	hyperparameter	MNIST	Fashion-MNIST	NORB	Segment	Mfeat	Prior	Advertisement	Gina agnostic	Spambase	Isolet	Sylva	Har	Abalone	Amazon	Gisette	Kin8nm	Madelon	Occupancy	Semeion	Yeast
AE	epochs	50	50	50	30	50	30	30	30	50	30	30	30	30	30	30	30	30	30	50	30
	learning rate	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	batch size	64	64	64	16	64	16	64	32	16	32	32	64	32	32	64	32	32	32	16	16
	hidden neurons	[2000,2000]	[2000]	[2000,2000]	[700]	[2500]	[1500,200,1000]	[1500]	[800,600]	[2500]	[1500]	[1000]	[1500]	[500,300]	[1200]	[2000,1500]	[1500]	[2000,1000,500,300]	[2500]	[1500,1000,500]	[2000,1000]
RBM	epochs	50	50	50	30	50	30	30	30	30	30	30	30	30	30	30	30	30	30	50	30
	learning rate	0.001	0.001	0.001	0.001	0.001	0.0001	0.001	0.0001	0.001	0.001	0.001	0.0001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	batch size	64	64	64	64	64	64	64	32	16	64	64	64	64	64	64	32	64	64	8	32
	Gibbs sampling	1	1	1	2	2	2	1	2	2	2	1	2	2	1	2	1	1	2	2	1
	hidden neurons	[2000,2000]	[2000]	[2000,2000]	[700]	[2500]	[1500,200,1000]	[1500]	[800,600]	[2500]	[1500]	[1000]	[1500]	[500,300]	[1200]	[2000,1500]	[1500]	[2000,1000,500,300]	[2500]	[1500,1000,500]	[2000,1000]
HELM	N1	700	700	3000	150	1500	1500	1200	1200	420	420	420	1100	500	1000	1000	1000	1000	1000	1000	500
	N2	700	700	3000	120	1200	1200	1000	1000	450	400	400	500	500	500	500	500	500	500	500	500
	N3	12000	12000	15000	2000	4000	2000	2000	2000	800	7000	1000	800	5000	5000	8000	6000	8000	5000	1500	1500
ELM-AE	hidden neurons	[700,15000]	[700,700,15000]	[3000,3000,15000]	[3000,2000]	[3000,1000]	[2000,1000]	[2000,1500]	[5000,2000]	[1500,2000]	[3000,2000]	[3000,2000]	[3000,2000]	[1500,2000]	[7000,2000]	[1500,2000]	[1500,2000]	[5000,2000]	[3000,2000]	[5000,2000]	[1500,2000]
BLS	feature nodes	10	10	100	350	130	150	130	380	130	130	380	120	200	100	100	100	100	100	100	100
	number of windows of feature nodes	10	10	10	6	6	6	6	10	6	6	10	6	20	20	10	20	20	20	20	20
	enhancement nodes	11000	11000	9000	2500	2500	9000	300	1000	2000	2000	1000	1000	2000	200	4000	2000	5000	5000	5000	2000
stacked BLS	feature nodes	[10,5,2,1,1]	[100,8,8]	[100,1,1]	[50,24,11]	[60,24]	[30,24,11]	[50]	[50]	[100,24]	[50,24]	[30]	[50,24]	[20,24]	[20,24,11]	[20,24]	[50,24,11]	[50,24]	[50,24]	[20]	[100,24,11]
	number of windows of feature nodes	[10,5,7,3,1]	[10,9,9]	[10,1,1]	[30,8,7]	[50,8]	[50,8,7]	[30]	[30]	[50,8]	[70,8]	[50]	[30,8]	[100,8]	[50,8,7]	[50,8]	[80,8,7]	[30,8]	[50,8]	[10]	[20,8,7]
	enhancement nodes	[11000,470,6,7,19]	[9000,380,380]	[3000,400,400]	[500,100,50]	[1500,100]	[200,100,50]	[500]	[500]	[500,100]	[600,100]	[300]	[500,100]	[200,100]	[100,100,50]	[300,100]	[500,100,50]	[500,100]	[200,100]	[1000]	[200,100,50]
NF-RVFL	iter	5	3	3	3	2	3	1	1	2	2	1	2	2	3	2	3	2	2	1	3
	Fuzzy Nodes	9000	9000	250	150	15	5	5	20	15	15	15	15	5	10	5	5	15	5	20	10
	Hidden Nodes	1000	1000	700	2000	2000	10	20	200	100	30	200	100	20	20	30	3	200	60	60	60
	DictSize	13	15	30	10	5	15	10	10	10	5	10	5	10	2	5	10	5	5	15	5
SLatDPL	alpha	0.05	0.05	0.6	0.05	0.05	0.05	0.05	0.05	0.005	0.05	0.05	0.05	0.05	0.05	0.005	0.5	0.005	0.005	0.000008	0.05
	beta	0.005	0.005	0.0005	0.005	0.005	0.05	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.05	0.05	0.0005	0.005
	gamma	0.3	0.3	0.2	0.7	0.5	0.3	0.8	0.8	0.5	0.3	0.8	0.8	0.1	0.8	0.5	0.8	0.05	0.8	0.8	0.1
	Iter	8	6	2	20	25	20	20	30	10	20	20	10	20	1	10	5	8	10	20	10
ViT	batch size	128	64	128	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	embedding dimension	128	128	256	128	128	256	256	256	256	64	64	512	512	512	256	256	64	128	256	128
	attention heads	16	16	4	16	16	16	16	16	16	4	4	16	16	8	16	16	4	8	4	8
	layers	3	3	1	3	3	5	5	3	3	3	5	5	5	3	1	1	3	1	3	1
	hidden neurons	256	512	1024	256	1024	1024	1024	2000	1024	1024	1024	1000	1000	3000	1000	1000	1024	1000	1000	1000
	learning rate	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.00001	0.0001	0.0001	0.001	0.00001	0.0001	0.001	0.0001	0.0001	0.001	0.001	0.0001	0.001
	epochs	40	50	50	200	200	200	200	300	300	500	200	400	300	500	300	300	100	300	400	300
BViT	batch size	64	64	32	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	embedding dimension	256	256	512	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	attention heads	8	8	8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	layers	4	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	hidden neurons	256	256	512	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	learning rate	0.001	0.001	0.001	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	epochs	30	30	60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Bi-PIL	hidden neurons	[2000,2000,2000]	[2000,2000,2000]	[2000,2000,1000]	[700,700,600]	[1500,200,1000]	[1500,200,1000,1000,1000]	[1500,1000,500]	[800,600,500,300,300]	[2500,600,500,300,300]	[1500,1400,500]	[1000,700,600,500]	[1500,1000,500]	[1500,1500,500,500,300]	[1200,1000,800]	[2000,1500,500,400,300]	[1500,1000,500,300,200]	[2000,1000,500,300,200]	[2500,1500,500,400,300,300,300,200,200,200]	[1500,1000,500,300,200]	[2000,1000,800,300,200]
	lambda in forward training	2^-30	2^-30	2^-30	1E-12	2^-30	1E-12	1E-11	1E-11	0.001	1E-12	1.00E-10	1.00E-11	1.00E-10	0.001	1.00E-12	1.00E-12	1.00E-10	0.00001	0.00001	0.0000001
	lambda in backward training	2^-30	2^-30	2^-30	1E-15	2^-30	0.1	0.1	0.5	1	0.1	0.1	1.00E-09	0.01	1.00E-10	0.1	0.1	0.1	0.001	1	0.1