Psychological Health Risk Prediction for Medical Interns Based on Multi-scale Deep Learning

Junqing Lin

Southwest Medical University, Luzhou, 646000, China

1529002456@qq.com

Abstract. Deep learning has become a key methodology in mental health development, especially for medical interns facing intense pressures. Early identification and individualized interventions are essential for addressing their mental health concerns. This paper presents a context-aware multi-scale deep learning framework for continuous psychological state modeling and risk prediction of medical interns, providing both theoretical and technical support for mental health management in medical education. The model combines a dynamic multi-scale convolutional feature extractor, a hierarchical recurrent neural network, and a context-aware meta-learning module, which collectively track interns' psychological changes over time. By bridging theory and practical exploration, the paper introduces an innovative approach to predicting mental health risks and contributing to intelligent interventions in medical education. The results show that the proposed model enhances psychological self-regulation in interns, strengthens the ability of educational management to monitor mental health dynamics, and transitions mental health services from a "passive response" to an "active warning" model, thus supporting the development of a student-centered psychological support system.

Keywords: Deep Learning, Medical Interns, Mental Health.

1 Introduction

According to the World Health Organization (WHO)[1], health is not merely the absence of disease or infirmity but a state of complete physical, mental, and social well-being, with the 21st century being referred to as the "century of great health." Guided by this concept, mental health, as an integral part of overall health, has gained increasing attention from various sectors of society. This is particularly true in the field of medical education, where the mental health of medical interns has become a growing concern, significantly influencing both their career development and the quality of care they provide.

Numerous studies have demonstrated that medical interns typically face multiple challenges during their internship, including academic pressure, clinical workload, interpersonal conflicts, and career uncertainties. As a result, the prevalence of anxiety, depression, and excessive stress among medical interns is notably higher compared to

H. Kannan et al. (eds.), *Proceedings of the 5th International Conference on Internet, Education and Information Technology (IEIT 2025)*, Atlantis Highlights in Social Sciences, Education and Humanities 40,
https://doi.org/10.2991/978-94-6463-803-5_68

their peers. These psychological difficulties not only impair cognitive function and the quality of life of interns but may also indirectly compromise patient safety and healthcare services[2].

Currently, mental health assessments in medical education primarily rely on standardized questionnaires and self-rating scales, such as the PHQ-9, GAD-7, and other tools[3]. While these tools offer valuable insights in initial screenings and clinical interventions, they exhibit limitations such as high subjectivity, poor timeliness, and insufficient adaptability to individual differences. These issues hinder their effectiveness in the dynamic and high-pressure environment of internships[4].

With the advancement of artificial intelligence, particularly deep learning technologies, researchers have begun exploring the development of automated, dynamic mental health prediction models. These models analyze multimodal data, such as text, speech, and physiological sensor signals, to more accurately assess individual psychological states[5]. This trend not only enhances the efficiency of psychological risk detection but also offers the potential for personalized interventions tailored to the needs of medical education groups, making it a promising frontier in mental health research.

In this study, we propose a multi-scale deep learning-based framework for mental health risk prediction among medical interns. The framework integrates multi-level semantic information and contextual features, aiming to improve the sensitivity and prediction accuracy of the model in detecting fluctuations in psychological states. The goal is to provide a more effective early warning system for the mental health of medical interns and offer technical support for universities and hospitals in developing personalized intervention strategies.

2 Theoretical Connotation of Multi-scale Deep Learning

Multi-scale deep learning refers to a class of neural network models designed to capture features across different levels of abstraction and spatial-temporal resolutions. Inspired by the hierarchical structure of the human perception system, multi-scale architectures differ from traditional single-scale models by integrating features extracted from various receptive fields. This enables the simultaneous focus on global patterns and local details, thereby enhancing the model's ability to model complex behaviors and psychological states.

In recent years, multi-scale convolutional neural networks (Multi-Scale CNN) and hierarchical recurrent networks (such as multi-layer LSTM, Bi-LSTM) have demonstrated exceptional performance in fields like emotion recognition, EEG signal analysis, and sentiment text modeling[6]. These models typically configure different convolutional kernel sizes or time windows in parallel or cascaded ways, effectively extracting multi-time scale features such as immediate psychological states (e.g., short-term stress responses) and long-term emotional fluctuations (e.g., chronic mood variations)[7].

In the context of mental health risk prediction, this multi-scale modeling approach holds significant importance. An individual's psychological state is highly dynamic and multidimensional, manifesting not only as short-term symptom fluctuations but also as long-term behavioral trajectories and emotional evolutions[8]. Multi-scale deep learning models can integrate short-term physiological signals or textual data with long-term behavioral records, enabling dynamic modeling and individualized prediction of psychological states, thus enhancing the sensitivity and stability of predictions.

Moreover, when multi-scale architectures are combined with attention mechanisms or graph neural networks, the model's feature selection, structural adaptability, and interpretability are significantly enhanced. This characteristic makes it particularly valuable in mental health monitoring and intervention scenarios, increasing its clinical usability and educational applicability[9]. Therefore, multi-scale deep learning not only aligns theoretically with the complexity of psychological modeling but also provides a robust technical foundation and research pathway for precise psychological interventions in medical education.

3 Methodology and Experimental Approach

To address the current limitations in monitoring the dynamic mental health status of medical interns—such as insufficient real-time tracking, delayed interventions, and a lack of personalized assessment—this study proposes a context-aware Multiscale Deep Neural Network (MscaleDNN) framework. The model integrates multiscale feature extraction, dynamic temporal modeling, and context-aware personalization mechanisms to enhance the timeliness, sensitivity, and adaptability of mental health risk prediction during clinical internships.

3.1 Research Participants and Model Data Collection

This study recruited 300 medical interns from 3 tertiary hospitals in Sichuan Province as participants. Data collection spanned the entire clinical internship period and utilized multiple modalities, including: 1) Psychometric assessments (e.g., PHQ-9, GAD-7), 2) Physiological indicators (e.g., heart rate variability), 3) Behavioral metrics (e.g., sleep duration, smartphone usage time). And all participants provided written informed consent in accordance with ethical guidelines.

All participants provided written informed consent in accordance with ethical guidelines. Preprocessing steps included normalization of numerical features, imputation of missing values, and segmentation of time-series data into daily windows to ensure consistency. Categorical variables (e.g., clinical department rotations) were transformed using one-hot encoding. Time-series alignment was performed using fixed daily windows to maintain uniformity across participants. Missing values (approximately 5.2%) were imputed using bidirectional GRU-based reconstruction models trained on complete sequences. The dataset was temporally split into training (first 70%), validation (next 15%), and testing sets (final 15%).

3.2 Model Architecture

The proposed MscaleDNN consists of three core components. First, the Dynamic Multiscale Convolutional Feature Extractor employs a combination of multi-kernel convolution and dilated convolution to capture both short-term fluctuations and long-term trends in psychological state dynamics across multiple temporal scales. Next, the Hierarchical Recurrent Temporal Modeler utilizes a two-layer Gated Recurrent Unit (GRU) architecture to model intra-week and inter-week changes in mental health status, thereby enhancing temporal representation and dependency learning[10]. Finally, the Context-Aware Meta-Learning Module integrates personalized contextual information—such as rotation workload and prior mental state history—through an attention-weighted mechanism. This design guides adaptive learning and improves both the model's generalization across individuals and its accuracy in identifying mental health risks.

3.3 Experimental Parameters

All models were implemented using PyTorch and trained on an NVIDIA V100 GPU with mixed-precision acceleration. The MscaleDNN model was trained using the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with an initial learning rate of 0.001, which was decayed by a factor of 0.5 every 10 epochs. Training was performed with a batch size of 32 over a maximum of 100 epochs. Early stopping was applied based on the validation loss with a patience of 5 epochs to prevent overfitting. The binary cross-entropy loss function was used as the training objective. To ensure training stability and avoid overfitting, L2 weight decay ($\lambda = 1e{-}4$) and gradient clipping with a maximum norm of 5 were applied. The entire training process took approximately 10 hours. For individual-level personalization, the model was fine-tuned using each intern's historical data, with an average adaptation time of 3-4 minutes per person.

3.4 Model Evaluation and Baseline Comparison

To comprehensively evaluate the performance of the proposed model, we employed a set of standard evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC). In addition, we used Cohen's κ to measure consistency between model predictions and clinical evaluations, the Area Under the Precision-Recall Curve (AUPRC) for early detection capability, Floating Point Operations (FLOP) per inference to assess computational efficiency, and Personalization Gain, which reflects the improvement in F1 score achieved through personalized fine-tuning compared to group-level modeling.

Furthermore, to validate the effectiveness of MscaleDNN, we conducted comparative analyses against five representative baseline models. The first is the Static Fixed-Scale CNN, a conventional convolutional neural network that utilizes three fixed kernel sizes (3, 5, and 7) with equal path weighting, serving as a benchmark for static multiscale feature extraction. The second model, LSTM-TD, is a Long Short-Term Memory network enhanced with a temporal attention mechanism, designed to capture longrange temporal dependencies and representing the current state-of-the-art in time series analysis for mental health. Third, we included MAML-Base, a model-agnostic meta-learning framework that is independent of specific network architectures and lacks multiscale capabilities, providing a baseline for evaluating meta-learning performance. Fourth, the Psychological Scales approach applies logistic regression to standardized clinical assessment tools such as PHQ-9 and GAD-7, representing a traditional psychoinformatic method. Lastly, the Early Fusion model integrates multimodal data at an early stage using standard GRU architectures, aiming to enhance predictive performance through early-stage data fusion. To ensure fair comparison, all baseline models were implemented using their original architectures, optimized via grid search on the validation set, and adjusted to match MscaleDNN's total parameter count (approximately 1.2 million).

By comparing MscaleDNN with these established methods, we demonstrate the empirical advantages of MscaleDNN and highlight its potential for further improvements in personalized mental health risk modeling.

4 Experiments and Results

To comprehensively assess the performance of MscaleDNN, we conducted comparative evaluations, ablation studies, and feature activation analyses.

4.1 Overall Performance Comparison

Table 1 presents a detailed comparison between MscaleDNN and several baseline models. MscaleDNN achieves a Cohen's Kappa coefficient of 0.725, outperforming the strongest baseline, LSTM-TD ($\kappa = 0.628$), by approximately 15.4%. In early detection, it reaches an AUPRC of 0.803, exceeding MAML-Base (AUPRC = 0.672) by 19.5%. Moreover, it is more computationally efficient, with a FLOPs count of 69.5×10^6, about 27.8% lower than LSTM-TD. Regarding personalization, MscaleDNN demonstrates a 28.7% performance improvement after adaptation—significantly higher than MAML-Base (18.9%), underscoring its robust subject-specific generalization.

Table 1. Performance comparison between MscaleDNN and baseline methods.

Method	Clinical Accuracy (κ)	Early Detection (AUPRC)	FLOPs ($\times10^6$)	Personalization Gain
Fixed-Scale CNN	0.576 ± 0.03	0.617 ± 0.04	82.1	+8.0%
LSTM-TD	0.608 ± 0.02	0.668 ± 0.03	96.3	+12.5%
MAML-Base	0.602 ± 0.04	0.672 ± 0.05	87.0	+18.9%
Psych Scale	0.531 ± 0.05	0.547 ± 0.06	2.6	+3.5%
Early Fusion	0.598 ± 0.03	0.635 ± 0.04	77.2	+9.5%
Mscale DNN	0.725 ± 0.02	0.803 ± 0.03	69.5	+28.7%

4.2 Multi-Scale Feature Analysis

To further validate the multi-scale feature extraction capability of the MscaleDNN model, we analyzed the dynamic convolutional activation patterns across different mental health indicators, as illustrated in Figure 1. The figure presents the activation responses of various psychological signals—such as electrodermal activity (EDA), heart rate variability (HRV), and sleep patterns—across multiple temporal scales. For high-frequency physiological signals such as EDA, MscaleDNN tends to favor smaller receptive fields (e.g., 3-5 samples), capturing rapid fluctuations effectively. In contrast, for slowly varying indicators like sleep patterns, the model exhibits a preference for larger receptive fields (e.g., 5-7 samples), enabling the extraction of long-term contextual information.

This adaptive feature extraction mechanism allows MscaleDNN to achieve superior performance across diverse signal types by dynamically adjusting to the temporal characteristics of each input modality.

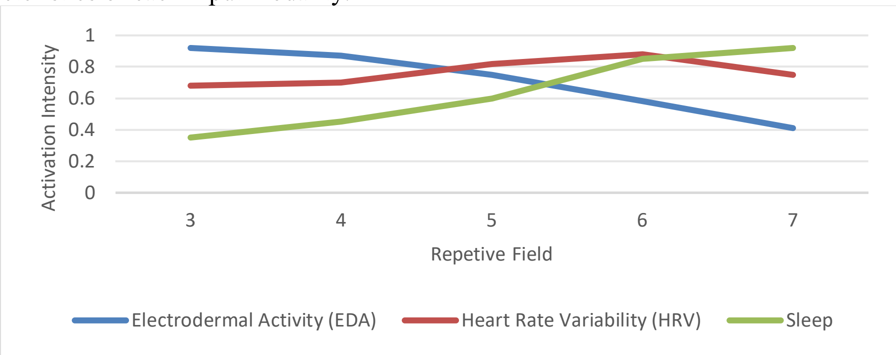

Fig. 1. Multi-scale convolutional activation trends for psychological signals.

In addition, MscaleDNN's hierarchical temporal modeling mechanism demonstrates strong effectiveness in capturing critical transitions between mental health states. By integrating both short-term and long-term GRU modules, the model is capable of accurately detecting acute stress events that evolve into progressive burnout patterns. This capability enables MscaleDNN to exhibit notable advantages in both early detection and continuous monitoring of mental health trajectories.

4.3 Personalized Modeling Capability

To evaluate the model's capability for personalized modeling, we employ a meta-learning module to adaptively fine-tune model parameters for individual interns, thereby optimizing performance on subject-specific prediction tasks. Figure 2 illustrates the parameter adaptation trajectories during the meta-learning process for four interns with distinct psychological profiles. For the intern exhibiting chronically low anxiety, the model prioritizes physiological stability features during adaptation. In contrast, for the intern with intermittent acute stress episodes, the model places greater emphasis on features related to workload volatility. These results highlight MscaleDNN's ability to dynamically tailor its representation and prediction strategies based on individual mental health patterns.

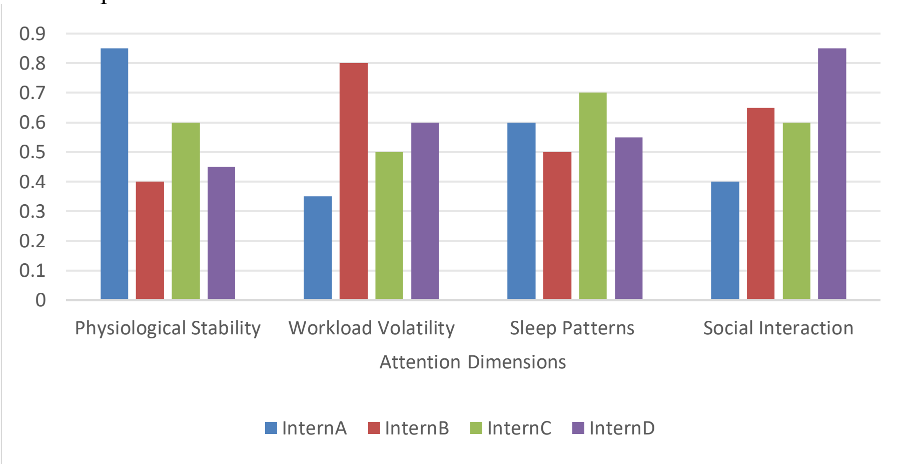

Fig. 2. Visualization of parameter adaptation trajectories for medical interns.

This suggests that MscaleDNN demonstrates strong adaptability in personalized modeling. It is capable of dynamically adjusting its model structure and parameters based on the varying mental health states of the interns, thereby improving its diagnostic accuracy.

4.4 Ablation Study Analysis

To further assess the contribution of each component within the MscaleDNN framework, an ablation study was conducted, with the results presented in Table 2. Removing the dynamic scaling module led to a 9.7% decrease in clinical accuracy. Additionally, the removal of the hierarchical RNN and meta-learning module resulted in varying degrees of impairment in early detection capability. These ablation experiments underscore the critical importance of the dynamic scaling mechanism, hierarchical RNN, and meta-learning modules, which are indispensable components of the MscaleDNN framework.

Table 2. Ablation study of MscaleDNN components

Configuration	Clinical Accuracy (κ)	Early Detection(AUPRC)	FLOP($\times 10^6$)
Full-scale DNN	0.72 ± 0.02	0.81 ± 0.03	68.9
Without Dynamic Scaling	0.65 ± 0.03	0.73 ± 0.04	82.4
Without Hierarchical RNN	0.68 ± 0.02	0.76 ± 0.03	71.2
Without Meta-learning	0.66 ± 0.03	0.74 ± 0.04	69.8
Single-scale Static	0.61 ± 0.04	0.69 ± 0.05	84.1

The ablation results further emphasize the significant contributions of the multi-scale feature extraction, hierarchical temporal modeling, and meta-learning modules to the model's performance. The integration of these modules allows MscaleDNN to more effectively adapt to individualized needs and enhance early warning capabilities when handling complex mental health data.

5 Conclusion

Addressing the growing need for psychological support in medical education, this study aims to explore an effective method for the early identification of mental health risks in medical interns. To achieve this objective, we propose the MscaleDNN framework, a multi-scale deep learning model that integrates a dynamic multi-scale convolutional feature extractor, a hierarchical recurrent neural network (RNN) temporal modeler, and a context-aware meta-learning module. These components work together to enable continuous modeling and individualized risk identification of interns' psychological states.

An empirical study involving 300 medical interns from 3 tertiary hospitals in Sichuan Province was conducted, and the results demonstrate that MscaleDNN offers high accuracy in detecting psychological state changes and predicting individual risks. Compared to traditional models, the framework shows clear advantages in time series modeling and capturing individual differences.

Future research could expand in several directions: increasing the sample size, incorporating multi-center and multimodal data (e.g., speech, facial expressions), and developing real-time feedback and intervention systems based on model predictions. Additionally, integrating this psychological risk prediction framework into the medical internship and education system could provide more timely and accurate psychological support for interns working in high-pressure environments.

This study presents a novel, practical, and intelligent approach to personalized risk prediction for the mental health of medical interns, contributing to the development of psychological early warning systems in high-pressure clinical education settings.

References

1. World Health Organization: Official records of the World Health Organization no. 2. Summary report on proceedings, minutes and final acts of the International Health Conference held in New York from 19 June to 22 July 1946. United Nations, Geneva (1948).
2. Abdulghani, H.M., Irshad, M., Al Zunitan, M.A., Al Sulihem, A.A., Al Dehaim, M.A., Al Esefir, W.A., Al Rabiah, A.M., Kameshki, R.N., Alrowais, N.A., Sebiany, A., Haque, S.: Prevalence of stress in junior doctors during their internship training: a cross-sectional study of three Saudi medical colleges' hospitals. Neuropsychiatric Disease and Treatment 10, 1879-1886 (2014).
3. Kroenke, K., Spitzer, R.L., Williams, J.B.W.: The PHQ-9: validity of a brief depression severity measure. Journal of General Internal Medicine 16(9), 606-613 (2001).
4. Zimmerman, M.: The value and limitations of self-administered questionnaires in clinical practice and epidemiological studies. World Psychiatry 23(2), 210-212 (2024).

5. Muetunda, F., Sabry, S., Jamil, M.L., Pais, S., Dias, G., Cordeiro, J.: AI-Assisted Diagnosing, Monitoring and Treatment of Mental Disorders: A Survey. ACM Transactions on Computing for Healthcare 5(4), Article 23 (2024).
6. Tripathi, S., Acharya, S., Sharma, R.D., Mittal, S., Bhattacharya, S.: Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: Proceedings of the AAAI Conference on Artificial Intelligence 31(1), 47464-752 (2017).
7. Chao, H., Dong, L.: Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals. IEEE Sensors Journal 21(2), 2024-2034 (2021).
8. Ding, Y., Robinson, N., Zeng, Q., Chen, D., Chai, A.A.P., Lee, T.S., Guan, C.: TSception: a deep learning framework for emotion detection using EEG. IEEE Transactions on Neural Systems and Rehabilitation Engineering 28(1), 195-204 (2020).
9. Cui, F., Wang, R., Ding, W., Chen, Y., Huang, L.: A Novel DE-CNN-BiLSTM Multi-Fusion Model for EEG Emotion Recognition. Mathematics 10(4), 582 (2022).
10. Teng, L., Li, H., Karim, S.: DMCNN: A Deep Multiscale Convolutional Neural Network Model for Medical Image Segmentation. Journal of Healthcare Engineering 2019, Article ID 8597606 (2019).