

Simple Feedforward Neural Networks are Almost All You Need for Time Series Forecasting

<https://arxiv.org/abs/2503.23621>

0. Introduction

- 시계열 예측은 금융, 에너지, 기상, 헬스케어 등 다양한 분야에서 중요
- 기존 모델들은 LSTM, GRU, Transformer 등 복잡하고 연산량 많음
- 복잡한 모델이 항상 더 나은 성능을 보이는 것은 아님
- 단순한 Feedforward Neural Network(SFNN)가 대부분 시계열 예측에서 충분히 강력함 보여줌
- 단순 구조임에도 다양한 데이터셋과 예측 작업에서 경쟁력 있는 성능 달성 가능

1. Overview

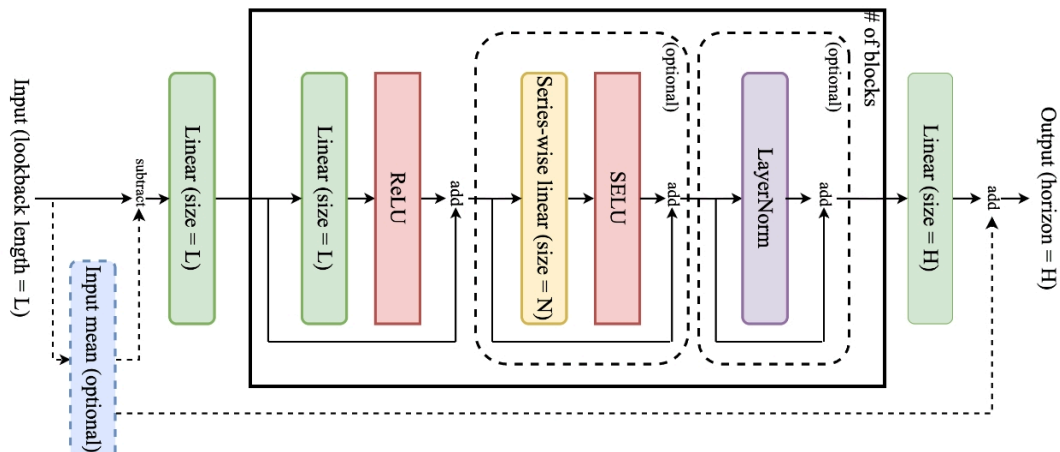
- SFNN 구조 : 입력 시계열 데이터를 Flatten 후 Fully Connected Layer 통과
- 모델 단순, 연산 효율 높음, 하드웨어 부담 적음
- 다변량/단일 변수, 단기/장기 예측 모두 적용 가능
- 기존 모델(LSTM, GRU, Transformer) 대비 성능 비교 중심

2. Challenges

- 시계열 장기 의존성 처리 어려움
- 다변량 시계열에서 변수 간 상관관계 학습 필요

- 단순 모델 구조로 인해 예측 성능 저하 가능성 존재
- 다양한 도메인에 범용적으로 적용 가능 여부 검증 필요

3. Method



- 입력 : 시계열 데이터를 Flatten
- 구조 : Dense Layer → 출력 (예측 Horizon 길이에 따라 Output Layer 크기 조정)
- 학습 : MSE 손실, Adam optimizer
- 비교 대상 : LSTM, GRU, Transformer 계열 모델

4. Experiments

- ETTh1, ETTh2 : 전력 소비 데이터, 1시간 단위 기록
- Electricity : 미국 전력 소비, 15분 단위 기록, 다변량
- Traffic : 교통량 시계열, 센서별 트래픽 데이터
- Weather : 기상 데이터, 온도/습도/풍속 포함
- 단기(1168시간) 예측 실험 수행
- SFNN과 LSTM/GRU/Transformer 모델 성능 비교

5. Results

Dataset	Horizon	SFNN	DUET [29]	iTransformer [23]
ETTh1	96	0.2790 ± 0.0003	$0.2775^{\dagger} \pm 0.0008$	0.3027 ± 0.0014
	192	$0.3161^{\dagger} \pm 0.0004$	0.3193 ± 0.0007	0.3450 ± 0.0023
	336	0.3448 ± 0.0018	0.3448 ± 0.0011	0.3786 ± 0.0022
	720	$0.3814^{\dagger} \pm 0.0002$	0.3908 ± 0.0006	0.4368 ± 0.0039
ETTm2	96	$0.1570^{\dagger} \pm 0.0007$	0.1627 ± 0.0013	0.1781 ± 0.0021
	192	$0.2113^{\dagger} \pm 0.0017$	0.2191 ± 0.0018	0.2421 ± 0.0028
	336	0.2640 ± 0.0025	0.2661 ± 0.0014	0.2879 ± 0.0030
	720	0.3410 ± 0.0054	0.3378 ± 0.0031	0.3753 ± 0.0090
ETTh1	96	$0.3503^{\dagger} \pm 0.0004$	0.3579 ± 0.0010	0.3883 ± 0.0014
	192	$0.3878^{\dagger} \pm 0.0019$	0.3964 ± 0.0043	0.4216 ± 0.0019
	336	$0.4119^{\dagger} \pm 0.0023$	0.4170 ± 0.0015	0.4722 ± 0.0064
	720	$0.4355^{\dagger} \pm 0.0010$	0.4503 ± 0.0159	0.5163 ± 0.0130
ETTh2	96	0.2739 ± 0.0026	0.2755 ± 0.0021	0.3027 ± 0.0019
	192	0.3368 ± 0.0034	$0.3333^{\dagger} \pm 0.0012$	0.3761 ± 0.0026
	336	0.3815 ± 0.0010	$0.3641^{\dagger} \pm 0.0026$	0.4173 ± 0.0062
	720	0.3983 ± 0.0028	$0.3831^{\dagger} \pm 0.0104$	0.4336 ± 0.0056
Solar Energy	96	$0.1537^{\dagger} \pm 0.0025$	0.1647 ± 0.0013	0.1844 ± 0.0041
	192	0.1832 ± 0.0018	0.1864 ± 0.0044	0.2000 ± 0.0025
	336	$0.1925^{\dagger} \pm 0.0023$	0.1950 ± 0.0017	0.2096 ± 0.0016
	720	$0.1959^{\dagger} \pm 0.0010$	0.2005 ± 0.0045	0.2124 ± 0.0014
Traffic	96	0.3484 ± 0.0002	0.3454 ± 0.0008	0.3468 ± 0.0022
	192	0.3697 ± 0.0025	0.3682 ± 0.0025	$0.3649^{\dagger} \pm 0.0028$
	336	0.3853 ± 0.0011	0.3917 ± 0.0010	$0.3835^{\dagger} \pm 0.0018$
	720	0.4310 ± 0.0008	0.4366 ± 0.0038	$0.4182^{\dagger} \pm 0.0031$
Electricity	96	$0.1259^{\dagger} \pm 0.0003$	0.1275 ± 0.0002	0.1480 ± 0.0003
	192	$0.1433^{\dagger} \pm 0.0003$	0.1467 ± 0.0002	0.1644 ± 0.0009
	336	$0.1568^{\dagger} \pm 0.0004$	0.1594 ± 0.0012	0.1785 ± 0.0008
	720	0.1865 ± 0.0002	0.1902 ± 0.0107	0.2093 ± 0.0014
1 st Count		19	6	3
1 st Count with $p < 5\%$		14	4	3
Avg. loss rel. to SFNN		1.000	1.011	1.101

Dataset(s)	ETTh1/ ETTm2	ETTh1/ETTh2/ Traffic/Electricity	Solar/ Weather	ILI	Exchange rate
Period	96	24 or 168	144	52	5
Look-back length L	96, 192, 384, 672, 1344	168, 336, 672, 1344	144, 288, 576, 1008	52, 104, 208	5, 10, 20, 40, 80, 160, 320

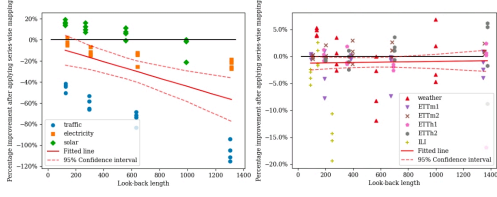
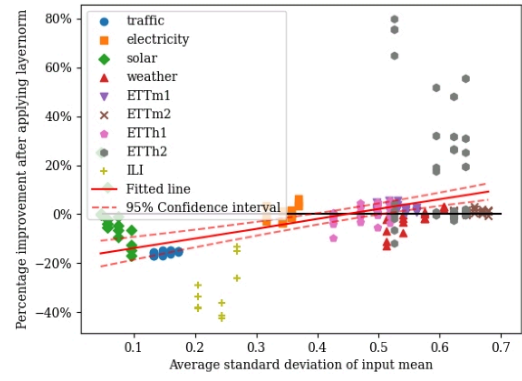
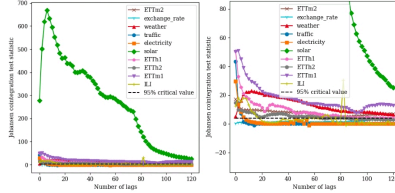


Figure 4: Percentage improvement from incorporating series-wise mapping, along with a linear regression line and its 95% confidence interval. The left panel presents results for datasets with a large number of series, while the right panel shows those for datasets with a small number of series.



Dataset	Horizon	SFNN	DUET [29]	iTransformer [23]
ETTh1	96	$0.2790^{\dagger} \pm 0.0003$	0.2909 ± 0.0008	0.3027 ± 0.0014
	192	$0.3196^{\dagger} \pm 0.0006$	0.3215 ± 0.0006	0.3467 ± 0.0011
	336	0.3514 ± 0.0001	0.3511 ± 0.0007	0.3848 ± 0.0020
	720	$0.3960^{\dagger} \pm 0.0003$	0.4083 ± 0.0007	0.4514 ± 0.0062
ETTh2	96	$0.1570^{\dagger} \pm 0.0007$	0.1657 ± 0.0017	0.1781 ± 0.0021
	192	$0.2139^{\dagger} \pm 0.0019$	0.2192 ± 0.0012	0.2421 ± 0.0028
	336	$0.2642^{\dagger} \pm 0.0012$	0.2661 ± 0.0014	0.2879 ± 0.0030
	720	$0.3493^{\dagger} \pm 0.0015$	0.3563 ± 0.0032	0.3761 ± 0.0042
ETTh1	96	$0.3503^{\dagger} \pm 0.0004$	0.3600 ± 0.0011	0.4042 ± 0.0020
	192	$0.3880^{\dagger} \pm 0.0005$	0.4091 ± 0.0005	0.4489 ± 0.0087
	336	$0.4138^{\dagger} \pm 0.0009$	0.4245 ± 0.0016	0.4722 ± 0.0064
	720	$0.4355^{\dagger} \pm 0.0010$	0.5232 ± 0.0362	0.7331 ± 0.0362
ETTh2	96	$0.2819^{\dagger} \pm 0.0005$	0.2846 ± 0.0022	0.3027 ± 0.0019
	192	$0.3588^{\dagger} \pm 0.0008$	0.3614 ± 0.0013	0.3761 ± 0.0026
	336	0.3945 ± 0.0020	0.3962 ± 0.0040	0.4173 ± 0.0062
	720	$0.4057^{\dagger} \pm 0.0027$	0.4171 ± 0.0030	0.4336 ± 0.0056
Solar Energy	96	0.1758 ± 0.0068	$0.1647^{\dagger} \pm 0.0013$	0.1878 ± 0.0052
	192	$0.1846^{\dagger} \pm 0.0019$	0.1912 ± 0.0016	0.2141 ± 0.0059
	336	$0.1935^{\dagger} \pm 0.0023$	0.2078 ± 0.0013	0.2263 ± 0.0028
	720	$0.1984^{\dagger} \pm 0.0019$	0.2077 ± 0.0015	0.2124 ± 0.0014
Traffic	96	$0.3484^{\dagger} \pm 0.0002$	0.3493 ± 0.0008	0.3566 ± 0.0006
	192	0.3697 ± 0.0025	0.3691 ± 0.0010	0.3752 ± 0.0012
	336	$0.3853^{\dagger} \pm 0.0011$	0.3922 ± 0.0003	0.3891 ± 0.0011
	720	0.4310 ± 0.0008	0.4379 ± 0.0027	$0.4204^{\dagger} \pm 0.0042$
Electricity	96	$0.1261^{\dagger} \pm 0.0003$	0.1311 ± 0.0008	0.1480 ± 0.0003
	192	$0.1441^{\dagger} \pm 0.0002$	0.1467 ± 0.0002	0.1645 ± 0.0010
	336	$0.1586^{\dagger} \pm 0.0003$	0.1655 ± 0.0013	0.1793 ± 0.0012
	720	$0.1895^{\dagger} \pm 0.0007$	0.1936 ± 0.0010	0.2112 ± 0.0052
1 st Count		24	3	1
1 st Count with $p < 5\%$		23	1	1
Avg. loss rel. to SFNN		1.000	1.028	1.117

Dataset	Horizon	N linears	SFNN	DUET [29]	iTransformer [23]
ILI	24	1.922	1.6710 \pm 0.0259	1.6675 \pm 0.0615	1.9931 \pm 0.1910
	36	1.868	1.6815 \pm 0.0367	1.6802 \pm 0.1122	1.8731 \pm 0.1252
	48	1.828	1.7336 \pm 0.0229	1.7960 \pm 0.1707	1.8491 \pm 0.1403
	60	1.859	1.8390 \pm 0.0405	1.7761 \pm 0.1440	2.2390 \pm 0.1729
Weather	96	0.1399	0.1412 \pm 0.0007	0.1494 \pm 0.0018	0.1749 \pm 0.0014
	192	0.1821	0.1874 \pm 0.0010	0.1887 \pm 0.0010	0.2248 \pm 0.0016
	336	0.2315	0.2409 \pm 0.0014	0.2354 \pm 0.0009	0.2816 \pm 0.0011
	720	0.3025	0.3040 \pm 0.0031	0.3075 \pm 0.0022	0.3584 \pm 0.0015
Exchange	96	0.0765	0.0786 \pm 0.0007	0.0791 \pm 0.0008	0.0815 \pm 0.0001
	192	0.1479	0.1503 \pm 0.0003	0.1637 \pm 0.0033	0.1693 \pm 0.0004
	336	0.2402	0.2606 \pm 0.0023	0.2850 \pm 0.0086	0.3093 \pm 0.0005
	720	0.5487	0.5770 \pm 0.0243	0.6593 \pm 0.0340	0.8240 \pm 0.0016

- 대부분 데이터셋에서 SFNN, LSTM/GRU/Transformer와 유사한 RMSE, MAE 기록
- 장기 예측에서는 Transformer가 약간 우세
- SFNN 연산 속도 빠르고 하드웨어 부담 적음
- 다변량 데이터에서도 SFNN 경쟁력 있음 확인

6. Insight

- 단순 Feedforward 모델도 충분히 강력함
- 데이터 특성과 목적에 맞게 모델 선택 필요
- 계산 효율성, 구현 용이성 고려 시 SFNN 활용 가능
- 향후 연구 : 시계열 프리트레이닝, 외생 변수 포함, 단순 모델 성능 향상 기법