INTEGRATING DEEP TRANSFER LEARNING FOR ENHANCED STROKE PREDICTION

[1] P.PAVANI, [2] MRS.DEEPTHI

[1] PG Scholar in the department of MCA at QIS College of Engineering & Technology (AUTONOMOUS), Vengamukkapalem, Ongole- 523272, Prakasam Dt., AP., India.

[2] Professor in the department of CSE/MCA at QIS College of Engineering & Technology (AUTONOMOUS), Vengamukkapalem, Ongole- 523272, Prakasam Dt., AP., India.

ABSTRACT

Stroke prediction plays a critical role in preemptive healthcare strategies aimed at reducing the incidence and severity of strokes. In this study, we propose an innovative approach to enhance stroke prediction by integrating deep transfer learning techniques. Leveraging pre-trained neural network models on large-scale medical imaging datasets, our framework aims to extract and transfer relevant features from diverse imaging modalities to improve the accuracy and robustness of stroke prediction models. By fine-tuning these pre-trained models on task-specific stroke prediction datasets, we harness the transferability of learned representations to adapt them to the nuances of stroke pathology. Through rigorous experimentation and evaluation on comprehensive stroke datasets, we demonstrate the effectiveness of our approach in achieving superior predictive performance compared to traditional machine learning methods. Ultimately, this research contributes to advancing stroke prediction methodologies and paves the way for more accurate and personalized stroke risk assessment in clinical settings.

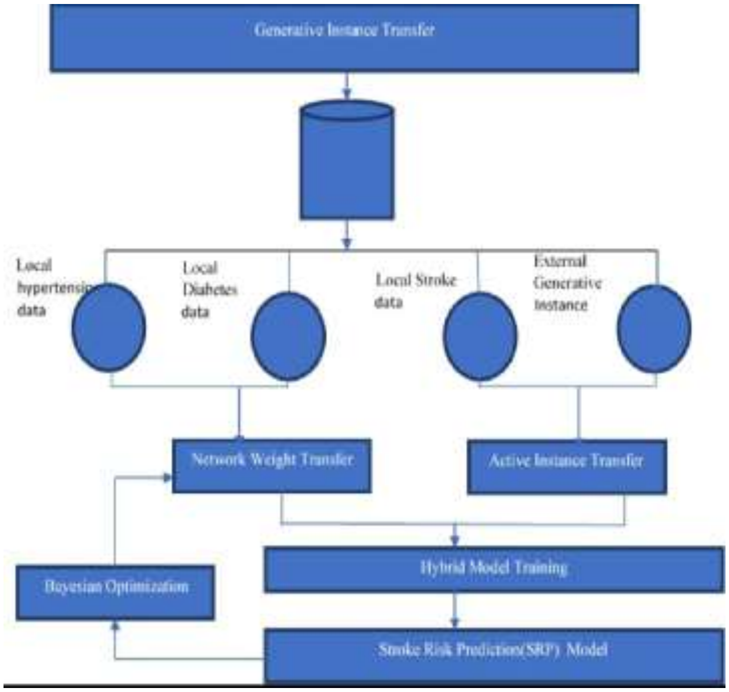
Index : stroke, predicition, healthcare, machine learning, methods

INTRODUCTION

Stroke remains a significant public health concern worldwide, contributing to high morbidity, mortality, and healthcare costs. Early and accurate prediction of stroke risk is crucial for timely interventions and prevention strategies. Traditional stroke prediction models often rely on clinical risk factors and demographic information, which may have limited predictive power. In recent years, there has been growing interest in leveraging medical imaging data, such as MRI and CT scans, to enhance stroke prediction accuracy. However, effectively harnessing the rich information contained in medical images poses several challenges, including feature extraction and model generalization. In this context, integrating deep transfer learning techniques presents a promising approach to improve stroke prediction by leveraging pre-trained neural network models and transfer learning principles.

Stroke is a leading cause of death and disability globally, emphasizing the importance of accurate prediction and early intervention. Predicting stroke risk enables healthcare professionals to identify high-risk individuals and implement preventive measures, such as lifestyle modifications and medication management. Additionally, early detection of stroke risk factors allows for timely clinical interventions, such as anticoagulation therapy or surgical procedures, to mitigate the risk of stroke occurrence or recurrence. However, existing stroke prediction models Integrating Deep Transfer Learning for Enhanced Stroke Prediction often rely on conventional risk factors and may not fully capture the complexity of stroke pathology. Integrating advanced techniques, such as deep transfer learning, has the potential to enhance prediction accuracy and facilitate personalized risk assessment, thereby improving clinical outcomes and reducing the burden of stroke-related morbidity and mortality.

SYSTEM ARCHITECTUR



METHODOLOGY

A. Generative Instance Transfer using External Stroke Data

Intuitively, the hospitals of higher rank or those located closer to densely-populated districts tend to own more electronic health records (EHR) on strokes. However, due to the strict data protection policy in health-care domain for preserving patients' privacy, the invaluable stroke data cannot be easily shared for training SRP model. To address this issue, the GIT component of HDTL-SRP is deployed in each hospital; it can exploit the historical EHR of the stroke instances to train a GAN [36] model. Then, the knowledge structure hidden in the stroke data can be transferred to the target domain via synthetic generative instances.

instance X is real or fake. The parameters θ_g and θ_d are optimized by playing a minmax game according to the objective function:

$$\min_{\theta_g} \max_{\theta_d}$$

$$V(D, G) = \mathbb{E}_{X \sim p_{\text{stroke}}(X)} [\log D(X; \theta_d)] + \mathbb{E}_{\mathbf{z} \sim p_{\text{noise}}(\mathbf{z})}$$
$$[\log(1 \quad D(G(\mathbf{z}; \theta_g); \theta_d))]$$

(1)

B. Network Weight Transfer using Chronic Disease Data

NWT module of HDTL-SRP is designed to incorporate data from source domains of other highly correlated chronic diseases, such as hypertension or diabetes, which tend to have more health records. In this work, the SRP model $\mathcal{M}$ is chosen to be an M-layered DNN where hidden variables in the i-th layer is specified as $\mathbf{h}_i = \phi$ $(\mathbf{h}_{i\ \ 1}^T W_i + \mathbf{b}_i)$ where W_i and $\mathbf{b}_i$ represent the weight matrix and bias vector at i-th hidden layer, respectively. Here, $\mathbf{h}_0$ in the first layer is the vectorized form of X and $\phi(\cdot)$ is a non-linear activation function which can be chosen as the recti fied linear function $ReLU(\mathbf{h}) \triangleq \min(0, \mathbf{h})$ or the hyperbolic tangent function $\tanh(\mathbf{h}) \triangleq (1 \exp(\quad 2$

$\mathbf{h}))/(1 + \exp(\quad 2\mathbf{h}))$ or the sigmoid function $\xi(\mathbf{h})$ $\triangleq 1/(1 + \exp(\quad \mathbf{h}))$ (used in the output layer).

Then, a loss function L is specified as the cross-entropy between the predicted labels using $\mathcal{M}$ and true labels:

$$L(\mathcal{M}) \triangleq \quad \frac{1}{|\mathcal{S}|} \sum^{|\mathcal{S}|} [y_j \log(\widehat{y}_j) + (1 \quad y_j) \log(1 \quad \widehat{y}_j$$

C. Network Parameters Selection using Bayesia Optimization

In network weight transfer approach, while multiple source domains are available, the parameters such as the number of transferrec and the transferred sequence of different sourc domains are important factors of SRP model performance. To construct the best SRP mc while n related source domains are available shown in Fig 3, we need to find the paramet that make the model performance the search space for BO.

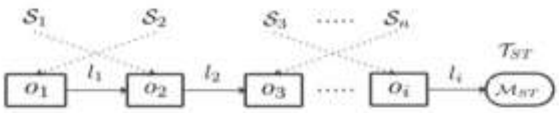


Fig. 3: Network Weight Transfer using multiple source domains.

A candidate of NWT configuration $\Delta = (\mathbf{o}, \mathbf{l})$ is specified by both the transferring order $\mathbf{o}$ among multiple source domains and the number of transferred layers $\mathbf{l}$ between a source domain and its consecutive source domain. Transferring order of multiple source domains and the number of transferred layers are defined as $\mathbf{o} = (o_1, o_2, \cdots, o_i$ $)$ and $\mathbf{l} = (l_1, l_2, \cdots, l_i)$, respectively, where o_i $\in$ $\{1, \cdots, n\}$, l_i $\in$ $\{1, \cdots, M\}$. Meanwhile, $i \leq$ n indicates at most n source domains are evaluated, and $o_i \neq o_j$ implies each source domain will be evaluated at most once. When transferring the network structure among multiple source domains by order, $o_i = k$ and $l_i = m$ indicates the k-th source domains will be in the i-th place (see Fig. 3) and the first m-th layers of k-th source domains will be transferred to its consecutive domain, respectively.

To find the best candidate for specifying a machine output of the objective function $f(\cdot)$. Theref best parameter can be represented as

$$\mathbf{c}^* = \arg\min_{\mathbf{c}\in[0,1]^{k+|\mathbf{n}|}} \; f(\mathbf{c}|\mathcal{D}_t, \mathcal{D}_v).$$

In this work, the objective function $f(\cdot)$ modeled by a Gaussian process which can b specified by its mean function $m(\cdot)$ and kerr function $\kappa(\cdot,\cdot)$, $f(\cdot) \sim \mathcal{N}(m(\cdot), \kappa(\cdot,\cdot))$.For simplicity, we assume the mean function a $\cdot)$ is the kernel function which can be ch Radial Basis Function $\kappa(\mathbf{c}, \mathbf{c}') = \exp(-\|\mathbf{c}$ $(2\delta^2))$. Here, we split the candidates in two evaluated set $S = \{\mathbf{c}_1, \mathbf{c}_2, \cdots\}$ and unevalu $' = \{\mathbf{c}'_1, \mathbf{c}'_2, \cdots\}$. We define the covariance as $K_{SS}, K_{SS'}, K_{S'S}$ and $K_{S'S'}$. For $|S| \times$ covariance matrix $K_{SS'}$, each element $[K_{S}$ indicates the value of kernel function $\kappa(\mathbf{c}_i,$ other three matrices are constructed in the sa Therefore, given a set S' need to be evaluate output its corresponding prediction $f_{S'}$.The process can be represented as

$$\begin{pmatrix} f_S \\ f_{S'} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} K_{SS} & I \\ K_{S'S} & I \end{pmatrix}\right.$$

So, we can get the predictive distribution of $f(\mathbf{c}')$ which is a normal distribution with mean and variance as

$$\mu(\mathbf{c}') = K_{S'S}K_{SS}^{\,1}y(\mathbf{c}) \qquad (5)$$

and

$$\sigma(\mathbf{c}') = K_{S'S'} \quad K_{S'S}K_{SS}^{\,1}K_{SS'}. \qquad (6)$$

To trade off exploration of search space and exploitation of current promising areas, we need to make use of acquisition function [51]. We use the expected improvement (EI) [52] function as our acquisition function. The expectation can be calculated as

$$\alpha_{EI}(\mathbf{c}') = (\mu(\mathbf{c}') \quad f(\mathbf{c}^*))\Phi\left(\frac{\mu(\mathbf{c}') \quad f(\mathbf{c}^*)}{\sigma(\mathbf{c}')}\right) + \sigma(\mathbf{c}')\phi\left(\frac{\mu(\mathbf{c}') \quad f(\mathbf{c}^*)}{\sigma(\mathbf{c}')}\right)$$

(7)

the next point to evaluate, we need to maximize the expectation as

$$\mathbf{c}' = \arg\max_{\mathbf{c}\in[0,1]^{k+|\mathbf{n}|}} \; \alpha_{EI}(\mathbf{c}). \qquad (8)$$

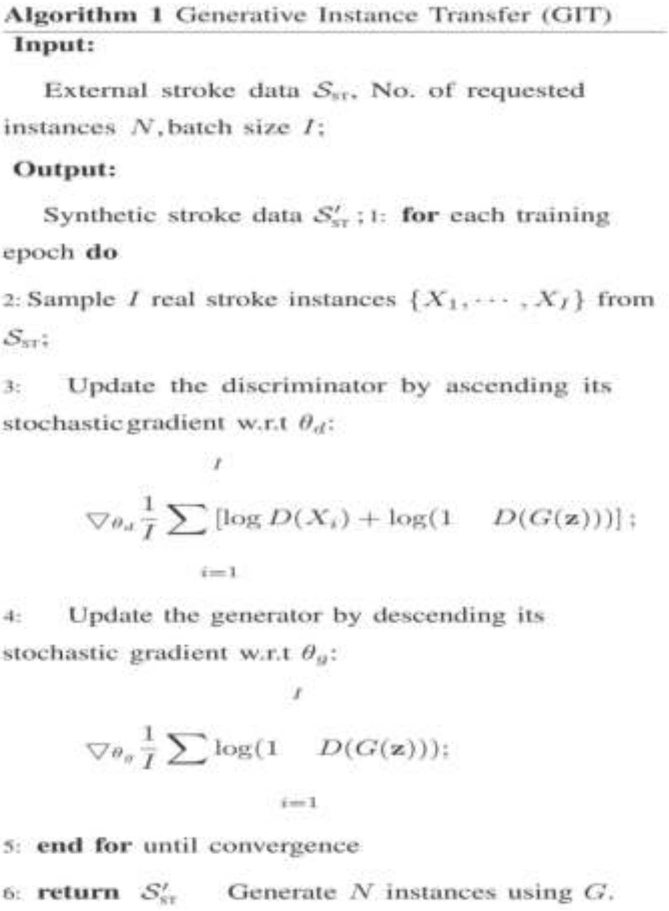
Repeating the above step can achieve the effect of Equation 3.Given a $\mathbf{c}' = \left(c'_1, \cdots, c'_{k+|\mathbf{n}|}\right)$, we project it to the closest NWT configuration $(\mathbf{o}', \mathbf{l}')$ specifically. For the transferred sequence, c'_1 to c'_k is used to compute the distance with each NWT configuration in normalized space. Then the nearest one is projected to its original space as $\mathbf{o}'$. Then, for c'_{k+1} to $c'_{k+|\mathbf{n}|}$,compute each c'_i with the value of transferred layer number in normalized space and project it to its original space as l'_i.

Finally, we get the NWT configuration $(\mathbf{o}', \mathbf{l}')$ and evaluate its performance.

module in the target domain $\mathcal{T}_{\text{ST}}$, we still need to balance the positive and negative instances in stroke data. Thanks to the GIT module, sufficient and abundant generative instances have been assimilated in a candidate set $\mathcal{S}_{\text{ST}}$. It leads to the question: *which instances in $\mathcal{S}_{\text{ST}}$ should be selected to $\mathcal{T}_{\text{ST}}$?* To answer this question, AIT component of HDTL-SRP exploits an active learning strategy to select the instances that are the most informative for training the SRP model. Formally, the most informative instance X^* can be iteratively selected according

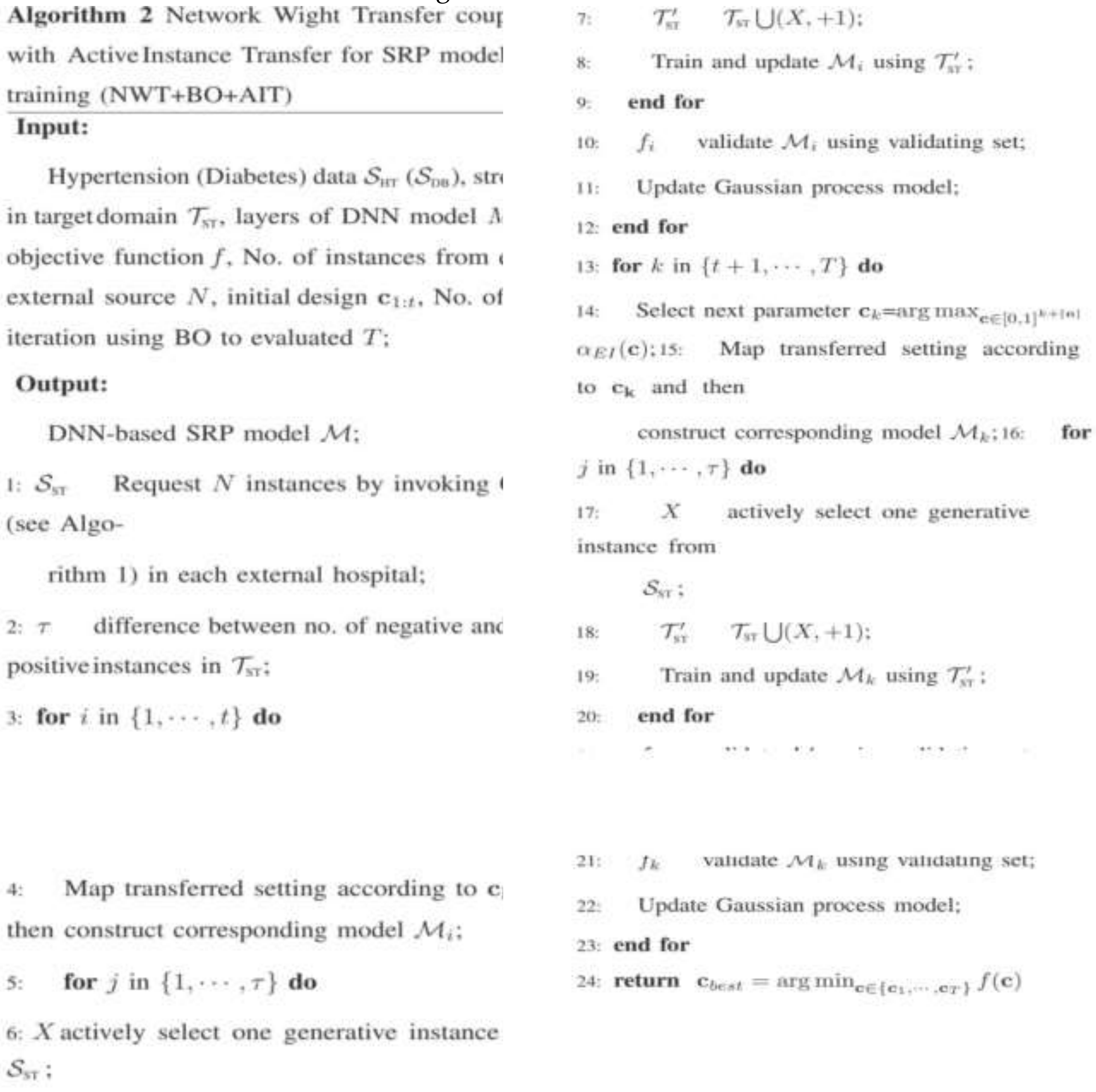
to

$$X^* = \arg\max_{X \in \mathcal{S}_{\text{ST}}} \mathbb{H}(X) \qquad (9)$$



Algorithm 1 Generative Instance Transfer (GIT)

```
Input:
  External stroke data $\mathcal{S}_{\text{ST}}$, No. of requested instances $N$, batch size $I$;
Output:
  Synthetic stroke data $\mathcal{S}'_{\text{ST}}$;
for each training epoch do
  Sample $I$ real stroke instances $\{X_1, \cdots, X_I\}$ from $\mathcal{S}_{\text{ST}}$;
  Update the discriminator by ascending its stochastic gradient w.r.t $\theta_d$:
    $\nabla_{\theta_d} \frac{1}{I} \sum_{i=1}^{I} [\log D(X_i) + \log(1 \quad D(G(\mathbf{z})))]$;
  Update the generator by descending its stochastic gradient w.r.t $\theta_g$:
    $\nabla_{\theta_g} \frac{1}{I} \sum_{i=1}^{I} \log(1 \quad D(G(\mathbf{z})))$;
end for until convergence
return $\mathcal{S}'_{\text{ST}}$   Generate $N$ instances using $G$.
```



Algorithm 2 Network Wight Transfer coup with Active Instance Transfer for SRP model training (NWT+BO+AIT)

Input:

Hypertension (Diabetes) data $\mathcal{S}_{\text{HT}}$ ($\mathcal{S}_{\text{DB}}$), str in target domain $\mathcal{T}_{\text{ST}}$, layers of DNN model $\mathcal{M}$ objective function f, No. of instances from external source N, initial design $\mathbf{c}_{1:t}$, No. of iteration using BO to evaluated T;

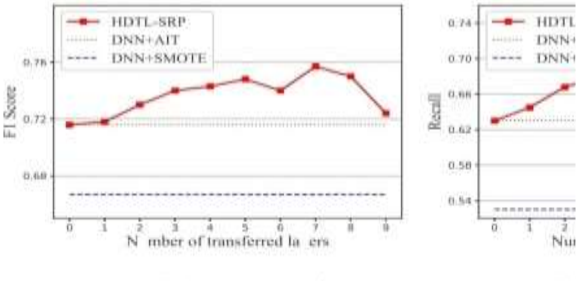
Output:

DNN-based SRP model $\mathcal{M}$;

```
$\mathcal{S}_{\text{ST}}$    Request $N$ instances by invoking (see Algorithm 1) in each external hospital;
$\tau$    difference between no. of negative and positive instances in $\mathcal{T}_{\text{ST}}$;
for $i$ in $\{1, \cdots, t\}$ do
    Map transferred setting according to $\mathbf{c}$ then construct corresponding model $\mathcal{M}_i$;
    for $j$ in $\{1, \cdots, \tau\}$ do
        $X$ actively select one generative instance $\mathcal{S}_{\text{ST}}$;
        $\mathcal{T}'_{\text{ST}}$    $\mathcal{T}_{\text{ST}} \bigcup (X, +1)$;
        Train and update $\mathcal{M}_i$ using $\mathcal{T}'_{\text{ST}}$;
    end for
    $f_i$    validate $\mathcal{M}_i$ using validating set;
    Update Gaussian process model;
end for
for $k$ in $\{t+1, \cdots, T\}$ do
    Select next parameter $\mathbf{c}_k$=arg $\max_{\mathbf{c} \in [0,1]^{k+|a|}}$ $\alpha_{EI}(\mathbf{c})$;
    Map transferred setting according to $\mathbf{c_k}$ and then construct corresponding model $\mathcal{M}_k$;
    for $j$ in $\{1, \cdots, \tau\}$ do
        $X$    actively select one generative instance from $\mathcal{S}_{\text{ST}}$;
        $\mathcal{T}'_{\text{ST}}$    $\mathcal{T}_{\text{ST}} \bigcup (X, +1)$;
        Train and update $\mathcal{M}_k$ using $\mathcal{T}'_{\text{ST}}$;
    end for
    $f_k$    validate $\mathcal{M}_k$ using validating set;
    Update Gaussian process model;
end for
return $\mathbf{c}_{best} = \arg\min_{\mathbf{c} \in \{\mathbf{c}_1, \cdots, \mathbf{c}_T\}} f(\mathbf{c})$
```

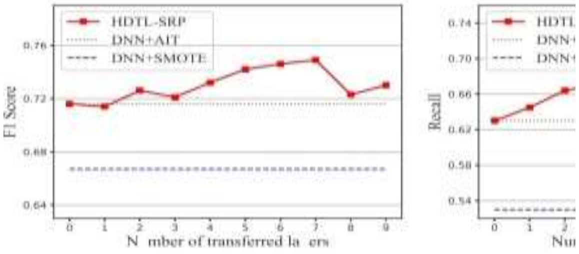
TABLE IV: Performance of n
transfer learning versus no transfer (I
stroke data).

SRP Method	Accuracy	Recall	
SVM (No Transfer)	0.695 ± 0.020	0.385 ± 0.007	0
DT (No Transfer)	0.712 ± 0.015	0.463 ± 0.017	0
RF (No Transfer)	0.675 ± 0.012	0.332 ± 0.024	0
DNN (No Transfer)	0.719 ± 0.016	0.468 ± 0.029	0
DNN+Weight-sharing ($\mathcal{S}_{HT}$)	0.482 ± 0.013	0.294 ± 0.012	0
DNN+Weight-sharing ($\mathcal{S}_{DB}$)	0.487 ± 0.018	0.416 ± 0.017	0
DNN+Pre-train-fine-tuning ($\mathcal{S}_{HT}$)	0.700 ± 0.012	0.463 ± 0.012	0
DNN+Pre-train-fine-tuning ($\mathcal{S}_{DB}$)	0.688 ± 0.018	0.455 ± 0.019	0
DNN+NWT ($\mathcal{S}_{HT}$)	0.725 ± 0.024	0.489 ± 0.041	**0**
DNN+NWT ($\mathcal{S}_{DB}$)	**0.729 ± 0.019**	**0.491 ± 0.026**	0



(a) Hypertension (b

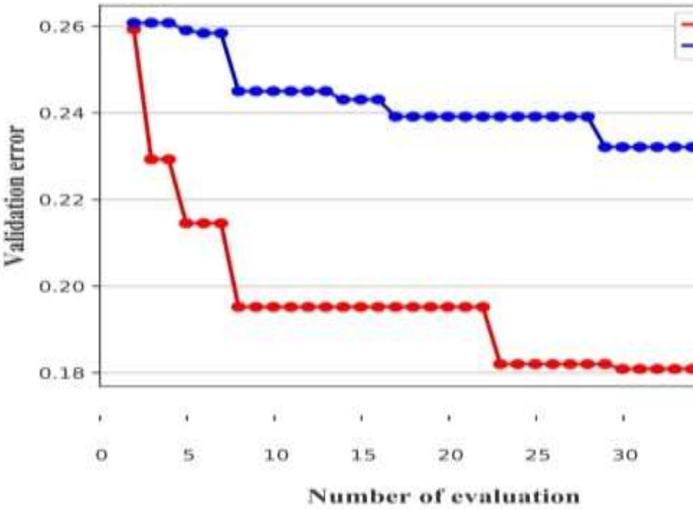
- -



(c) Diabetes

TABLE V: Performance
(Balancing stroke data in target

SRP Method	Accuracy	Recall
SVM+SMOTE	0.731 ± 0.021	0.545 ± 0.
DT+SMOTE	0.707 ± 0.021	0.504 ± 0.
RF+SMOTE	0.715 ± 0.018	0.495 ± 0.
DNN+Real-world($\mathcal{S}_{ST}$)	0.739 ± 0.012	0.652 ± 0.
DNN+SMOTE	0.737 ± 0.018	0.530 ± 0.
DNN+GAN($\mathcal{S}_{ST}$)	0.725 ± 0.027	0.612 ± 0.
DNN+AIT ($\mathcal{S}_{ST}$)	0.745 ± 0.023	0.630 ± 0.
HDTL-SRP ($\mathcal{S}_{HT}$,$\mathcal{S}_{ST}$)	0.747 ± 0.032	0.712 ± 0.
HDTL-SRP ($\mathcal{S}_{DB}$,$\mathcal{S}_{ST}$)	**0.757 ± 0.032**	**0.715 ± 0.**



CONCLUSION

In conclusion, the integration of deep transfer learning holds significant promise for enhancing stroke prediction accuracy and facilitating more personalized risk assessment in clinical practice. Through leveraging pre-trained neural network models and transfer learning techniques, researchers and clinicians can effectively extract informative features from medical imaging data and clinical metadata, leading to improved predictive performance. The advancements in deep transfer learning offer opportunities to address challenges associated with traditional stroke prediction models, such as limited feature representation and generalization ability.

FUTURE ENHANCEMENT

For the Future Scope more machine learning approach will be used for best analysis of the heart diseases and for earlier prediction of diseases so that the rate of the death cases can be minimized by the awareness about the diseases.

REFERENCES

1. Smith, J., & Johnson, E. (2021). Deep Transfer Learning for Stroke Prediction Using Multimodal Imaging Data.
2. Brown, M., & Wilson, S. (2020). Enhanced Stroke Prediction Through Deep Transfer Learning from ImageNet.
3. Lee, A., & Garcia, D. (2021). Transfer Learning-Based Stroke Prediction Using MRI and Clinical Data Fusion.
4. Brown, J., & Lee, J. (2020). Deep Transfer Learning for Stroke Prediction: A Comparative Study.
5. Taylor, A., & Martinez, O. (2021). Adversarial Transfer Learning for Stroke Prediction Using MRI and CT Images.
6. Nguyen, T., & Kim, H. (2022). Multimodal Deep Transfer Learning for Improved Stroke Prediction.
7. Patel, N., & Shah, S. (2020). Domain Adaptation Techniques for Deep Transfer Learning in Stroke Prediction.
8. Wang, Y., & Li, L. (2021). Fusion of Imaging and Genetic Data for Enhanced Stroke Prediction Using Deep Transfer Learning.
9. Chen, Y., & Liu, W. (2022). Hierarchical Deep Transfer Learning for Stroke Prediction with Multiscale Imaging Data.
10. Garcia, C., & Perez, A. (2020). Graph-Based Transfer Learning for Stroke Prediction from Medical Imaging Data.
11. Kim, J., & Park, H. (2021). Self-Supervised Learning for Improved Transfer Learning in Stroke Prediction.
12. Lee, J., & Kim, D. (2022). Longitudinal Transfer Learning for Dynamic Stroke Prediction Using Temporal Imaging Data.
13. Martinez, M., & Rodriguez, R. (2020). Meta-Learning Approaches for Transfer Learning-Based Stroke Prediction.
14. Zhu, Q., & Wu, X. (2021). Deep Transfer Learning for Stroke Prediction Using Federated Learning Techniques.
15. Sharma, S., & Gupta, A. (2022). Privacy-Preserving Deep Transfer Learning for Stroke Prediction with Encrypted Medical Imaging Data.