Deep Learning Reproducibility

Hojune Shin, Jaewon shin, Sanghoon Han

Department of Data Science, SeoulTech, Korea

1 Introduction

In modern biomedical data analysis, dealing with small sample sizes is often necessary, which can degrade the performance and generalizability of models. To address this issue, the "Weight Predictor Network with Feature Selection (WPFS)" model has been proposed. This model aims to enhance prediction accuracy through feature selection using two auxiliary networks: the Weight Predictor Network (WPN) and the Sparsity Network (SPN).

2 Main Content of the Paper

The WPFS model leverages WPN and SPN to learn the unique patterns in the data. The WPN predicts weights to select important features, while the SPN maintains the sparsity of the selected features to reduce model complexity. The original paper compares the MLP model and the WPFS model, where the MLP model uses data without embedding techniques, while the WPFS model applies embedding techniques such as SVD and Dot Histogram.

3 Experiments Conducted

In this study, we reproduced the experiments from the original paper and evaluated the model's performance under various conditions through the following experiments:

3.1 Comparison of MLP Model and WPFS Model

- **Data Preprocessing**: Min-max scaling preprocessing was applied to the MLP model, and additional experiments without embedding techniques were conducted (including experiments using SVD, NMF, and Dot Histogram).
- **Purpose**: To verify the results of the original paper and evaluate the performance changes with and without embedding techniques.

3.2 Evaluation of Performance with and without Auxiliary Networks

- **Purpose**: To compare the performance of the WPFS model with and without the auxiliary networks (WPN, SPN).
- **Method**: Experiments were conducted to verify if the results were consistent with the original paper and to evaluate the performance changes with and without the auxiliary networks.

3.3 Reproduction of the Paper's Experiments and Hyperparameter Tuning

- **Purpose**: To reproduce the hyperparameter settings of the original paper and evaluate the performance of the WPFS model with various hyperparameter combinations.
- **Method**: The model was trained based on the MLP structure proposed in the original paper, and experiments were conducted by changing the input values to various embedding values (SVD, PCA, Dot Histogram, etc.).

Through these experiments, we aim to analyze the strengths and weaknesses of the WPFS model and evaluate its practical performance on small-scale biomedical data. The importance of reproduction studies lies in verifying the results of the paper and determining how consistently the same methods work on different datasets or conditions. Therefore, this study aims to validate the effectiveness of the WPFS model by reproducing the results of the original paper and identifying potential improvements.

4 Experiment 1

4.1 Method

In this study, we intended to use the GitHub code provided in the paper, but incorporating the preprocessing steps proved complex. Thus, we adopted the MLP structure from the paper and trained a basic model. This allowed us to compare the learning and validation results of the MLP model from the paper and the provided code. Subsequently, experiments were conducted by changing the input values to various embeddings, using feature selection and feature extraction techniques such as SVD, PCA, and Dot Histogram. For SVD, a few features were selected, and the results were analyzed.

4.2 MLP Model Information

The structure of the MLP model used in the paper is as follows:

- **Three Layers**:
  - First layer: 100 neurons
  - Second layer: 100 neurons
  - Third layer: 10 neurons
- Each layer uses the LeakyReLU activation function.
- The output layer uses the softmax activation function to output class probabilities.
- All layers use batch normalization and dropout (p = 0.2) to prevent overfitting.

4.3 Training and Validation

- **Loss Function**: Weighted cross-entropy loss
- **Optimizer**: AdamW
- **Learning Rate**: A scheduler was used to linearly decrease the initial learning rate from 3e-3 to 3e-4.
- **Batch Size**: 8
- **Epochs**: 500
- **Cross-Validation**: 5-fold cross-validation was repeated 5 times, totaling 25 experiments, with 10

4.4 Results and Analysis

In some cases, the accuracy was higher than the results of the original paper, while in others, it was lower. However, given the small number of models and the issue of overfitting with small data, it is more efficient to try various methods rather than strictly following the paper's approach. If the model itself is encrypted for data encryption, and the results cannot be verified, the effectiveness of the model proposed in the paper could be questioned.

4.5 Results Table

4.6 Analysis

The experimental results varied depending on different factors. Some experiments showed performance exceeding the results of the paper, while others did not. These results suggest that the characteristics of the dataset and the hyperparameter settings of the model greatly influence performance.

Dataset	WPFS Accuracy	MLP Accuracy	Highest Accuracy (Method)
CLL	0.723	0.759	MLP
Lung	0.925	0.964	MLP
Prostate	0.903	0.893	WPFS
SMK	0.620	0.645	MLP
Toxicity	0.700	0.800	MLP

Table 1: Comparison of WPFS and MLP accuracy

4.7 Conclusion

This study evaluated the performance of the MLP model compared to the WPFS model proposed in the original paper. The experiments conducted with various embedding techniques and hyperparameter settings showed that the MLP model achieved higher accuracy on some datasets, suggesting that preprocessing can significantly affect model performance regardless of the presence of embedding techniques. Further research with more diverse datasets and model settings is needed.

Through these experiments, we validated the effectiveness of the WPFS model and evaluated its practical performance on small-scale biomedical data. Reproduction studies are crucial for verifying the results of a paper and determining how consistently the same methods work on different datasets or conditions.

5 Experiment 2

5.1 Method

In this study, we conducted experiments in the same manner as the paper. We evaluated the results when each of the two auxiliary networks (WPN, SPN) was absent. This allowed us to assess the performance changes due to the presence of the auxiliary networks and compare the results with those of the paper. The specific experimental methods are as follows:

Aspect	**Details**
Model Structure Setting	We modified the model creation code to control the use of auxiliary networks (`src/model.py` → `create_model`). The parameters of the `FirstLinearLayer` function of the WPFS model were changed to create a new function, and experiments were conducted by removing the WPN network.
Parameter Setting	The feature dimension was set to six different settings: [100, 30], [100, 50], [100, 100], [100, 100, 30], [100, 100, 50], [100, 100, 100].
Experiment Repetition	The final results were obtained by averaging the results of 25 experiments, consistent with the paper.

Table 2: Experiment Details

5.2 Results

The performance of the WPFS model, WPN-removed model, and SPN-removed model for each feature dimension setting is as follows:

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.7809	0.7661	0.7395
lung	0.9606	0.9911	0.9642
prostate	0.8927	0.8918	0.8155
smk	0.6111	0.6778	0.6100
toxicity	0.6922	0.8813	0.6996

Table 3: Comparison of models with feature dimension [100, 30]

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.7138	0.7535	0.7914
lung	0.9588	0.9777	0.9677
prostate	0.8927	0.8564	0.8636
smk	0.6280	0.6403	0.5704
toxicity	0.7083	0.8563	0.7246

Table 4: Comparison of models with feature dimension [100, 50]

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.7481	0.7341	0.7735
lung	0.9464	0.9911	0.9464
prostate	0.8718	0.8927	0.8218
smk	0.6226	0.6869	0.5115
toxicity	0.7033	0.8757	0.7521

Table 5: Comparison of models with feature dimension [100, 100]

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.6465	0.7328	0.7663
lung	0.9517	0.9499	0.9286
prostate	0.8164	0.9152	0.9027
smk	0.5742	0.6224	0.5958
toxicity	0.7081	0.8274	0.6822

Table 6: Comparison of models with feature dimension [100, 100, 30]

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.7347	0.7817	0.6894
lung	0.9267	0.9642	0.9688
prostate	0.8827	0.8827	0.8827
smk	0.5671	0.6905	0.4819
toxicity	0.6799	0.8396	0.7038

Table 7: Comparison of models with feature dimension [100, 100, 50]

Dataset	WPFS	rm_WPN	rm_SPN
cll	0.7687	0.7434	0.7187
lung	0.9571	0.9392	0.9777
prostate	0.8536	0.8627	0.8064
smk	0.5854	0.5455	0.5764
toxicity	0.7061	0.8135	0.7028

Table 8: Comparison of models with feature dimension [100, 100, 100]

5.3 Analysis

- The WPFS model proposed in the paper did not always achieve the highest accuracy.
- The model with the best performance varied by dataset, and consistency was not observed.

5.4 Future Research Directions

- A more detailed analysis of the research proposed in the paper is needed, and improvements in code and performance should be pursued.
- The best model generated automatically during code execution should be analyzed to derive optimal parameter settings.

5.5 Conclusion

This experiment re-evaluated the performance of the WPFS model and assessed the impact of the auxiliary networks. It was confirmed that the presence of auxiliary networks could significantly influence model performance, highlighting the need for further research and analysis to clarify these effects.

6 Experiment 3

6.1 Reproduction of the Paper's Experiments and Hyperparameter Tuning

6.2 Existing Environment Settings

The existing parameter settings of the WPFS model are as follows. This setup comprises a 3-layer feedforward neural network, auxiliary WPN, and SPN networks, with parameters and training settings for each network.

- **Classification Network**:
  - Structure: 3-layer feedforward neural network
  - Neurons per layer: 100, 100, 10
  - Output activation function: Softmax
  - Internal activation function: LeakyReLU
  - Dropout: 0.2

  - Batch normalization: Used

- **Weight Predictor Network (WPN)**:
  - Structure: 4-layer feedforward neural network
  - Neurons per layer: 100
  - Final layer activation function: Tanh
  - Internal activation function: LeakyReLU
  - Dropout: 0.2
  - Batch normalization: Used
- **Sparsity Network (SPN)**:
  - Structure: 4-layer feedforward neural network
  - Neurons per layer: 100
  - Final layer activation function: Sigmoid
  - Internal activation function: LeakyReLU
  - Dropout: 0.2
  - Batch normalization: Used

Training Settings:

- Batch size: 8
- Optimization method: AdamW
- Learning rate: 3e-3
- Weight decay: 1e-4
- Learning rate scheduler: Linear decay (from 3e-3 to 3e-4 over 500 epochs)
- Total epochs: 500

6.3 Experimental Results

Experiments were conducted using the code and data provided on the paper's GitHub. The results for each model with the hyperparameters provided in the paper are as follows:

Dataset	WPFS Accuracy	DietNets	FsNe	CAE	MLP
CLL	0.723367003	0.748215488	0.571986532	0.590707	0.758855219
Lung	0.924933862	0.923148	0.676124339	0.687089947	0.964087302
Prostate	0.902727273	0.826363636	0.843636364	0.773636364	0.892727273
SMK	0.619736842	0.558392	0.534385965	0.588216	0.644678363
Toxicity	0.700297619	0.677380952	0.473015873	0.530952	0.799801587

Table 9: Comparison of WPFS and other models' accuracy

In this experiment, the WPFS model had the lowest average rank, but in our study, the MLP model had the lowest average rank.

6.4 WPFS Hyperparameter Tuning

Experiments were conducted with various hyperparameter settings for batch size, learning rate, dropout rate, and weight decay. The settings with the highest accuracy for each dataset are as follows:

Experiments with various hyperparameters for the model proposed in the paper showed improved performance compared to the results obtained with the code provided in the paper. Especially for the Lung dataset, higher accuracy was observed.

Dataset	ACC	Batch Size	Learning Rate	Dropout Rate	Weight Decay
CLL	0.794276094	16	0.003	0.3	0.0001
Lung	0.978439153	16	0.003	0.3	1.00E-05
Prostate	0.893636364	16	0.0003	0.1	1.00E-05
SMK	0.655760234	16	0.03	0.3	0
Toxicity	0.832638889	16	0.003	0.1	0.0001

Table 10: Hyperparameter settings with highest accuracy for each dataset

6.5 Analysis

- The WPFS model proposed in the paper did not always achieve the highest accuracy.
- The model with the best performance varied by dataset, and consistency was not observed.
- It was confirmed that performance could be optimized through various hyperparameter settings.

6.6 Future Research Directions

- A more detailed analysis of the research proposed in the paper is needed, and improvements in code and performance should be pursued.
- The best model generated automatically during code execution should be analyzed to derive optimal parameter settings.
- Additional datasets should be used to evaluate the generalizability of the model and verify performance in various scenarios.

This study focused on re-evaluating the performance of the WPFS model and optimizing performance through various hyperparameter settings. Through this, the effectiveness of the WPFS model was validated, and its practical performance on small-scale biomedical data was evaluated. Reproduction studies are crucial for verifying the results of a paper and determining how consistently the same methods work on different datasets or conditions.