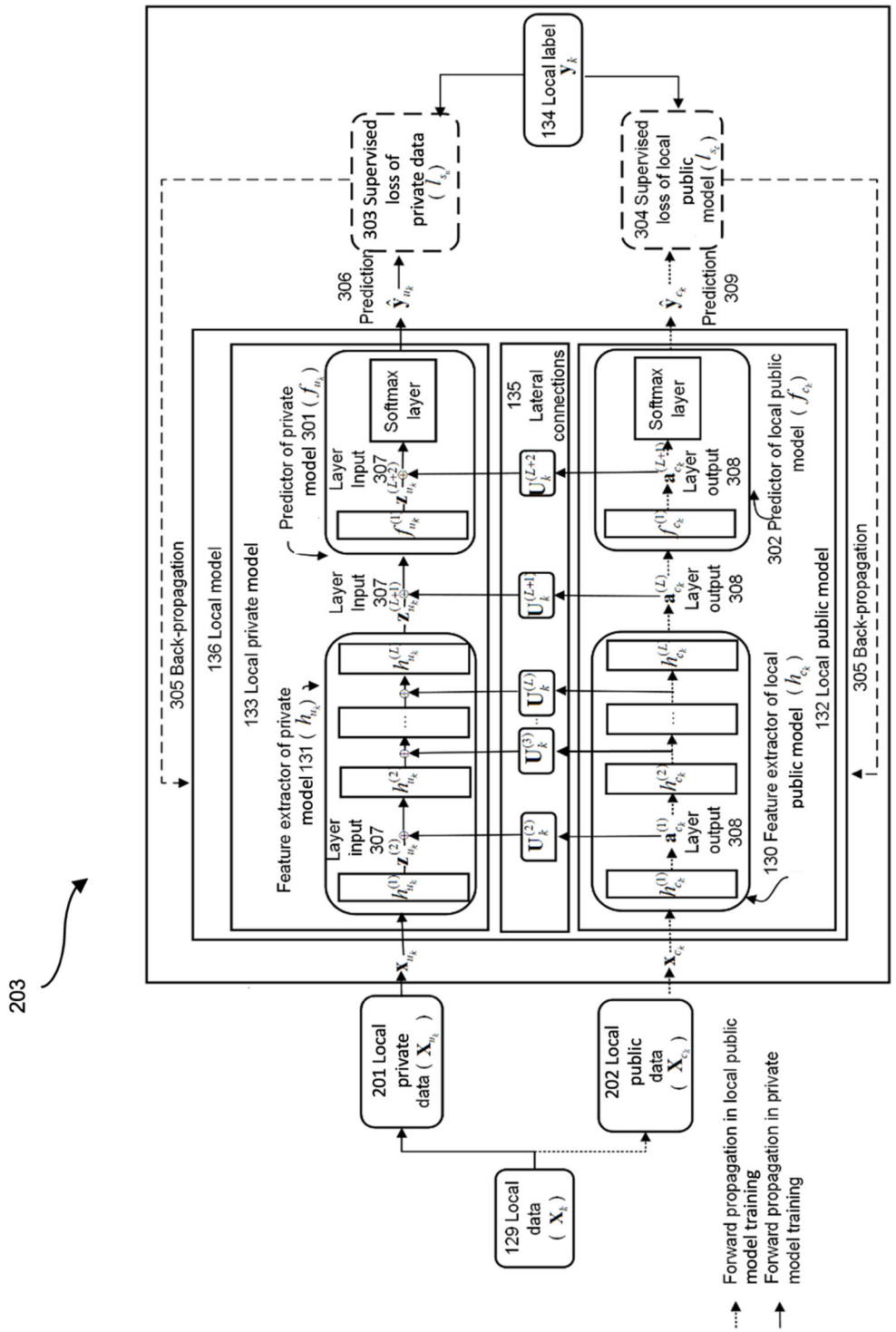
203
129 Local data (X_k)
201 Local private data (X_{u_k})
202 Local public data (X_{c_k})
Forward propagation in local public model training
Forward propagation in private model training
305 Back-propagation
136 Local model
133 Local private model
Feature extractor of private model 131 (h_{u_k})
Predictor of private model 301 (f_{u_k})
Layer input 307
Layer Input 307
Softmax layer
306 Prediction $\hat{y}_{u_k}$
303 Supervised loss of private data (l_{u_k})
134 Local label y_k
135 Lateral connections
130 Feature extractor of local public model (h_{c_k})
132 Local public model
Layer output 308
302 Predictor of local public model (f_{c_k})
304 Supervised loss of local public model (l_{c_k})
309 Prediction $\hat{y}_{c_k}$

FIG. 2

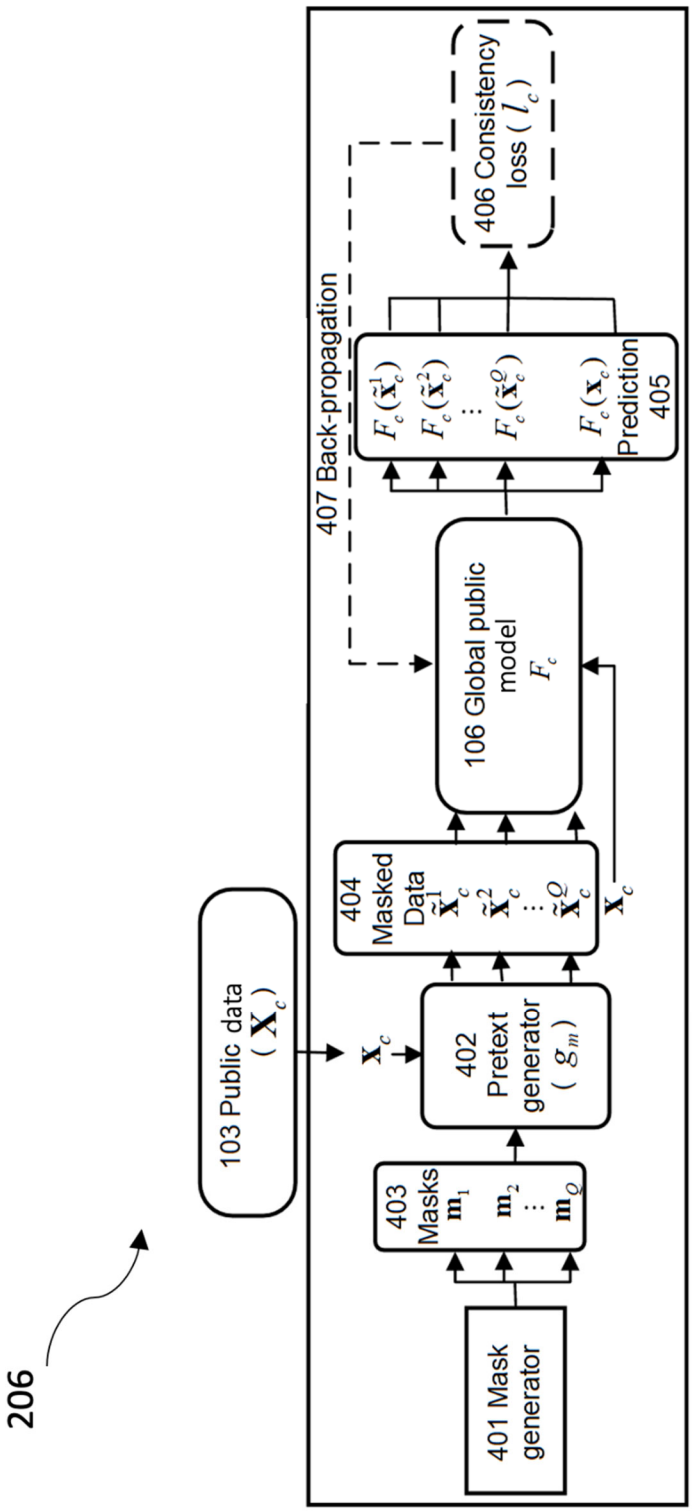
206
103 Public data ($\mathbf{X}_c$)
401 Mask generator
403 Masks $\mathbf{m}_1$ $\mathbf{m}_2$ $\mathbf{m}_Q$
$\mathbf{X}_c$
402 Pretext generator (g_m)
404 Masked Data $\tilde{\mathbf{X}}_c^1$ $\tilde{\mathbf{X}}_c^2$ $\tilde{\mathbf{X}}_c^Q$ $\mathbf{X}_c$
106 Global public model F_c
407 Back-propagation
$F_c(\tilde{\mathbf{x}}_c^1)$ $F_c(\tilde{\mathbf{x}}_c^2)$ $F_c(\tilde{\mathbf{x}}_c^Q)$ $F_c(\mathbf{x}_c)$ Prediction 405
406 Consistency loss (l_c)

FIG. 3

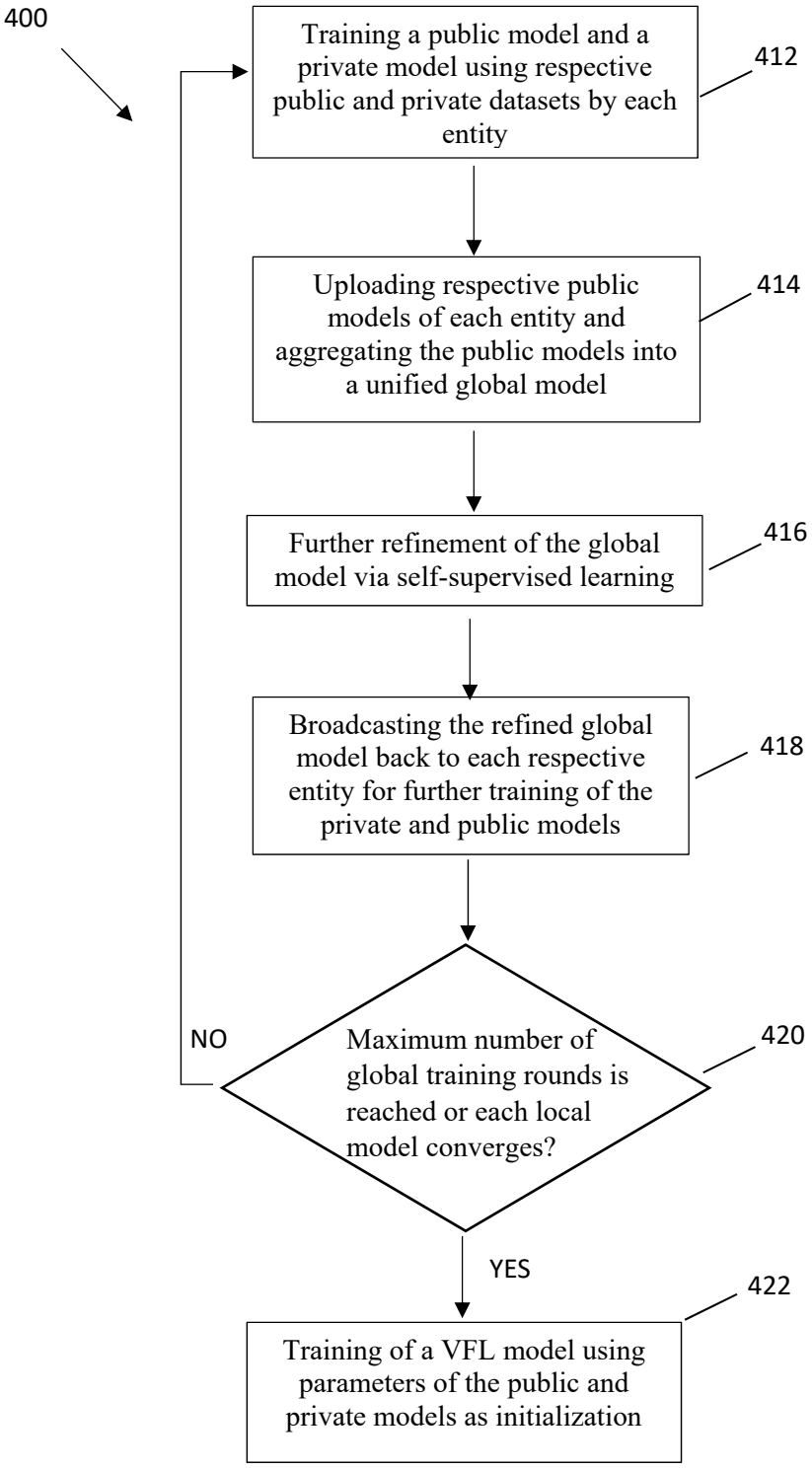
400
Training a public model and a private model using respective public and private datasets by each entity
412
Uploading respective public models of each entity and aggregating the public models into a unified global model
414
Further refinement of the global model via self-supervised learning
416
Broadcasting the refined global model back to each respective entity for further training of the private and public models
418
Maximum number of global training rounds is reached or each local model converges?
420
NO
YES
Training of a VFL model using parameters of the public and private models as initialization
422

FIG. 4

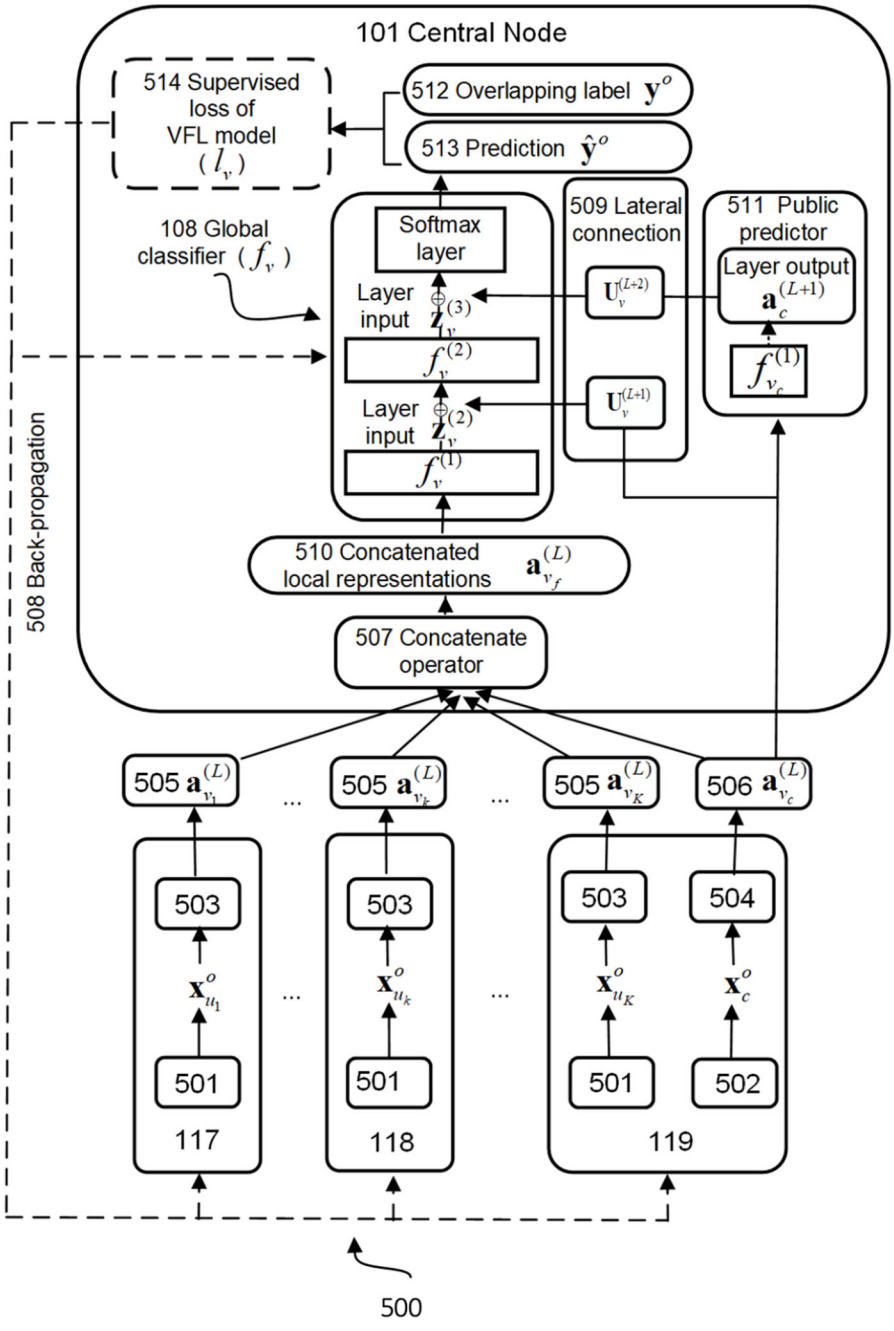
101 Central Node
514 Supervised loss of VFL model (l_v)
512 Overlapping label $\mathbf{y}^o$
513 Prediction $\hat{\mathbf{y}}^o$
108 Global classifier (f_v)
Softmax layer
Layer input $\mathbf{z}_v^{(3)}$
$f_v^{(2)}$
Layer input $\mathbf{z}_v^{(2)}$
$f_v^{(1)}$
509 Lateral connection
$\mathbf{U}_v^{(L+2)}$
$\mathbf{U}_v^{(L+1)}$
511 Public predictor
Layer output $\mathbf{a}_c^{(L+1)}$
$f_{v_c}^{(1)}$
510 Concatenated local representations $\mathbf{a}_{v_f}^{(L)}$
507 Concatenate operator
508 Back-propagation
505 $\mathbf{a}_{v_1}^{(L)}$
505 $\mathbf{a}_{v_k}^{(L)}$
505 $\mathbf{a}_{v_K}^{(L)}$
506 $\mathbf{a}_{v_c}^{(L)}$
503
504
$\mathbf{x}_{u_1}^o$
$\mathbf{x}_{u_k}^o$
$\mathbf{x}_{u_K}^o$
$\mathbf{x}_c^o$
501
502
117
118
119
500

FIG. 5

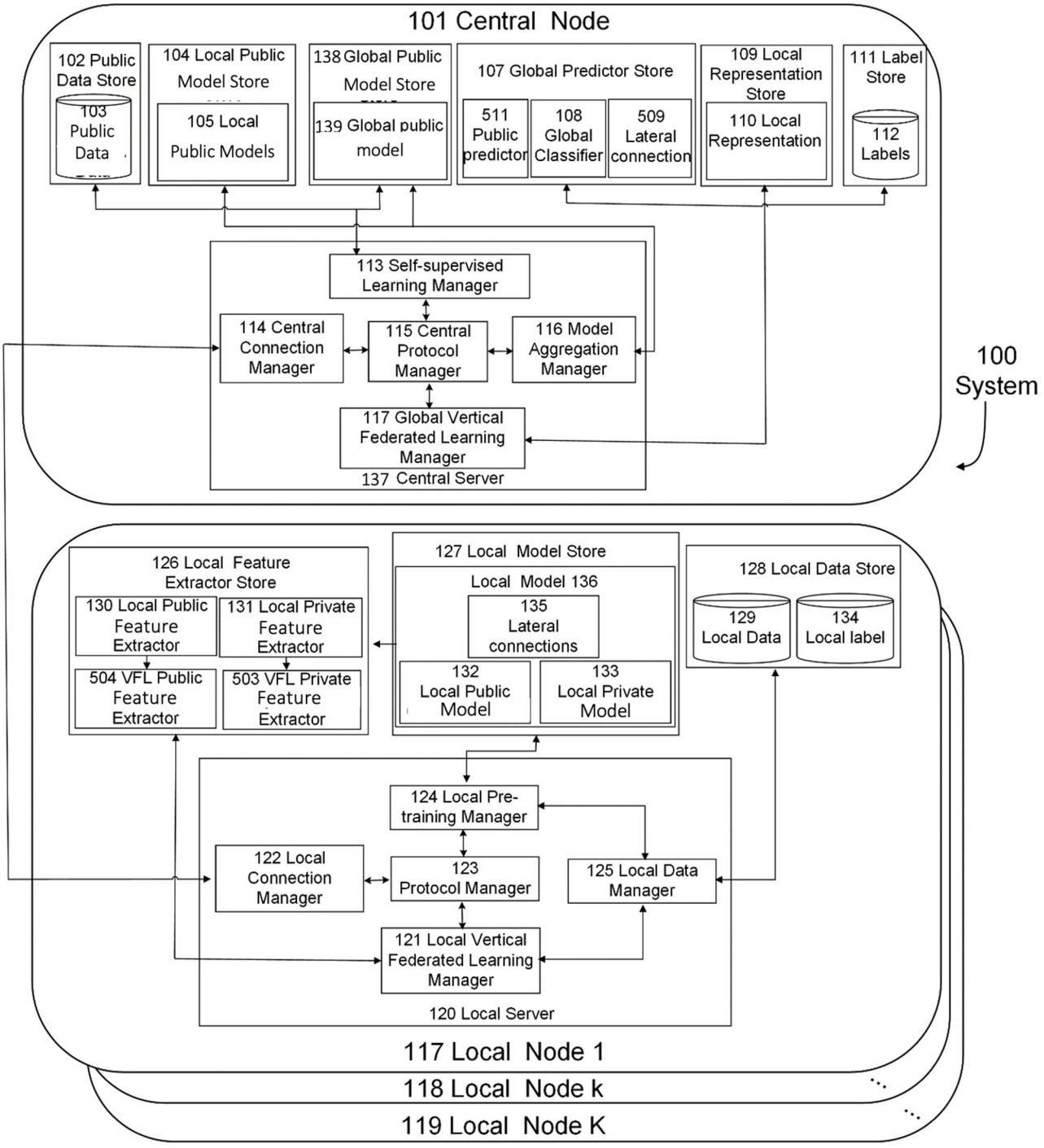
101 Central Node
102 Public Data Store
103 Public Data
104 Local Public Model Store
105 Local Public Models
138 Global Public Model Store
139 Global public model
107 Global Predictor Store
511 Public predictor
108 Global Classifier
509 Lateral connection
109 Local Representation Store
110 Local Representation
111 Label Store
112 Labels
113 Self-supervised Learning Manager
114 Central Connection Manager
115 Central Protocol Manager
116 Model Aggregation Manager
117 Global Vertical Federated Learning Manager
137 Central Server
100 System
126 Local Feature Extractor Store
130 Local Public Feature Extractor
131 Local Private Feature Extractor
504 VFL Public Feature Extractor
503 VFL Private Feature Extractor
127 Local Model Store
Local Model 136
135 Lateral connections
132 Local Public Model
133 Local Private Model
128 Local Data Store
129 Local Data
134 Local label
124 Local Pre-training Manager
122 Local Connection Manager
123 Protocol Manager
125 Local Data Manager
121 Local Vertical Federated Learning Manager
120 Local Server
117 Local Node 1
118 Local Node k
119 Local Node K

FIG. 6

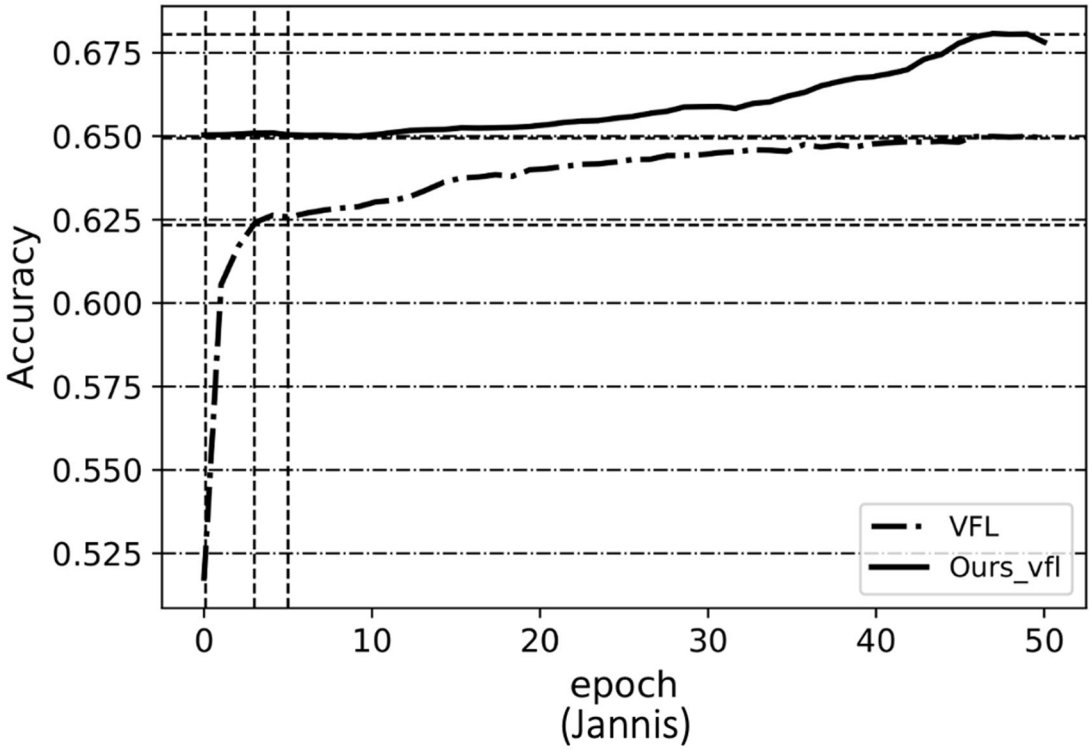
0.675
0.650
0.625
0.600
0.575
0.550
0.525
Accuracy
0
10
20
30
40
50
epoch
(Jannis)
VFL
Ours_vfl

FIG. 7A

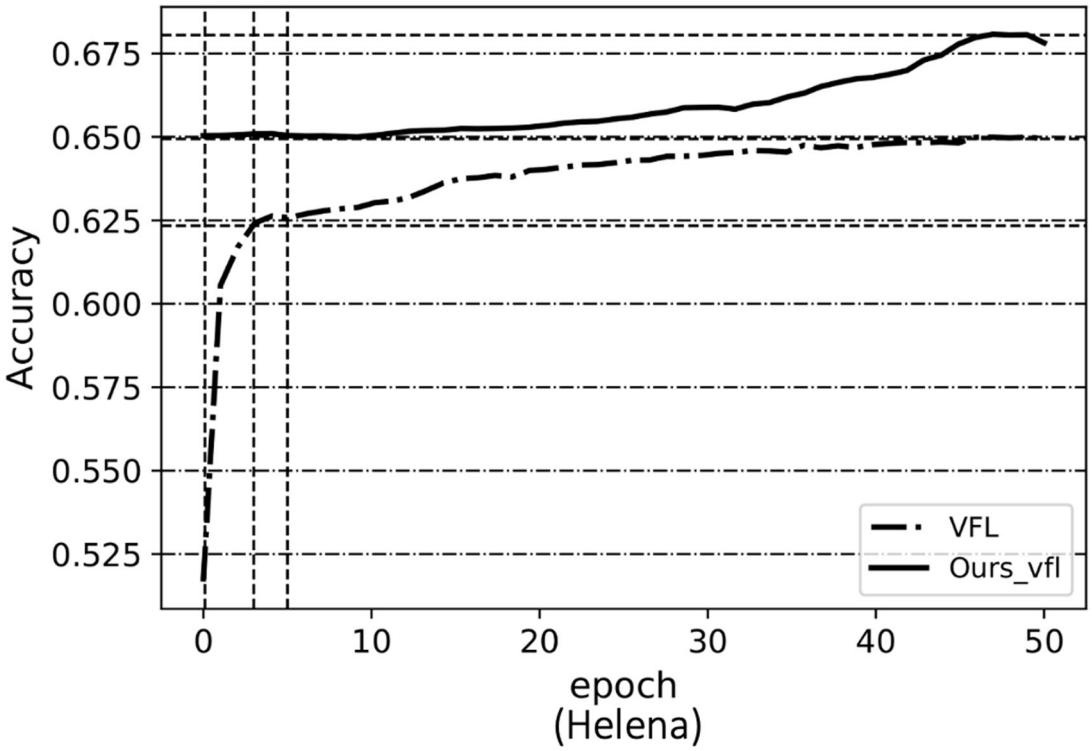
0.675
0.650
0.625
0.600
0.575
0.550
0.525
Accuracy
0
10
20
30
40
50
epoch
(Helena)
VFL
Ours_vfl

FIG. 7B

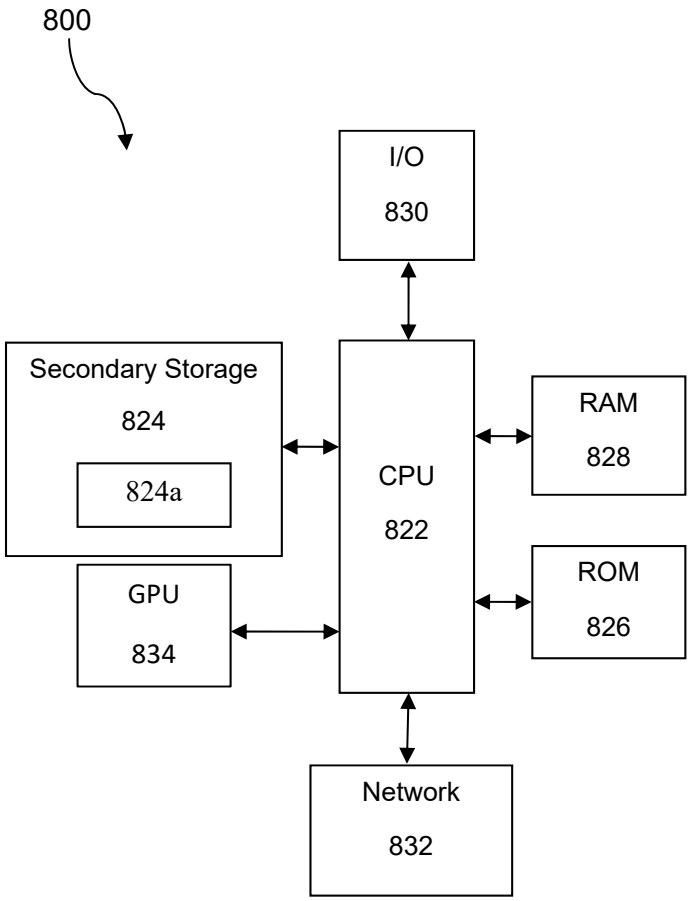
800
I/O
830
Secondary Storage
824
824a
CPU
822
RAM
828
GPU
834
ROM
826
Network
832

FIG. 8