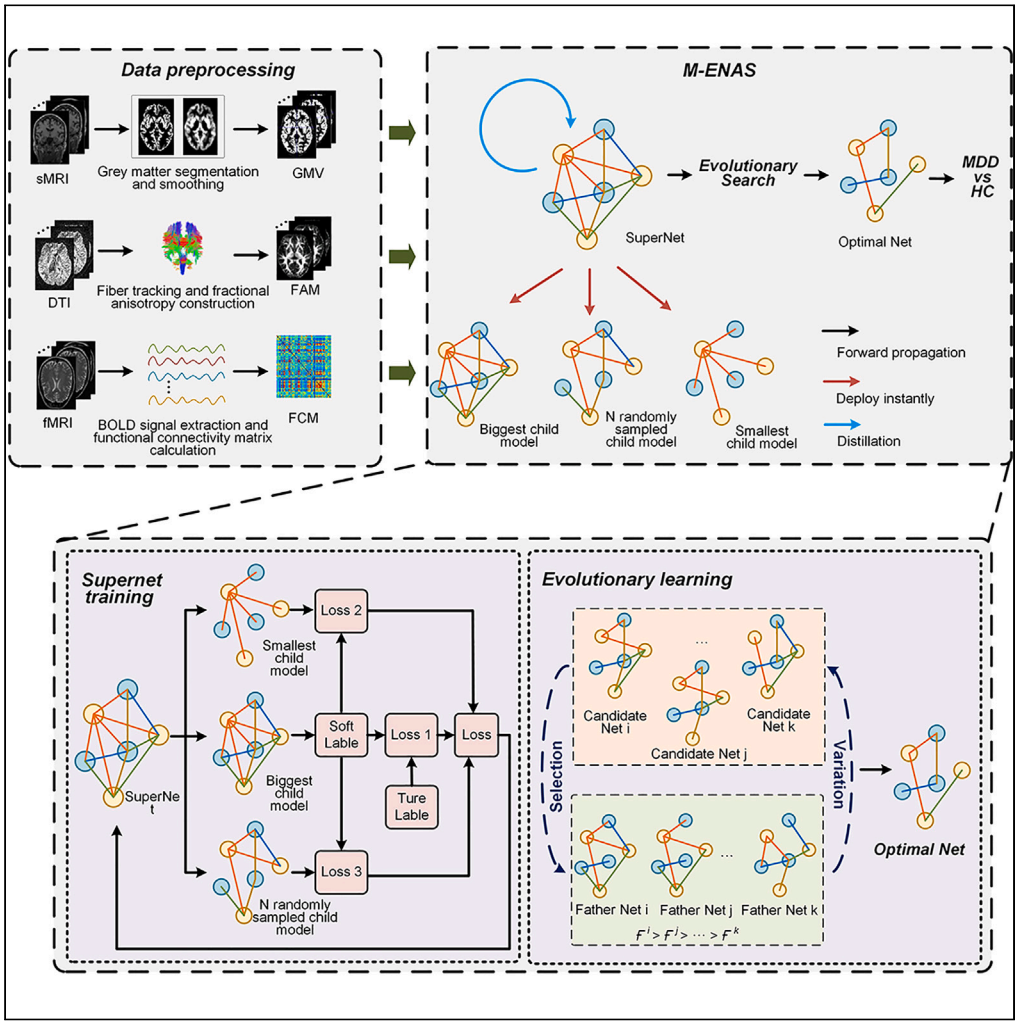
iScience

CellPress
OPEN ACCESS

Article

Evolutionary neural architecture search for automated MDD diagnosis using multimodal MRI imaging



Tongtong Li, Ning Hou, Jiandong Yu, ..., Sujie Ma, Jiansong Zhou, Bin Hu

yaozj@lzu.edu.cn (Z.Y.)
18993803505@163.com (S.M.)
zhoujs2003@csu.edu.cn (J.Z.)
bh@lzu.edu.cn (B.H.)

Highlights

We proposed an M-ENAS framework for the automatic diagnosis of MDD

M-ENAS method effectively extracts and fuses multimodal neuroimaging features

Applications on two datasets unanimously demonstrated the effectiveness of M-ENAS

Our findings show that somatomotor network regions are crucial in MDD diagnosis

Li et al., iScience 27, 111020
October 18, 2024
https://doi.org/10.1016/j.isci.2024.111020

iScience

CellPress
OPEN ACCESS

Article

Evolutionary neural architecture search for automated MDD diagnosis using multimodal MRI imaging

Tongtong Li,[1,2,9] Ning Hou,[3,9] Jiandong Yu,[1,2] Ziyang Zhao,[1,2] Qi Sun,[1,2] Miao Chen,[1,2] Zhijun Yao,[1,2,*] Sujie Ma,[4,*] Jiansong Zhou,[5,*] and Bin Hu[1,2,6,7,8,10,*]

SUMMARY

Major depressive disorder (MDD) is a prevalent mental disorder with serious impacts on life and health. Neuroimaging offers valuable diagnostic insights. However, traditional computer-aided diagnosis methods are limited by reliance on researchers' experience. To address this, we proposed an evolutionary neural architecture search (M-ENAS) framework for automatically diagnosing MDD using multi-modal magnetic resonance imaging (MRI). M-ENAS determines the optimal weight and network architecture through a two-stage search method. Specifically, we designed a one-shot network architecture search (NAS) strategy to train supernet weights and a self-defined evolutionary search to optimize the network structure. Finally, M-ENAS was evaluated on two datasets, demonstrating that M-ENAS outperforms existing hand-designed methods. Additionally, our findings reveal that brain regions within the somatomotor network play important roles in the diagnosis of MDD, providing additional insight into the biological mechanisms underlying the disorder.

INTRODUCTION

Major depressive disorder (MDD) is a prevalent mental disorder characterized by persistent feelings of sadness and despair, which can lead to profound self-doubt and suicidal thoughts or behavior.[1] As reported by the World Health Organization (WHO), around 350 million individuals worldwide suffer from depression, and more than 1 million deaths by suicide are attributed to MDD each year.[2] In recent years, global pandemics and escalating societal pressures have resulted in a significant increase in the number of patients with MDD, imposing a substantial burden on healthcare systems.[3,4] However, traditional diagnostic methods for depression rely heavily on clinical interviews and questionnaire assessments, which are susceptible to subjective biases and demand considerable time from healthcare professionals, thereby exacerbating the social healthcare burden.[5] Therefore, developing an efficient and reliable diagnostic model for the automatic diagnosis of MDD has become an urgent task.

In recent decades, rapid advancements in neuroimaging technologies have provided evidence for structural and functional changes in the brain in a non-invasive manner, facilitating the assessment of physiological changes. Structural magnetic resonance imaging (sMRI) and diffusion tensor imaging (DTI) imaging depict macroscopic and microscopic states of the brain structure, respectively.[6–8] The resting-state functional magnetic resonance imaging (rs-fMRI) captures the dynamic processes of brain activity by recording the blood-oxygen-level-dependent (BOLD) signals.[9] The development of various imaging technologies has provided a more comprehensive understanding of the structure and function of the brain. This advancement not only stimulates further exploration in neuroscience but also opens additional avenues for clinical diagnosis and treatment.[10,11]

Each neuroimaging modality provides a distinct perspective and interpretation of abnormal changes in the brain. Previous studies[12–14] have demonstrated that integrating complementary information from different neuroimaging modalities can improve the performance of diagnostic models for disease detection. However, the design of the traditional multi-modal deep learning models often relies on the researchers' experience and insight, which does not guarantee an optimal solution and is time-consuming.

[1]School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China
[2]Gansu Provincial Key Laboratory of Wearable Computing, Lanzhou University, Lanzhou 730000, China
[3]Medical Department, The Third People's Hospital of Tianshui, Tianshui 741000, China
[4]Sleep Department, The Third People's Hospital of Tianshui, Tianshui 741000, China
[5]National Clinical Research Center for Mental Disorders, and National Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha 410000, China
[6]School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China
[7]CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[8]Joint Research Center for Cognitive Neurosensor Technology of Lanzhou University & Institute of Semiconductors, Chinese Academy of Sciences, Lanzhou 730000, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: yaozj@lzu.edu.cn (Z.Y.), 18993803505@163.com (S.M.), zhoujs2003@csu.edu.cn (J.Z.), bh@lzu.edu.cn (B.H.)
https://doi.org/10.1016/j.isci.2024.111020

Table 1. Comparison of different inputs in the M-ENAS

Data modalities	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1-score*	*Parameter*(M)
sMRI	66.67%	60.00%	81.82%	53.85%	0.6923	2.96
fMRI	62.50%	62.50%	45.46%	76.92%	0.5263	2.96
DTI	66.87%	71.43%	45.46%	84.62%	0.5556	2.96
sMRI+fMRI	66.67%	60.00%	81.82%	53.85%	0.6923	4.47
sMRI+DTI	70.83%	64.29%	81.82%	61.54%	0.7200	4.47
fMRI+DTI	66.67%	61.54%	72.73%	61.54%	0.6667	4.47
sMRI+fMRI+DTI	75.00%	66.67%	90.91%	61.54%	0.7692	5.61

Network architecture search (NAS) is a method for automatically exploring the optimal network architecture using a search strategy within a pre-defined search space.[15,16] It allows a comprehensive exploration of the network architecture space to find the efficient model suitable for a specific task. Moreover, the NAS method has been applied to multi-modal tasks. Xiao et al.[14] designed an automatic search network architecture "DLS-DARTS" model to extract and fuse multi-modal information for glioma grading. Liu et al.[17] developed a multi-task NAS model to mine complementary information in multi-modal MRI images for tumor segmentation and identification. Dai et al.[18] employed the AdaNet-based NAS model to extract discriminative features for fMRI signal classification.

Although the aforementioned studies have shown that NAS algorithms hold significant promise for the extraction and integration of multi-modal features, existing NAS methods for neuroimaging are still in their infancy due to the complexity of the search space[19–21] and the scarcity of data.[22,23] Under the circumstances, we proposed an evolutionary neural architecture search (M-ENAS) method for the automatic diagnosis of MDD. Specifically, the M-ENAS architecture determines the optimal weights and network architecture through a two-stage NAS approach. We trained the weights of the supernet through a one-shot strategy and designed an evolutionary learning search strategy to explore the optimal network architecture. For the detection of MDD, multi-modal features were further integrated and identified through a feature concatenation module and fully connected layers (FC) with a softmax activation function, respectively. Finally, the M-ENAS was evaluated on two separate datasets, demonstrating that it outperforms existing hand-designed deep learning models. In addition, we investigated the regions of interest selected by M-ENAS using occlusion analysis and found that the brain regions within the somatomotor network (SMN) play important roles in the diagnosis of MDD. However, we must acknowledge that this method still presents practical implementation challenges in hospital settings due to lower accuracy compared to clinicians and requires high-performance computational resources.

The contributions of the study are as follows.

(1) We developed an evolutionary neural architecture search framework named M-ENAS for the automatic diagnosis of MDD.
(2) We designed a two-stage architecture search strategy to explore the optimal architecture.
(3) We evaluated the performance of the M-ENAS on two datasets, achieving competitive performance compared to hand-designed deep learning models.
(4) Our findings highlight that brain regions within the SMN play critical roles in MDD diagnosis.

RESULTS

Experimental results

In this section, we comprehensively evaluated the performance of the M-ENAS through comparative experiments. The reported results are the mean values of stratified 5-fold cross-validation. Moreover, we provided the number of parameters of different models to assess the effectiveness of the M-ENAS.

Table 1 reports the comparison of different inputs in the M-ENAS. The optimal performance (accuracy = 75.00%, F1-score = 0.7692) was achieved when all three modalities were used simultaneously as inputs. The fusion of two modalities improved the model's performance (accuracy $\geq$ 66%), indicating that multi-modal information fusion enhances the diagnostic performance by leveraging the complementary information between modalities.

In Table 2, we compared M-ENAS with existing hand-designed state-of-the-art methods on two different datasets, including site-specific dataset and an open-access multi-site dataset, to validate the generalization of the M-ENAS, and achieved competitive performance on both two datasets.

Table 3 reports the comparison of the different encodings between classical hand-designed deep learning models (AlexNet,[27] VGGNet,[28] ResNet,[29] GoogLeNet,[30] DenseNet,[31] Vision Transformer [VIT],[32] Swin-Transformer,[33] and Repvit[34]) and the M-ENAS. Experimental results showed that the M-ENAS method has better potential by exploring the optimal architectures within a pre-defined search space (see STAR Methods). To estimate the robustness of the M-ENAS, we further compared it with mainstream NAS models (Darts[35] and FB-Net[36]), and the results were shown in Table 4. Finally, comparing with different subnet search strategies (random search and coarse-to-fine[37]), the effectiveness of the self-designed evolutionary search strategy was demonstrated (see Table 5).

In addition, Figures 1 and 2 present the confusion matrix of M-ENAS and the receiver operating characteristic (ROC) curves for comparisons of M-ENAS with different inputs, respectively. The confusion matrices for all algorithms are provided in Figures S4–S7.

Table 2. Comparison of the M-ENAS with existing state-of-the-art methods on two different datasets

Dataset	Models	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1-score*
Our dataset	Wang et al. 2023.[24]	72.4%	**75.0%**	50.0%	**88.2%**	0.600
	Li et al., 2024.[5]	73.52%	74.98%	66.25%	67.88%	0.7005
	M-ENAS (Ours)	**75.00%**	66.67%	**90.91%**	61.54%	**0.7692**
REST-*meta*-MDD	Zhu et al., 2023.[25]	72.1%	62.5%	**79.2%**	67.1%	0.699
	Dai et al., 2024.[26]	**74.7%**	–	61.4%	**80.5%**	–
	M-ENAS (Ours)	73.68%	**69.95%**	74.13%	72.88%	**0.7198**

A detailed analysis of experimental results is provided in the discussion section.

The searched optimal network architecture

The final optimal network architecture was determined through supernet training using a one-shot strategy and subnet optimization with a self-designed evolutionary algorithm. As shown in Table 6, the feature extraction backbones for sMRI, DTI, and fMRI branches were obtained separately, which are detailed in terms of block depth, width, and convolution kernel size. Moreover, we reported the number of features in each modality to help illustrate the complexity of the model. Experimental results indicate that the representation of higher-order features requires more complex structures and parameters (see blocks 4–6 in Table 6), which is consistent with previous research studies.[38,39] For the optimal subnets obtained through random search and coarse-to-fine methods, the network structures can be found in Tables S2 and S3.

DISCUSSION

Interpretation of our findings

In this study, we developed an M-ENAS framework for the automatic diagnosis of MDD using multi-modal MRI. Experimental results demonstrated that the proposed M-ENAS achieved competitive performance through a series of comparative experiments.

As shown in Table 1, with single-modality input, the model using sMRI or DTI as input presented better performance. This may be attributed to the fact that MDD is considered a structural change that underlies alterations in brain function.[40] The performance of DTI was slightly better than sMRI, suggesting that MDD is more likely to induce micro-structural changes.[41] This is essential for the investigation of MDD and the discovery of biomarkers.[42] Moreover, in the dual-modality experiments, combining fMRI and DTI as model inputs yielded superior results compared with other modal combinations, consistent with the results of the single-modality experiments. Additionally, Table 1 shows that the introduction of multiple-modality information helps to improve the model's performance by leveraging complementary information between inter- and intra-modality.[43] However, it is important to note that multi-modal fusion may also introduce redundant information, which may limit the improvement in model performance.[44]

Table 2 demonstrates that the M-ENAS method exhibits strong competitiveness on both site-specific dataset and a publicly available dataset. However, some metrics are slightly lower than those of state-of-the-art methods, which may be attributed to the design of the search space. Traditional convolutional neural network (CNN)-based modules are limited in their ability to capture long-range dependencies.[45] Introducing the non-local blocks, such as transformer modules, will help to capture the long-range dependencies information and improve the model's feature extraction capabilities.[46]

Furthermore, Tables 3 and 4 reveal that the combination of the one-shot supernet construction approach and evolutionary algorithm contributes to exploring the complementary information from multi-modal neuroimaging data. This suggests that both the construction of the supernet and the optimization of the child network are two crucial stages in NAS.[15,47] In particular, we explored the impact of voxel-based

Table 3. Comparison of different encodings between classical deep learning models and M-ENAS using multi-modal inputs

Models	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1-score*	*Parameter*(M)
AlexNet	68.82%	**74.93%**	55.26%	80.56%	0.6159	15.64
VGGNet	70.00%	71.05%	66.25%	73.33%	0.6676	138
ResNet	66.76%	68.34%	56.25%	76.11%	0.6119	199.72
GoogLeNet	62.50%	60.00%	54.55%	69.23%	0.5714	148.73
DenseNet	71.18%	74.80%	60.00%	**81.11%**	0.6557	51.51
VIT	66.67%	50.00%	52.08%	79.62%	0.5784	86.0
Swin-Transformer	69.11%	73.06%	53.75%	82.77%	0.6138	82.71
Repvit	69.56%	57.14%	50.00%	80.00%	0.5333	7.11
M-ENAS (Ours)	**75.00%**	66.67%	**90.91%**	61.54%	**0.7692**	**5.61**

Table 4. Comparison of classical NAS models with M-ENAS

Models	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1-score*	*Times* (s/epoch)	*Parameter*(M)
3D Darts	62.50%	60.00%	54.55%	**69.23%**	0.5714	745.57	8.56
3D FB-Net V1	66.67%	60.00%	81.82%	53.85%	0.6923	**97.20**	5.77
M-ENAS (Ours)	**75.00%**	**66.67%**	**90.91%**	61.54%	**0.7692**	694.14	**5.61**

fMRI features like amplitude of low-frequency fluctuations (ALFF) and regional homogeneity (ReHo) on classification performance (see Table S4 in the supplemental information). We found that three different resolutions of fMRI features exhibited similar performance in single-input experiments. On the other hand, the feature of functional connectivity matrix (FCM) still exhibited superior classification performance in multimodal fusion analysis, suggesting that region-level features provide more valuable information for the diagnosis of MDD.

In addition, Tables 5 and 6 show that the optimal subnet architecture was obtained using a customized evolutionary algorithm. Compared to models architectures obtained from random search and coarse-to-fine search strategies (see Tables S2 and S3 in the supplemental information), the self-designed evolutionary algorithm identified more complex and efficient feature encoding structures, including a number of convolutional kernels and channels. Notably, the model's inference time and resource consumption increase with the number of convolutional layers and parameters. In practice, this cost is acceptable as it also enhances the model's feature extraction capabilities and robustness.

In brief, M-ENAS is a promising approach for multi-modal neuroimaging feature extraction, fusion, and classification. This study has designed an evolutionary NAS algorithm for MDD diagnosis, providing evidence for the application of NAS in MDD diagnosis.

Occlusion analysis

In this section, we explored the regions of interest selected by M-ENAS in MDD diagnosis using occlusion analysis. Specifically, we investigated the impact on performance when different brain regions were occluded separately using the anatomical automatic labeling (AAL) atlas.[48] Furthermore, we showed the contribution of different subnetworks in MDD diagnosis with reference to the Thomas yeo seven-network cortical parcellation,[49] as shown in Figure 3.

Our findings indicated that brain regions within the SMN play important roles in the diagnosis of MDD. The SMN is a complex neural control system that plays a crucial role in the continuous sensory process of environmental stimuli and proprioceptive information.[50] Previous studies have indicated that abnormalities in the SMN are a common transdiagnostic feature in the diagnosis of psychiatric disorders.[51,52] A significant reduction in intra-network connectivity within the SMN in MDD patients is associated with the severity of depression and the frequency of depressive episodes.[53,54] Gorwood P et al. indicated that psychomotor disturbances progressively worsen with the accumulation of depressive episodes.[55] Additionally, dysfunction in SMN regions may be associated with physical symptoms of recurrent MDD, such as pain and fatigue.[56] In summary, abnormalities in the SMN system may be the underlying cause of non-emotional symptoms, and our findings provide additional evidence to support this hypothesis.

Limitations of the study

Although the proposed model has achieved promising performance, it still faces several limitations and challenges. First, despite the promising performance achieved by the M-ENAS, it is important to acknowledge that it currently falls short of what is considered suitable for real clinical applications. Second, M-ENAS requires high-performance computational resources, making practical implementation challenging in hospital settings. Third, achieving an ideal balance between the design of the search space and search strategy is often challenging, potentially resulting in low search efficiency or failure to find the global optimal solution. Fourth, small-scale data are insufficient to ensure the predictive performance of the model, as the training of supernet relies on a large amount of data. Fifth, NAS methods currently lack sufficient interpretability, making it difficult to extract valuable design principles from search results. Finally, although the stratified cross-validation strategy ensures balance between classes, it does not take into account other participant characteristics. As a result, the unavailability of certain clinical information may limit the generalizability of our findings.

RESOURCE AVAILABILITY

Lead contact

Further information and requests can be directed to the lead contact, Bin Hu (bh@lzu.edu.cn).

Table 5. Comparison of different subnet search strategies

Search strategies	*Accuracy*	*Precision*	*Recall*	*Specificity*	*F1-score*	*Parameter*(M)
Random Search	70.83%	62.50%	90.91%	53.85%	0.7407	3.44
Coarse-to-fine	66.67%	61.54%	72.73%	61.44%	0.6667	**3.06**
Evolutionary (Ours)	**75.00%**	**66.67%**	**90.91%**	**61.54%**	**0.7692**	5.61

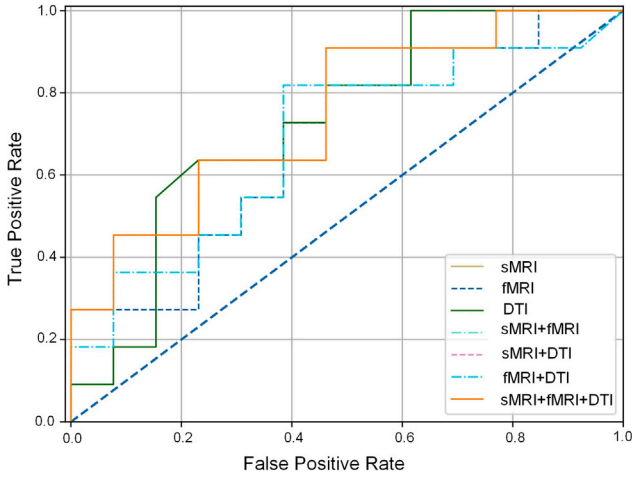


Figure 1. The ROC curves of the comparison of different inputs in the M-ENAS

Materials availability

This study did not generate new physical materials.

Data and code availability

- The imaging datasets used in this study are publicly available and listed in the key resources table.
- The original code and pre-trained models have been deposited in the code repository (https://github.com/TTLi1996/M-ENAS).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported in part by the Science and Technology Innovation 2030-Major Project (2021ZD0200701), by the National Natural Science Foundation of China (U21A20520), in part by the National Natural Science Foundation of China (62227807), in part by the National Key Research and Development Program of China (2019YFA0706200), in part by the Science and Technology Program of Gansu Province (23YFGA0004 and 24JRRA506), and in part by the Fundamental Research Funds for the Central Universities (lzujbky-2024-it16). In addition, we sincerely thank the REST-meta-MDD consortium for generously providing valuable data.

AUTHOR CONTRIBUTIONS

Conceptualization, T.L.; methodology, T.L. and J.Y.; validation, T.L. and J.Y.; investigation, T.L. and J.Y.; data curation, N.H. and S.M.; writing – original draft preparation, T.L.; writing – review and editing, T.L., Z.Z., Q.S., and M.C.; supervision, Z.Y.; project administration, Z.Y.; funding acquisition, J.Z., Z.Y., and B.H. All authors have read and agreed to the published version of the manuscript.

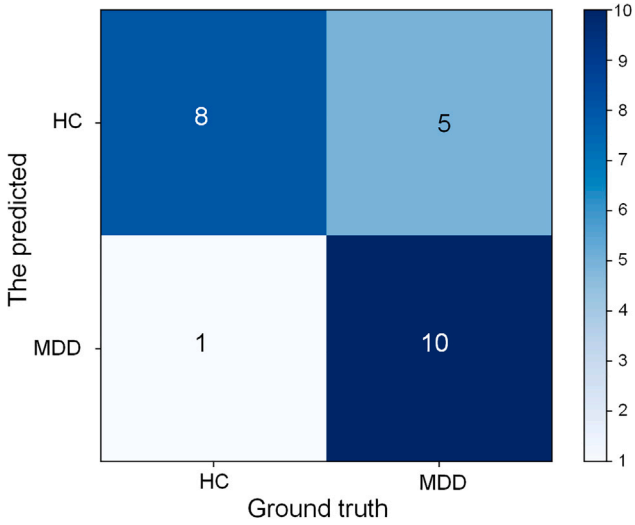


Figure 2. The confusion matrix of M-ENAS

Table 6. The searched optimal network architecture using a self-designed evolutionary algorithm

		sMRI branch				DTI branch				fMRI branch			
Name	SE Att	Depths	Channels	Kernel Sizes	Number of feature	Depths	Channels	Kernel sizes	Number of feature	Depths	Channels	Kernel sizes	Number of feature
First_Conv	–	1	32	3	45359104	1	40	3	45660160	1	32	3	107648
Bolck 1	✓	1	16	3	22679552	2	24, 16	3, 3	18264064	2	24, 16	3, 3	53824
Bolck 2	✓	2	32, 32	3, 3	11339776	2	32, 24	3, 3	6849024	3	32, 32, 32	3, 3, 3	26912
Bolck 3	✓	2	48, 48	3, 3	4252416	3	40, 40, 40	3, 5, 3	2853760	2	48, 40	3, 5	9000
Bolck 4	✓	2	80, 88	5, 3	1949024	3	88, 88, 80	5, 5, 3	1426880	2	88, 88	3, 5	5632
Bolck 5	✓	2	128, 128	3, 5	2834944	4	112, 120, 120, 120	3, 5, 3, 3	2140320	4	112, 112, 112, 112	5, 5, 3, 5	7168
Bolck 6	✓	2	200, 216	5, 3	1195992	2	208, 192	5, 3	856128	3	200, 192, 192	3, 3, 5	3072
Bolck 7	✓	2	352, 352	3, 3	1949024	1	320	3	1426880	2	320, 352	5, 5	5632
Last_Conv	–	1	1280	1	7087360	1	1280	1	5707520	1	1280	1	20480

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Data acquisition and preprocessing
  - M-ENAS
  - One-shot strategy with supernet training
  - Stacked cell architecture space
  - Evolutionary learning search strategy
  - Feature fusion and classification
  - Implementation details
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111020.

Received: May 30, 2024
Revised: August 17, 2024
Accepted: September 20, 2024
Published: September 24, 2024

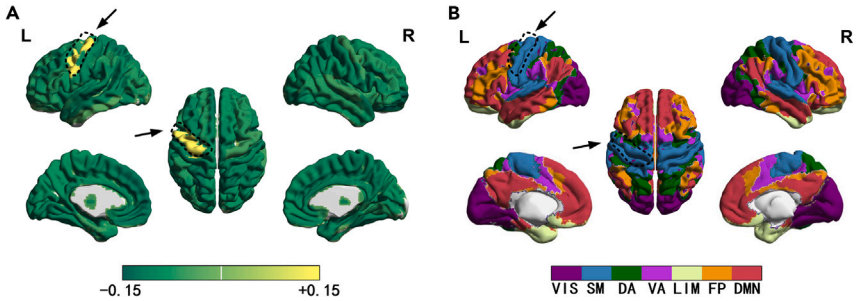


Figure 3. Investigating the contribution of different brain regions through occlusion analysis

(A) The impacts on performance when different brain regions are occluded using the AAL atlas.

(B) Yeo seven-network parcellation (black arrows indicate the SMN regions).

REFERENCES

1. Vos, T., Lim, S.S., Abbafati, C., Abbas, K.M., Abbasi, M., Abbasifard, M., Abbasi-Kangevari, M., Abbastabar, H., Abd-Allah, F., and Abdelalim, A. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The lancet *396*, 1204–1222.
2. Woelfer, M., Kasties, V., Kahlfuss, S., and Walter, M. (2019). The role of depressive subtypes within the neuroinflammation hypothesis of major depressive disorder. Neuroscience *403*, 93–110.
3. Santomauro, D.F., Herrera, A.M.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., Abbafati, C., Adolph, C., Amlag, J.O., and Aravkin, A.Y. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet *398*, 1700–1712.
4. Zhang, L., Cai, H., Bai, W., Zou, S.-Y., Feng, K.-X., Li, Y.-C., Liu, H.-Z., Du, X., Zeng, Z.-T., Lu, C.-M., et al. (2022). Prevalence of suicidality in clinically stable patients with major depressive disorder during the COVID-19 pandemic. J. Affect. Disord. *307*, 142–148.
5. Li, T., Guo, Y., Zhao, Z., Chen, M., Lin, Q., Hu, X., Yao, Z., and Hu, B. (2024). Automated diagnosis of major depressive disorder with multi-modal MRIs based on contrastive learning: a few-shot study. IEEE Trans. Neural Syst. Rehabil. Eng. *32*, 1566–1576.
6. Gonuguntla, V., Yang, E., Guan, Y., Koo, B.B., and Kim, J.H. (2022). Brain signatures based on structural MRI: Classification for MCI, PMCI, and AD. Hum. Brain Mapp. *43*, 2845–2860.
7. Wu, D., and Mori, S. (2023). Structural Neuroimaging: From Macroscopic to Microscopic Scales. In Handbook of Neuroengineering (Springer), pp. 2917–2951.
8. Thiel, K., Meinert, S., Winter, A., Lemke, H., Waltemate, L., Breuer, F., Gruber, M., Leenings, R., Wüste, L., Rüb, K., et al. (2023). Reduced fractional anisotropy in bipolar disorder v. major depressive disorder independent of current symptoms. Psychol. Med. *53*, 4592–4602.
9. Mousavian, M., Chen, J., and Greening, S. (2020). Depression Detection Using Atlas from fMRI Images (IEEE), pp. 1348–1353.
10. Schumann, G., Andreassen, O.A., Banaschewski, T., Calhoun, V.D., Clinton, N., Desrivieres, S., Brandlistuen, R.E., Feng, J., Hese, S., Hitchen, E., et al. (2023). Addressing global environmental challenges to mental health using population neuroscience: A review. JAMA Psychiatr. *80*, 1066–1074.
11. Zhu, H., Li, T., and Zhao, B. (2023). Statistical learning methods for neuroimaging data analysis with applications. Annu. Rev. Biomed. Data Sci. *6*, 73–104.
12. Zhang, Y.-D., Dong, Z., Wang, S.-H., Yu, X., Yao, X., Zhou, Q., Hu, H., Li, M., Jiménez-Mesa, C., Ramirez, J., et al. (2020). Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. Inf. Fusion *64*, 149–187.
13. Zhou, T., Thung, K.-H., Liu, M., Shi, F., Zhang, C., and Shen, D. (2020). Multi-modal latent space inducing ensemble SVM classifier for early dementia diagnosis with neuroimaging data. Med. Image Anal. *60*, 101630.
14. Zheng, G., Zheng, W., Zhang, Y., Wang, J., Chen, M., Wang, Y., Cai, T., Yao, Z., and Hu, B. (2023). An attention-based multi-modal MRI fusion model for major depressive disorder diagnosis. J. Neural. Eng. *20*, 066005.
15. Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., and Wang, X. (2021). A comprehensive survey of neural architecture search: Challenges and solutions. ACM Comput. Surv. *54*, 1–34.
16. Cong, S., and Zhou, Y. (2023). A review of convolutional neural network architectures and their optimizations. Artif. Intell. Rev. *56*, 1905–1969.
17. Liu, X., Yao, C., Chen, H., Xiang, R., Wu, H., Du, P., Yu, Z., Liu, W., Liu, J., and Geng, D. (2023). BTSC-TNAS: A neural architecture search-based transformer for brain tumor segmentation and classification. Comput. Med. Imaging Graph. *110*, 102307.
18. Dai, H., Ge, F., Li, Q., Zhang, W., and Liu, T. (2020). Optimize CNN Model for FMRI Signal Classification via Adanet-Based Neural Architecture Search (IEEE), pp. 1399–1403.
19. Lu, Z., Cheng, R., Jin, Y., Tan, K.C., and Deb, K. (2024). Neural architecture search as multiobjective optimization benchmarks: Problem formulation and performance assessment. IEEE Trans. Evol. Comput. *28*, 323–337.
20. Dong, X., Kedziora, D.J., Musial, K., and Gabrys, B. (2024). Automated deep learning: Neural architecture search is not the end. FNT. in Machine Learning *17*, 767–920.
21. He, Y., and Xiao, L. (2024). Structured pruning for deep convolutional neural networks: A survey. IEEE Trans. Pattern Anal. Mach. Intell. *46*, 2900–2919.
22. Cui, F., Huang, X., Jing, Y., Luo, Y.-j., Liu, J., and Gu, R. (2022). How resource sharing resists scarcity: the role of cognitive empathy and its neurobiological mechanisms. Cereb. Cortex *32*, 5330–5342.
23. Xia, M., Si, T., Sun, X., Ma, Q., Liu, B., Wang, L., Meng, J., Chang, M., Huang, X., Chen, Z., et al. (2019). Reproducibility of functional brain alterations in major depressive disorder: Evidence from a multisite resting-state functional MRI study with 1,434 individuals. Neuroimage *189*, 700–714.
24. Wang, J., Li, T., Sun, Q., Guo, Y., Yu, J., Yao, Z., Hou, N., and Hu, B. (2023). Automatic Diagnosis of Major Depressive Disorder Using a High-and Low-Frequency Feature Fusion Framework. Brain Sci. *13*, 1590.
25. Zhu, M., Quan, Y., and He, X. (2023). The classification of brain network for major depressive disorder patients based on deep graph convolutional neural network. Front. Hum. Neurosci. *17*, 1094592.
26. Dai, P., Shi, Y., Lu, D., Zhou, Y., Luo, J., He, Z., Chen, Z., Zou, B., Tang, H., Huang, Z., and Liao, S. (2024). Classification of recurrent major depressive disorder using a residual denoising autoencoder framework: Insights from large-scale multisite fMRI data. Comput. Methods Programs Biomed. *247*, 108114.
27. Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. *25*, 1097–1105.
28. Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations.
29. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
30. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826.
31. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
32. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., and Gelly, S. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations.
33. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012–10022.
34. Wang, A., Chen, H., Lin, Z., Han, J., and Ding, G. (2024). Repvit: Revisiting mobile cnn from vit perspective. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15909–15920.
35. Liu, H., Simonyan, K., and Yang, Y. (2018). Darts: Differentiable architecture search. In International Conference on Learning Representations.
36. Wu, B., Dai, X., Zhang, P., Wang, Y., Sun, F., Wu, Y., Tian, Y., Vajda, P., Jia, Y., and Keutzer, K. (2019). Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10734–10742.
37. Yu, J., Jin, P., Liu, H., Bender, G., Kindermans, P.-J., Tan, M., Huang, T., Song, X., Pang, R., and Le, Q. (2020). Bignas: Scaling up Neural Architecture Search with Big Single-Stage Models (Springer), pp. 702–717.
38. Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., Lu, T., Luo, P., and Shao, L. (2021). Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 568–578.
39. Lin, Q., Li, T., Cao, C., Cao, Y., Man, Z., and Wang, H. (2021). Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images. Sci. Rep. *11*, 4223.
40. Yao, Z., Zou, Y., Zheng, W., Zhang, Z., Li, Y., Yu, Y., Zhang, Z., Fu, Y., Shi, J., Zhang, W., et al. (2019). Structural alterations of the brain preceded functional alterations in major depressive disorder patients: evidence from multimodal connectivity. J. Affect. Disord. *253*, 107–117.
41. Cullen, K.R., Klimes-Dougan, B., Muetzel, R., Mueller, B.A., Camchong, J., Houri, A., Kurma, S., and Lim, K.O. (2010). Altered white matter microstructure in adolescents with major depression: a preliminary study. J. Am. Acad. Child Adolesc. Psychiatry *49*, 173–183.e1.
42. Kennis, M., Gerritsen, L., van Dalen, M., Williams, A., Cuijpers, P., and Bockting, C. (2020). Prospective biomarkers of major depressive disorder: a systematic review and meta-analysis. Mol. Psychiatry *25*, 321–338.

43. Xu, P., Zhu, X., and Clifton, D.A. (2023). Multimodal learning with transformers: A survey. IEEE Trans. Pattern Anal. Mach. Intell. *45*, 12113–12132.
44. Mai, S., Zeng, Y., and Hu, H. (2023). Multimodal information bottleneck: Learning minimal sufficient unimodal and multimodal representations. IEEE Trans. Multimedia *25*, 4121–4134.
45. Fu, Z., Li, J., Ren, L., and Chen, Z. (2023). SLDDNet: Stage-wise short and long distance dependency network for remote sensing change detection. IEEE Trans. Geosci. Remote Sens. *61*, 1–19.
46. Yuan, F., Zhang, Z., and Fang, Z. (2023). An effective CNN and Transformer complementary network for medical image segmentation. Pattern Recogn. *136*, 109228.
47. Hu, Y., Wang, X., Li, L., and Gu, Q. (2021). Improving one-shot NAS with shrinking-and-expanding supernet. Pattern Recogn. *118*, 108025.
48. Rolls, E.T., Huang, C.-C., Lin, C.-P., Feng, J., and Joliot, M. (2020). Automated anatomical labelling atlas 3. Neuroimage *206*, 116189.
49. Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.
50. Herman, A.M., Palmer, C., Azevedo, R.T., and Tsakiris, M. (2021). Neural divergence and convergence for attention to and detection of interoceptive and somatosensory stimuli. Cortex *135*, 186–206.
51. Huang, C.-C., Luo, Q., Palaniyappan, L., Yang, A.C., Hung, C.-C., Chou, K.-H., Zac Lo, C.Y., Liu, M.-N., Tsai, S.-J., Barch, D.M., et al. (2020). Transdiagnostic and illness-specific functional dysconnectivity across schizophrenia, bipolar disorder, and major depressive disorder. Biol. Psychiatry. Cogn. Neurosci. Neuroimaging *5*, 542–553.
52. Kebets, V., Holmes, A.J., Orban, C., Tang, S., Li, J., Sun, N., Kong, R., Poldrack, R.A., and Yeo, B.T.T. (2019). Somatosensory-motor dysconnectivity spans multiple transdiagnostic dimensions of psychopathology. Biol. Psychiatry *86*, 779–791.
53. Liu, J., Fan, Y., Zeng, L.-L., Liu, B., Ju, Y., Wang, M., Dong, Q., Lu, X., Sun, J., Zhang, L., et al. (2021). The neuroprogressive nature of major depressive disorder: evidence from an intrinsic connectome analysis. Transl. Psychiatry *11*, 102.
54. Yang, H., Chen, X., Chen, Z.-B., Li, L., Li, X.-Y., Castellanos, F.X., Bai, T.-J., Bo, Q.-J., Cao, J., Chang, Z.-K., et al. (2021). Disrupted intrinsic functional brain topology in patients with major depressive disorder. Mol. Psychiatry *26*, 7363–7371.
55. Gorwood, P., Richard-Devantoy, S., Baylé, F., and Cléry-Melun, M.-L. (2014). Psychomotor retardation is a scar of past depressive episodes, revealed by simple cognitive tests. Eur. Neuropsychopharmacol *24*, 1630–1640.
56. Bair, M.J., Robinson, R.L., Katon, W., and Kroenke, K. (2003). Depression and pain comorbidity: a literature review. Arch. Intern. Med. *163*, 2433–2445.
57. Yan, C.-G., Chen, X., Li, L., Castellanos, F.X., Bai, T.-J., Bo, Q.-J., Cao, J., Chen, G.-M., Chen, N.-X., Chen, W., et al. (2019). Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. Proc. Natl. Acad. Sci. USA *116*, 9078–9083.
58. Jacob, Y., Morris, L.S., Huang, K.-H., Schneider, M., Rutter, S., Verma, G., Murrough, J.W., and Balchandani, P. (2020). Neural correlates of rumination in major depressive disorder: A brain network analysis. Neuroimage. Clin. *25*, 102142.
59. Hellewell, S.C., Welton, T., Maller, J.J., Lyon, M., Korgaonkar, M.S., Koslow, S.H., Williams, L.M., Rush, A.J., Gordon, E., and Grieve, S.M. (2019). Profound and reproducible patterns of reduced regional gray matter characterize major depressive disorder. Transl. Psychiatry *9*, 176.
60. Zhao, Z., Zhang, Y., Chen, N., Li, Y., Guo, H., Guo, M., Yao, Z., and Hu, B. (2022). Altered temporal reachability highlights the role of sensory perception systems in major depressive disorder. Prog. Neuro-Psychopharmacol. Biol. Psychiatry *112*, 110426.
61. Chen, N., Guo, M., Li, Y., Hu, X., Yao, Z., and Hu, B. (2023). Estimation of discriminative multimodal brain network connectivity using message-passing-based nonlinear network fusion. IEEE/ACM Trans. Comput. Biol. Bioinform. *20*, 2398–2406.
62. Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.
63. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M.P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit. Med. *3*, 136.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Multimodal MRIs	This paper	https://github.com/TTLi1996/M-ENAS
REST-*meta*-MDD	Yan et al.[57]	https://rfmri.org/REST-meta-MDD
Software and algorithms		
Python 3.8	Python Software Foundation	https://www.python.org/
Pytorch 1.13.0	PyTorch Foundation	https://pytorch.org/
Pycharm	JetBrains	https://www.jetbrains.com/pycharm/
M-ENAS	This paper	https://github.com/TTLi1996/M-ENAS

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

We collected three modalities of neuroimaging data, including sMRI, fMRI, and DTI, from 128 participants at Gansu Provincial Hospital, including 62 patients with MDD and 66 age- and sex-matched healthy controls (HCs). All MDD patients and HCs underwent a clinical diagnosis using the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID) and non-patient edition, respectively. Furthermore, we included 116 participants in this study, consisting of 54 MDD patients and 62 HCs after data preprocessing (see method details). Notably, all participants signed informed consent forms, indicating their understanding of this study. This work was supported by the Ethics Committee of Gansu Provincial Hospital (Approval No. 2017-071).

Among this MDD patients, 11 patients (20.3%) were free of antidepressants or other psychotropic medications for at least 4 weeks (8 weeks for fluoxetine) prior to MRI scanning,[58] 9 patients (16.7%) were treated with monotherapy (using one of sertraline, lexapro, venlafaxine, or fluoxetine), and 32 patients (59.3%) received combination therapy (using two or more of the following: paroxetine, citalopram, fluvoxamine, fluoxetine, amitriptyline, venlafaxine, or trazodone).

In addition, to validate the generalizability of the M-ENAS, we expanded the validation of our model by incorporating an open-access dataset: REST-*meta*-MDD (https://rfmri.org/REST-meta-MDD),[57] which is currently the largest public MDD dataset. Of note, we selected non-replicated subjects with fMRI scan times ≥170-time points from this database[14] and used the corresponding gray matter volume (GMV) and FCM as inputs. In total, 1179 MDD and 1008 HCs from 20 different cohorts were included in the analysis. The table below outlines the demographic characteristics of the participants.

Demographic characteristics of the participants

	Our dataset		REST-*meta*-MDD	
	HCs	MDD	HCs	MDD
Number	62	54	1,008	1,179
Male/female	27/35	30/24	426/582	425/754
Age (±years)	33.5 ± 12.17	33.0 ± 11.62	36.9 ± 15.9	36.8 ± 14.8
HAMD (17-item)	–	17.62 ± 5.95	–	–
HAMA	–	17.19 ± 7.58	–	–

HCs, healthy controls; MDD, major depressive disorder; HAMD, Hamilton depression rating scale; HAMA, Hamilton anxiety scale.

METHOD DETAILS

Data acquisition and preprocessing

All participants underwent sMRI, rs-fMRI and DTI scans using a 3.0T MRI scanner (Siemens). Participants were instructed to keep their eyes closed and avoid sleeping or thinking during the scanning process. The parameter settings and cautions during image scanning are provided in ref. [5].

Structural Magnetic Resonance Imaging preprocessing

We employed a voxel-based morphometry analysis method to extract GMV as a feature of sMRI using CAT12 (https://neuro-jena.github.io/cat), retaining the default parameter settings in the preprocessing. The standard preprocessing pipeline has been previously described.[59]

fMRI preprocessing

Rs-fMRI images were preprocessed using the DPARSF toolbox (http://www.restfmri.net). The standard preprocessing process was reported in referent.[60] Of note, after extracting BOLD signals from 116 brain regions using AAL atlas, the static FCM was calculated according to Equation 1.

$$Corr(r_i, r_j) = \frac{\mathrm{cov}(r_i, r_j)}{\sigma_{ri}\sigma_{rj}} \quad \text{(Equation 1)}$$

where r_i and r_j are BOLD signals of different brain regions, the cov(·) is the covariance between r_i and r_j, while σ_{ri} and σ_{rj} denote the standard deviation of r_i and r_j, respectively.

Diffusion Tensor Imaging preprocessing

We preprocessed DTI data using the PANDA toolbox (http://www.nitrc.org/projects/panda). The fractional anisotropy mapping (FAM) was extracted and used as a feature of the DTI images, where the preprocessing pipelines have been previously described.[61]

Furthermore, the exclusion criteria were as follows: (a) data with missing modalities and excessive disturbance (head movement rotation >2°), (b) the presence of psychiatric disorders other than depression, (c) the presence of comorbidities, (d) participants with a history of illicit drug abuse (such as heroin). In total, we included 116 participants in this study, consisting of 62 HCs and 54 MDD patients with a HAMD scores >7.

M-ENAS

As shown in Figure S1 (in the supplemental information), the M-ENAS architecture consists of 2 main components, including data preprocessing and M-ENAS model construction. Details are as follows.

Data preprocessing

This component focuses on data preparation. The sMRI, DTI, and fMRI were preprocessed to calculate GMV, FAM, and FCM, respectively. These data were then fed in parallel as inputs to the M-ENAS model.

M-ENAS framework

This component aims to construct a NAS model for the diagnosis of MDD. Firstly, the set of child network structures used for architectural search was generated by customizing the search space. Then, the weights of the supernet were trained using a One-shot training strategy. Finally, the optimal network was determined as the target network using a self-designed evolutionary search strategy.

In summary, the M-ENAS architecture obtains the optimal weights and the child network architecture through a two-stage architecture search approach. A high-quality network architecture is important for adapting to the data due to the distinct initialization of small and big child models. We employ a One-shot training technique to train the supernet weights, and then search for the optimal child model using a self-designed evolutionary algorithm. Details are illustrated in Figure S2.

One-shot strategy with supernet training

The training stage focuses on training a robust supernet, as shown in Figure S2 (a). A high-quality supernet plays a crucial role in generating or controlling the parameters of a child model. In this stage, the Sandwich Rule and Inplace Distillation strategy were adopted for supernet weight training, alleviating the problem of high computational overhead of traditional supernet training.

Sandwich Rule: The sandwich rule samples the biggest model, the smallest model, and N (=2) child models in each training step. The weights of the supernet were updated by aggregating and back-propagating gradients from the child models. Specifically, the biggest child model represents the model with the highest parameter settings, including model depth, width, convolution kernel size, etc. By sampling different child models, this strategy improves both the lower and upper bounds of the model, leading to an overall performance improvement in the supernet's performance.

Inplace distillation

The inplace distillation takes the predicted results of the biggest child model as soft labels to supervise other child models. Specifically, during the same training batch, the soft labels were supervised using ground truth to calculate cross-entropy loss (i.e., Loss 1, see Figure S2A; Equation 2). Simultaneously, all other child models were supervised using these soft labels to calculate distillation loss (i.e., Loss2 and Loss3, see Figure S2A; Equations 3 and 4). Furthermore, the loss of the M-ENAS model is a comprehensive loss, as shown in Equation 5. Of note, predictions from the largest child model can be used in the gradient update process to "teach" all other child models.

$$Loss1 = CrossEntropyLoss(soft\_label, ground\_truth) \quad \text{(Equation 2)}$$

$$Loss2 = DistillationLoss(prediction\_smallest, soft\_label) \quad \text{(Equation 3)}$$

$$Loss3 = DistillationLoss(prediction\_n_i, soft\_label) \quad \text{(Equation 4)}$$

$$Loss = Loss1 + Loss2 + Loss3 \quad \text{(Equation 5)}$$

where 'soft_label' is the prediction of the biggest child model, the 'prediction_smallest' and 'predicted_n_i' represent predicted labels from the smallest child model and N sampled child models, respectively.

In summary, we train supernet weights using the Sandwich Rule and the Inplace Distillation strategy, which not only optimizes the training process but also reduces the computational burden.

Stacked cell architecture space

We present a pre-defined search space with a stacked cell architecture. Specifically, 7 search cells were stacked to form a network backbone. To better focus the model on regions of interest, we introduced the squeeze-and-excitation attention (SE Att)[62] in each cell. On the other hand, we aim to improve the model's ability to capture fine-grained local information by exploring different combinations of network parameters in each cell, such as network depth, channel number, and convolutional kernel size. Detailed settings of the search space are reported in Table S1 (in the supplemental information).

Evolutionary learning search strategy

The target architecture was obtained in the search space using a self-designed evolutionary learning search algorithm. As shown in Figure S2B. Concretely, the N (=10) subnet was picked as the parent model from the trained supernet to generate the next generation. The next-generation models were continuously generated from the parent model through random crossover and mutation. For the crossover, two subnets were randomly picked from the parent model as the initial net for the crossover. For the mutation, a candidate mutation operation randomly varies the depth of the network with a probability $P_d = 0.2$. It also modifies the number of channels and the convolution kernel size in each block. The optimal network architecture was obtained after 100 iterations of crossover and mutation.

In summary, we aimed to maximize the model's performance while minimizing its size. We employed a self-designed evolutionary search strategy to explore the optimal model based on the principle of survival of the fittest.

Feature fusion and classification

The multi-modal features were integrated and recognized through the feature fusion module and feature classification module, respectively, as presented in Figure S3. Specifically, the features from sMRI, DTI and fMRI were extracted using the optimal network. The extracted features were then integrated by concatenating them along the channel dimensions.[63] Finally, the integrated features were input into a feature classification module, comprising an FC layer with a softmax function, to produce predictions for recognizing between HCs and MDD.

Implementation details

In this study, we employed a stratified 5-fold cross-validation approach to train and test the M-ENAS. In other words, the dataset was divided into five non-overlapping subsets with a ratio of 3:1:1 for training, validation, and test subsets, respectively. Moreover, the Adam optimizer with an initial learning rate of 0.001 and cosine annealing weight decay was used to optimize the training process. Additionally, to prevent the risk of model overfitting, an early stopping strategy was implemented. The training was stopped when the validation accuracy failed to improve for 10 consecutive epochs, reaching the patience threshold.

The experiments related to this study were compiled using PyTorch-1.13 and executed on an NVIDIA A100 GPU, running on Ubuntu 20.04.

QUANTIFICATION AND STATISTICAL ANALYSIS

We applied a variety of evaluation metrics to comprehensively assess the performance of the proposed model, including accuracy, precision, recall, specificity, F1-score, and the number of training parameters, which are calculated according to Equations 6, 7, 8, 9, and 10, respectively.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{(Equation 6)}$$

$$Precision = \frac{TP}{TP+FP} \quad \text{(Equation 7)}$$

$$Recall = \frac{TP}{TP+FN} \quad \text{(Equation 8)}$$

$$Specificity = \frac{TN}{TN+FP} \quad \text{(Equation 9)}$$

$$F1 - \text{score} = 2 \times \frac{Prec \times Rec}{Prec+Rec} \quad \text{(Equation 10)}$$

where the TP, FP, TN, and FN are True Positive, False Positive, True Negative, and False Negative, respectively.

In addition, the reported experimental results are mean using the stratified 5-fold cross-validation and can be found in Tables 1, 2, 3, 4, and 5.