International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

# Deep Learning-Based Domain Adaptation Healthcare Method for Predicting Mortality Risk of ICU Patient

**Dr. (Mrs) Jomy John[1], N. Geetha[2], Dr. Veera Talukdar[3], Dr. Pravin B. Waghmare[4], Sreenivasulu Reddy L.[5], Suneetha Bandeela[6], Ankur Gupta[7]**

**Submitted**: 25/05/2023 **Revised**: 15/07/2023 **Accepted**: 26/07/2023

**Abstract:** Deep learning models often encounter issues such as insufficient labelled training data, shifts in overall data distribution, and shifts in data distribution between categories, leading to a drop in model prediction accuracywhen employed in prediction tasks in the medical area. A deep learning-based domain adaptation marginal softmax aware additive cosine margin loss prediction (DDSCM) model is presented to address such issues. The presented approachutilizes a bidirectional memory network along with an attention modulein order to retrieve the crucial features. The approach offers the concept of generative adversarial networks to reduce data distribution shifts between comprehensive data in the form of domain adversarial. The idea of metric learning is then included in the method to reduce the data distribution shift between categories even further by optimizing the decision boundary which thereby increase the domain adaptation effect and prediction accuracy of the model. At last, real-world medical data collection is used to conduct the mortality risk prediction job. The experimental findings were compared to the baseline models of the other five. The proposed method solves the problem of data distribution shift better than existing baseline models and obtains a better classification result.

***Keywords:*** *Deep Learning, Unsupervised, Domain Adaptation, Softmax, Cosine Margin Loss, BI-LSTM, Medical*

## 1 Introduction

Public health emergencies often put enormous pressure on social medical resources. For example, the outbreak of COVID-19 in early 2020 caused a shortage of medical staff and a run on medical resources. One of the reasons is that people infected with the new coronavirus are prone to "inflammatory storm" [1], which leads to rapid deterioration of the disease and an increased risk of death. Medical staffs need to invest a lot of energy to observe and track the changes of patients' physiological conditions, and they need to deploy different medical equipment according to the degree of death risk of patients. For example, extracorporeal membrane oxygenation equipment can win precious time for rescue, but the number is relatively small, which is suitable for patients with severe cardiopulmonary failure. If a deep learning model can be constructed by using the patient's vital sign data to issue an early warning with the risk of death, the energy of medical staff can be saved, medical equipment can be reasonably allocated in a timely manner, and the utilization rate of medical resources can be increased[2-7].

The effective application of deep learning models is based on a huge number of labeled training data, and often requires the test and training data to obey the same distribution, which is often not satisfied in practical applications. Due to various realistic conditions, the collected training data has certain

[1]*Associate Professor, Department of Computer Science, P M Government College, Chalakudy, Kerala, India*
*jomyeldos@gmail.com*
[2]*Assistant Professor, Department of Information Technology, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India*
*geetha.n@cit.edu.in*
[3]*Professor, Department of Computer Science, D Y Patil International University, Akurdi Pune, Maharashtra, India*
*bhaskarveera95@gmail.com*
[4] *HOD, Department of Civil Engineering, Acharya Shrimannarayan Polytechnic, Pipri, Wardha (MSBTE, Mumbai), India*
*pravin7717@gmail.com*
[5]*Associate Professor, Department of Mathematics, School of Advanced Sciences, Kalasalingam Academy of Research and Education, Krishnankoil, Tamil Nadu, India*
*sreenivasulureddy.svu@gmail.com*
[6] *Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh - 522502, India*
*suneetha.charita@gmail.com*
[7] *Assistant Professor, Department of Computer Science and Engineering, Vaish College of Engineering, Rohtak, Haryana, India*
*ankurdujana@gmail.com*

limitations, for example, a certain age group [8], a certain department or a certain complication may occupy the majority. This limitation makes deep learning models unable to make general predictions for data in other situations. To address this issue, domain adaptation techniques can make use of the source domain and destination, and the target domain receives the knowledge gained from the source domain.

However, the application of domain adaptation technique to the task of predicting mortality risk in intensive care patients still encounters three difficulties: the data distribution offset (DDO), the data distribution offset between categories (DDOC), and the diversity and complexity of time series data sex. The DDO and the DDOC between categories are shown in Figure 1:

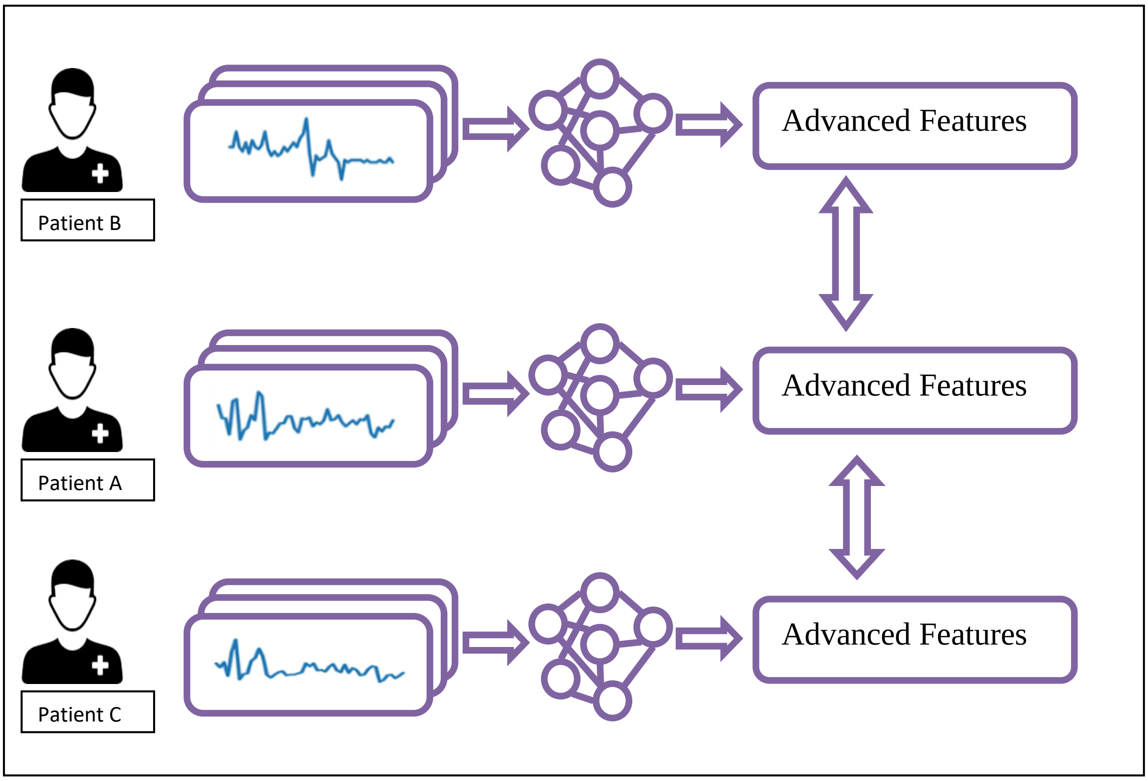


**Fig 1:** Data Distribution Shift

The DDO refers to the fact that the overall data distribution of the source domain and the destination domain are often different. For example, the elderly may make up the majority of the data collected in the intensive care unit Figure 1. The vital signs of middle-aged and elderly patients A and young patients B are not similar, which means that the data distribution of the source domain with the elderly patients as the main body and the destination domain with the young patients as the main body is different. Taking the MIMIC-III (medical information mart for intensive care III) data set in the medical field as an example, blood pressure, as one of the important indicators reflecting the physiological condition of patients, has different distributions among patients of different ages. As shown in Figure 2, the mean blood pressure of the patients gradually decreased with age. The differences in the distribution of these physiological indicators cause the models trained on the data of elderly patients to not generalize well to the data of young patients. The domain adaptation method can appropriately reduce the distance among the high-level features of patient A and patient B, and eliminate the influence of the overall data distribution shift.

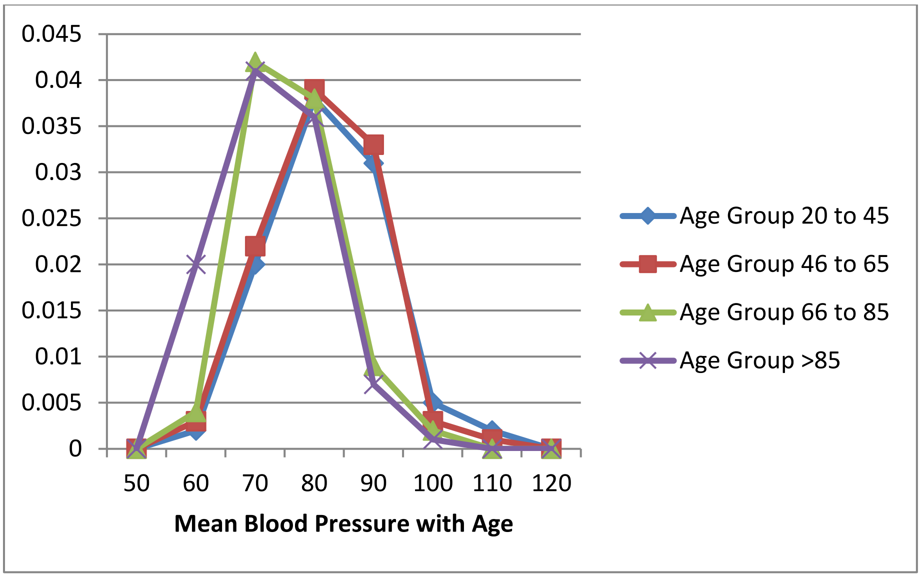


**Fig 2:** Distribution of Mean Blood Pressure with Age

The DDOC refers to the fact that the data distribution of the same category is often different between different domains. No matter which domain the features come from, the features of the same category should be close to each other, and the features of different categories should be far apart. Therefore, it is necessary to perform category adaptation on the basis of domain adaptation.As illustrated in Figure 1, the age of patient A and patient C are similar, and the data distribution is also similar. However, the survival results of the two are different and belong to different classes, so class adaptation is needed to increase the distance between the high-level features of the two.

The difficulty caused by the diversity and complexity of time series data refers to the fact that various physiological indicators of patients constitute different channels with different data distributions, and the time series trends of different channels together describe the physiological status of patients. A deep learning model can only perform domain adaptation after understanding the complex temporal dependencies between different time steps and effectively extracting high-level features.

This paper proposes an unsupervised domain adaptation method Deep Learning-based Domain adaptation marginal Softmax aware additive Cosine Margin loss prediction (DDSCM) based on domain adversarial and additive cosine margin loss. Domain adversarial is a method similar to generative adversarial network, which can solve the overall data distribution difficulty in shifting. The concept of metric learning is introduced via the additive cosine interval loss, which can solve the difficulty caused by the shift of data distribution between categories. In addition, this paper uses a bidirectional long-short-range network based on attention approach as a feature extractor to cope with the diversity and complexity of time-series data.

## 2 Related Work

The problem of unsupervised domain adaptation involves using data and labels of the source domain and the data of the target domain to train a deep learning model, hoping that the model can achieve the highest possible accuracy on target domain. In contrast to many other transfer learning strategies [9-13], domain adaptation does not require labels on the target domain, which further reduces the pressure of obtaining labeled data.

The deep learning model can be simply regarded as two parts: feature extractor and classifier. If the feature extractor of the deep learning model can extract domain invariant features from the data between different domains, then the classifier trained on the source domain can be well suited to the target domain. Domain-invariant features refer to features that are expressive and discriminative in both source and destination domains, and contain knowledge that can be shared between source and destination domains. There are two typical methods for obtaining domain-invariant features and reducing the offset of the total data distribution:

1) Feature mapping method: The distance constraint is imposed on the high-level features extracted from the source and target domain respectively by the

deep learning model, so that the high-level features that the neural network has learned are distributed similarly. Examples include DDC (deep domain confusion) [14], DAN (deep adaptation network) [15] and other methods use the maximum mean difference to measure the distribution difference between high-level features, Deep CORAL [16] method uses CORAL distance to measure high-level features distribution difference between.

2) Domain adversarial method:The concept of the generative adversarial network is utilized, andthe domain discriminator is introduced to identifyif the advanced features learned by the deep learning model belong to the source domain or the target domain. The feature extractor and the domain discriminator are balanced by adversarial training. When the domain discriminator cannot distinguish which domain the feature comes from, it means that the feature extractor has extracted the feature with domain invariance. Example includes Adversarial Discriminative Domain Adaptation [17], Domain Adversarial Neural Networks [18] etc.

Recently, domain adversarial methods have attracted much attention due to their excellent performance. To minimize the data distribution shift between categories and further improve the effect of unsupervised domain adaptation, some works have introduced the idea of metric learning based on domain adversarial methods. For example, Reference [19] and Reference[20] introduced triplet loss in the domain adaptation task to minimize intra-class distance and maximize inter-class distance to a certain extent.

However, the calculation of triplet loss needs to traverse a huge number of sample pairs, which enhances the amount of extra calculation, and it is necessary to select the appropriate size of the hidden layer features is used as the optimization object of triple loss, which increases the burden of adjusting hyperparameters.

## 3 Proposed Work

To solve the difficulties encountered in applying domain adaptation methods to mortality risk prediction tasks and the deficiencies of related work, this paper presents an unsupervised domain adaptation method DDSCM based on domain adversarial and additive cosine margin loss. This strategy utilizes labelled data from the source domain and unlabeled data from the target domain for training to increase the model's accuracy in the target domain when sample labels are not available. The method is mainly composed of feature extractor G, domain discriminator D and additive cosine interval loss classifier C. Its architecture is shown in Figure 3. The data flow of source domain and destination domain is represented by arrows with solid and dashed lines, respectively.

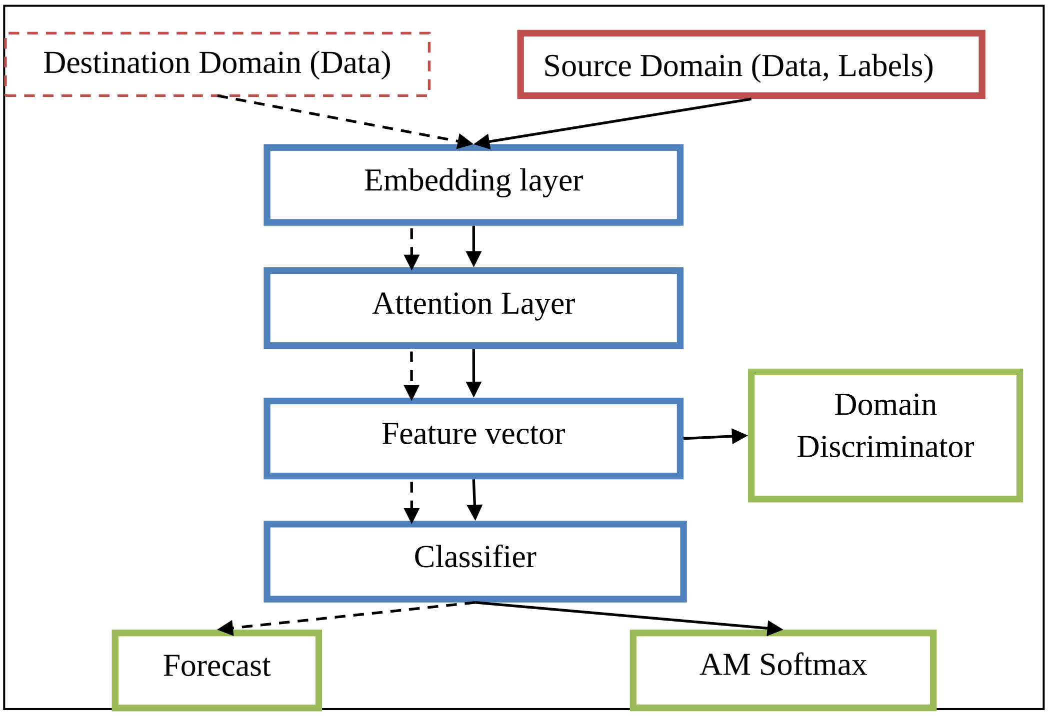


**Fig 3:**Overall Structure of the Proposed Method

### 3.1 Problem Definition

The purpose of this study is to apply an unsupervised domain adaptation method to the task of predicting mortality risk in intensive care patients. Various medical equipments in the intensive care unit records various vital signs of patients at regular intervals, and these records can be naturally regarded as time series data.

**Definition 1**:- Unsupervised domain adaptation task on time series data. Given a source domain with ns labeled samples $D_s = \{(x_i^s, y_i^s)\}_{i=0}^{n_s}$ and a destination domain with $n_t$ unlabeled samples $D_t = \{x_i^t\}_{i=0}^{n_t}$, where $x_i^s, x_i^t \in R^{m\times d}$ refers to the Time series data, d refers to the length of time series data, m refers to the dimension of each time step data. Unsupervised domain adaptation seeks to reduce empirical risk on the target domainwhen $D_s \neq D_t$.

### 3.2 Feature Extraction

The feature extractor is responsible for extracting effective high-level features from the input data. To cope with the difficulties brought by the complexity and diversity of time series data, this paper selects a bidirectional long and short-term memory network with an attention mechanism as the feature extractor. Among them, the bidirectional long short term memory (BiLSTM) network is used as the embedding layer to initially obtain the input features and capture basic time series information. The embedding layer transforms the input $x \in R^{m\times d}$ into the output $H \in R^{u\times d}$, that is, the feature dimension of each time step changes from m to u, and contains the required information.

To better extract time series information, this paper uses self-attention mechanism [21] to calculate the attention value $a_i, i = 1, 2, \cdots, d$ for each of the time interval, and then according to the attention value all time intervals are weighted and summed. The attention mechanism pay more attention to crucial time steps, thusthedeep learning model can extract more expressive features.

Denote $W_1 \in R^{n_a\times u}$, $W_2 \in R^{r\times n_a}$ as the parameter matrix, $n_a$ as the hidden layer vector dimension for calculating attention, and r refers to the count of attention heads. The Softmax operation is applied on each row vector to make the sum of the attention values at each time step equal to 1. The attention matrix $A \in R^{\mathrm{r}\times\mathrm{d}}$ is calculated as

$$A = Softmax(W_2 \tanh(W_1 H)) \qquad (1)$$

At last, the output of the attention layer is the output of the entire feature extractor $G(x) \in R^{\mathrm{r}\times\mathrm{u}}$, which is expressed as

$$M = AH^T \qquad (2)$$

### 3.3 Domain Confrontation

Domain discriminator performs domain adaptation in the form of domain confrontation, learn domain-invariant features, and try to solve the problem of overall data distribution shift. Domain adversarial draws on the idea of generative adversarial network, which makes the feature extractor and the domain discriminator, compete with each other. When the domain discriminator is unable todetermineif the feature comes from the source or the target domain, the feature extractor learns how to extract domain-invariant features.

Denoting the probability distributions w.r.t. the source and the destination domains as $p(X_s)$ and $p(X_t)$ respectively, the optimization objective w.r.t.the domain discriminator D can be expressed as

$$\max_D V(D) = E_{x^t \sim p(X_t)}\left[LogD\left(G(x^t)\right)\right] + E_{x^s \sim p(X_s)}[\log(1 - D(G(x^s)))] \qquad (3)$$

The optimization objective w.r.t. the feature extractor G can be expressedas

$$\min_G V(G) = E_{x^t \sim p(X_t)}\left[LogD\left(G(x^t)\right)\right] + E_{x^s \sim p(X_s)}[\log(1 - D(G(x^s)))] \qquad (4)$$

The optimization objectives G and D can be combined and written in the minimax optimization form:

$$\min_G \max_D V(D, G) = E_{x^t \sim p(X_t)}\left[LogD\left(G(x^t)\right)\right] + E_{x^s \sim p(X_s)}[\log(1 - D(G(x^s)))] \qquad (5)$$

Both D and G are deep learning models, which are often tuned using gradient descent to reduce the loss function. Denoting the model parameters of the D and G as $\theta_G$ and $\theta_D$, the loss function $L_{disc}(\theta_G, \theta_D)$ of the domain discriminator can be written as

$$L_{disc}(\theta_G, \theta_D) = -\frac{1}{n_s}\sum_{i=1}^{n_s}\log(1 - D(G(x^s))) - \frac{1}{n_t}\sum_{i=1}^{n_t} LogD\left(G(x^t)\right) \qquad (6)$$

In the process of adversarial training, the optimization of the D and G is carried out alternately, which is formally expressed as

$$\hat{\theta}_D = \arg\min_{\theta_D} L_{disc}(\theta_G, \theta_D), \hat{\theta}_G = \arg\min_{\theta_G} L_{disc}(\theta_G, \theta_D) \qquad (7)$$

Domain adaptation in a domain adversarial manner can take advantage of the powerful ability of generative adversarial networks to fit data distributions and better extract domain-invariant features.

### 3.4 Additive Cosine Margin Loss Classifier

The concept of metric learning is introduced by the additive margin softmax (AM-Softmax) loss, which can enhance the distinguishability between samples of different classes. As the final classification loss function, it can concurrentlydecrease the intra-class distance and increase the inter-class distance end-to-end, and does not need to spend time to select which layer of features in the model as the optimization target. Compared with the triple loss function, it does not need to calculate the distance between the sample pairs additionally, which saves the time required for training. In addition, optimizing the intra-class distance and inter-class distance end-to-end in the angle space canproducesuperior results than optimizing the hidden layer vector with triplet loss.

Denote $n$ as the count of training samples in the current batch, $y_i$ as the class label of the sample $x_i$, and there are c classes in total. II (•) is an indicative function, which is 1 when the expression in the brackets is true, and 0 when the expression is false. $p(j|x_i)$is the probability that the sample $x_i$ given by the model belongs to the jth class. The Softmax loss function $L_S$ can be expressed as

$$L_s = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} \mathrm{II}(y_i = j)\log p(j|x_i) \quad (8)$$

Consider the input of the sample $x_i$ to the last layer in the deep learning model as$f_i$, W is the weight matrix of the last layer, and $W_j$ is the row vector of the corresponding output category j w.r.t. weight matrix. Omitting bias terms, $Softmax\ loss\ function$ is further written as

$$L_s = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} \mathrm{II}(y_i = j)\log \frac{e^{W_j^T f_i}}{\sum_{j=1}^{c} e^{W_j^T f_i}} = -\frac{1}{n}\sum_{i=1}^{n} log \frac{e^{W_j^T f_i}}{\sum_{j=1}^{c} e^{W_j^T f_i}} \quad (9)$$

### 3.5 Training Process

The part of the proposed method contains trainable parameters is the feature extractor, the domain discriminator and the classifier, whose parameters are denoted as $\theta_G$, $\theta_C$, $\theta_D$ respectively. The final loss function $L(\theta_G,\ \theta_C,\ \theta_D)$ can be obtained from equations (6) and (9):

$$L(\theta_G, \theta_C, \theta_D) = L_{AMS}(\theta_G, \theta_C) - \lambda L_{disc}(\theta_G, \theta_D) \quad (10)$$

Where λ is the balance factor,it is mainly used to adjust the ratio of $L_{AMS}$and $L_{disc}$.

The detailed training process of the proposed method is shown in the DDSCM algorithm. Firstly, the parameters of feature extractor, domain discriminator and classifier are randomly initialized. During the training process, these parameters are optimized alternately in the form of gradient descent. In this paper, the Adam optimizer is used to complete the task of gradient descent. In each iteration of adversarial training, to enable the domain discriminator to better guide the feature extractor to generate domain-invariant features, it is necessary to increase the number of updates of the domain discriminator, that is, after the domain discriminator updates N disc times, the feature extractor and classifier are only updated once. The update of domain discriminator refers to updating the parameters w.r.t. domain discriminator through back-propagation after calculating$L_{disc}(\theta_G, \theta_D)$. The update of the feature extractor and the classifier is also similar to calculate the respective loss functions and then update the parameters through back-propagation. When loss function converges, the trained model is obtained. At this time, the data w.r.t. target domain is input into model, and final predicted value is obtained.

**Algorithm 1:**DDSCM based on domain adversarial and additive cosine margin loss.

**Input:** Labeled source domain data $D_s = \{(x_i^s, y_i^s)\}_{i=0}^{n_s}$ unlabeled destination domain data$D_t = \{(x_i^t)\}_{i=0}^{n_t}$, balance factor λ, feature extractor, domain discriminator and classifier the parameters $\theta_G$, $\theta_C$, $\theta_D$ the number of updates of the domain discriminator in each iteration$N_{disc}$, the optimizer Adam for gradient descent;

**Output:** The predicted value of the target domain data$\{y_i^t\}_{i=0}^{n_t} = 0$ , the parameters of the optimized feature extractor, domain discriminator and classifier$\ddot{\theta}_G, \ddot{\theta}_D, \ddot{\theta}_C$.

$Step$ 1: $Randomly\ initialize\ \theta_G, \theta_C, \theta_D$;
$Step$ 2: $repeat$
$Step$ 3: $for\ i\ =\ 1,2,?, N_{disc}\ do$
$Step$ 4: $Calculate\ \ L_{disc}(\theta_G, \theta_D)\ according\ to\ formula$ (6);
$Step$ 5: $\theta D\ \leftarrow Adam\ (\nabla_{\theta_D} L_{disc}(\theta_G, \theta_D))$ ;
$Step$ 6: $end\ for$

$Step\ 7$: $Calculate\ L(\theta_G, \theta_C, \theta_D)\ according\ to\ formula\ (10)$;

$Step\ 8$: $\theta_C \leftarrow Adam\ (\nabla_{\theta_D} L(\theta_G, \theta_C, \theta_D))$ ;

$Step\ 9$: $\theta_G \leftarrow Adam\ (\nabla_{\theta_D} L(\theta_G, \theta_C, \theta_D))$ ;

$Step\ 10$: $until\ the\ model\ parameters\ converge$

$Step\ 11$: $\ddot{\theta}_G, \ddot{\theta}_D, \ddot{\theta}_C \leftarrow \theta_G, \theta_C, \theta_D$;

$Step\ 12$: $\{y_i^t\}_{i=0}^{n_t} \leftarrow \{C(G(x_i^t))\}_{i=0}^{n_t}$;

$Step\ 13$: $return\ \{y_i^t\}_{i=0}^{n_t}, \ddot{\theta}_G, \ddot{\theta}_D, \ddot{\theta}_C$

## 4 Experiment

In this paper, the MIMIC-III dataset [22] is used for experiments. The MIMIC-III dataset is a public clinical database maintained by the Massachusetts Institute of Technology, containing about 60,000 hospitalization records between 2001 and 2016, each record including demographic characteristics, medical intervention records, imaging reports, vital signs records, nursing records and other information [2]. Generally speaking, the condition of patients within 48 hours after entering the intensive care unit is more critical, so this paper selects the data within 48 hours after the patient enters the intensive care unit to predict the patient's survival outcome.

According to the work presented in Reference[2], this paper extracts 76-dimensional features from the MIMIC-III data set, including heart rate, diastolic and systolic blood pressure, SpO2, capillary refill rate and other 60-dimensional continuous features, 12-dimensional discrete features such as Glasgow coma scale, and 4-dimensional constants about patient information. After data cleaning and preprocessing, the final input data has 48 time steps, and each time step has 76-dimensional features. Reference [23] attempted transfer learning between acute hypoxic respiratory failure patients of different ages. This proposed method follows the Reference [23] experimental settings, and divides all the patients in the ICU database of the MIMIC-III dataset into 4 groups namely Youth, Middle-aged, Elderlyand the Old Ageaccording to their age, as prescribed in Table 1:

**Table 1:**The Division of Different Domains of MIMIC-III Dataset

| Domain Name | Age Group | Number of People |
|---|---|---|
| Youth | 20~45 | 2763 |
| Middle-aged | 46~65 | 6963 |
| Elderly | 66~85 | 9100 |
| Oldage | >85 | 2313 |

Since the ratiow.r.t. the positive samples and the negative samples in the data set is quite different, and it belongs to a binary classification problem, to avoid the influence of the imbalance of positive &negative samples on the evaluation index, this experiment uses the area value under the ROC curve (receiver operating characteristic curve) (area under curve, AUC) as the evaluation standard. In this paper, five methods are used to compare with the proposed DDSCM method:

1) DAN [15]: The features are extracted using the BiLSTM network combined with the self-attention mechanism, and the transfer learning method DAN is adopted based on the feature map. In this method, the maximum mean difference is used to measure the dissimilarity amongthe characteristic distribution of the source domain and the characteristic distribution of the destination domain.
2) CORAL [16]: Features are extracted using a BiLSTM network combined with a self-attention mechanism, and a feature map-based transfer learning method Deep CORAL [16] is used. This method uses CORAL distance to measure the dissimilarity among the source domain feature distribution and the destination domain feature distribution.
3) BiLSTM [17]: A BiLSTM network combined with a self-attention mechanism is used as the baseline, trained on the source domain and tested on the destination domain, without using any unsupervised domain adaptation learning method.
4) ADA [18]: The same feature extractor is used as the methods described in (1)to (3), and the domain

adversarial method is used, using the Softmax loss function.

5) Tri-ADA [19]: The same feature extractor is used as the methods described in 1 to 4. A domain adversarial approach is used. And on this basis, a triple loss function is added to solve the problem of data distribution shift between categories.

In this paper, the unsupervised domain adaptation task is carried out in pairs between four different age groups, that is, between four domains. The experimental results are depicted in Table 2 along with Figure 4~7. For example, the data of young patients is taken as the source domain, the data of middle-aged patients is taken as the destination domain, and the unsupervised domain adaptation task is recorded as young → middle-aged.

The DDSCM method proposed in this paper achieves the highest AUC values in 10 of the 12 unsupervised domain adaptation tasks, indicating the effectiveness of the method. The BiLSTM method does not use any unsupervised domain adaptation method and thus performs poorly. For domains that are far apart, the performance of the BiLSTM method drops significantly. For example, for the task middle-aged → young, the AUC value of the BiLSTM method is 0.867, while for the task middle-aged → elderly, the AUC value of the BiLSTM method drops to 0.754. Domains that are far apart mean that the ages are far apart, and the distribution of the data is more significant, so it has a more impact on the accuracy of the model.

The CORAL method and the DAN method use the feature map-based migration method to solve the problem of global data distribution differences. It can be seen from the results that these two methods have a certain improvement compared with the BiLSTM method. Compared with the method based on feature mapping, the ADA method introduces domain confrontation, which can better reduce the global data distribution difference, so the effect is better. Tri-ADA performs domain adaptation in the form of domain adversarial, and adds triple loss to reduce the data distribution difference between classes. Compared with the CORAL and DAN methods, the experimental results have a certain improvement. In order to align the data distribution between categories more finely, the DDSCM method proposed in this paper introduces an additive cosine margin loss, which is compared with the ADA method and the Tri-ADA method. The accuracy has been further improved, which shows the effectiveness of the proposed method.

In order to directly reflect the superiority of the proposed method, the three methods of BiLSTM, Tri-ADA and DDSCM are trained separately, and the output features of the last layer of the classifier of each method are projected to the angle space for visualization. The data of young patients was selected as the source domain, and the data of elderly patients was selected as the destination domain. One of the reasons for the decreased accuracy of the BiLSTM method in the target domain is the distribution shift between categories. When training in the source domain, there are clear boundaries between features of different categories. However, decision boundary is not wide enough, and classification error due to the distribution shift in the target domain test.

**Table 2:** Results of Mortality Risk Prediction Based on Unsupervised Domain Adaptation

| Source Domain→ Destination Domain | BiLSTM | CORAL | DAN | ADA | Tri-ADA | DDSCM |
|---|---|---|---|---|---|---|
| Youth → Middle Age | 0.803 | 0.822 | 0.832 | 0.834 | 0.831 | 0.842 |
| Young → Elderly | 0.781 | 0.794 | 0.781 | 0.790 | 0.795 | 0.808 |
| Youth → Old Age | 0.756 | 0.763 | 0.759 | 0.766 | 0.765 | 0.781 |
| Middle-Age→ Youth | 0.867 | 0.874 | 0.879 | 0.883 | 0.880 | 0.889 |
| Middle Age → Elderly | 0.808 | 0.825 | 0.828 | 0.831 | 0.835 | 0.835 |
| Middle Age → Old Age | 0.754 | 0.766 | 0.779 | 0.784 | 0.780 | 0.793 |
| Elderly → Young | 0.845 | 0.863 | 0.870 | 0.875 | 0.873 | 0.868 |
| Elderly → Middle Age | 0.832 | 0.848 | 0.853 | 0.857 | 0.858 | 0.862 |
| Elderly → Old Age | 0.763 | 0.792 | 0.794 | 0.797 | 0.795 | 0.807 |
| Old Age → Young | 0.826 | 0.841 | 0.840 | 0.856 | 0.861 | 0.856 |
| Old Age → Middle Age | 0.804 | 0.801 | 0.796 | 0.799 | 0.812 | 0.819 |
| Old Age → Elderly | 0.775 | 0.802 | 0.799 | 0.805 | 0.803 | 0.815 |

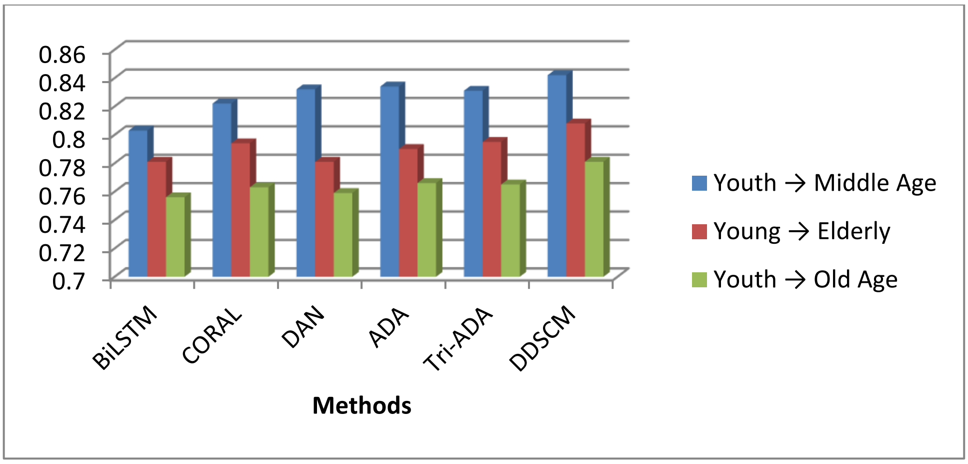


**Fig 4:**Comparisons Young to other Aged Group with Different Methods

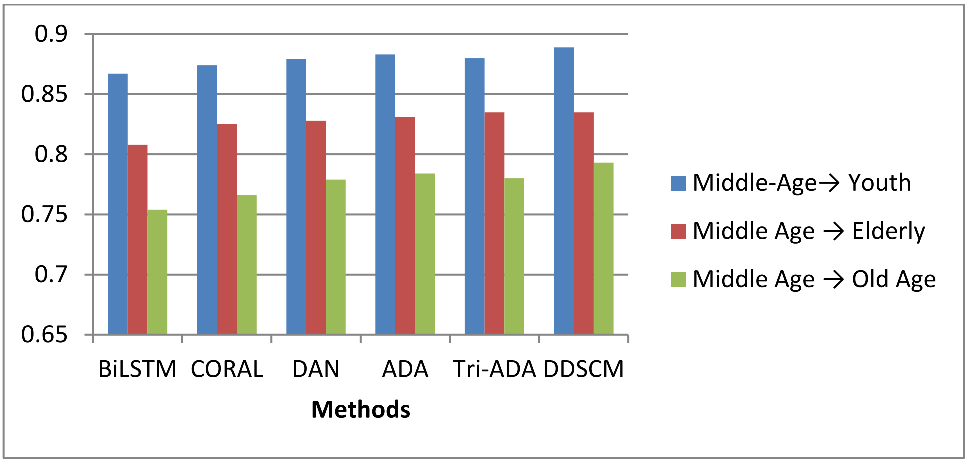


**Fig 5:**Comparisons Middle-Aged to other Aged Group with Different Methods

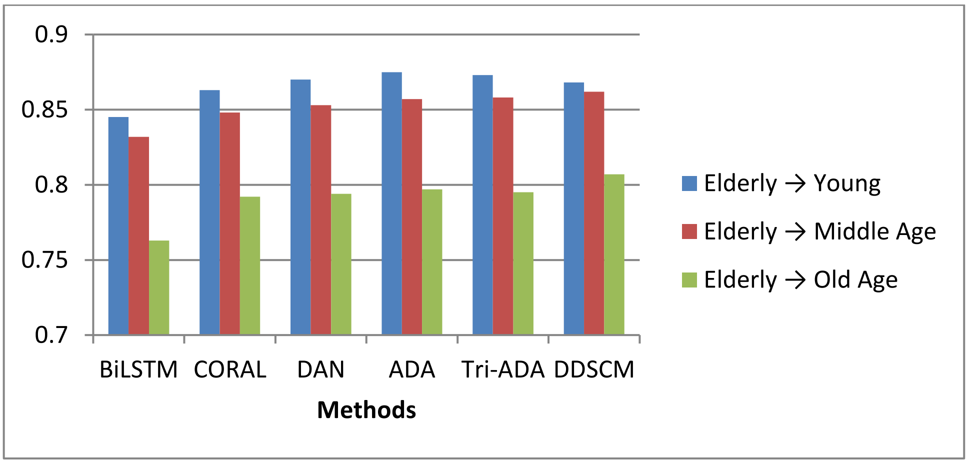


**Fig 6:** Comparisons Elderly to other Aged Group with Different Methods

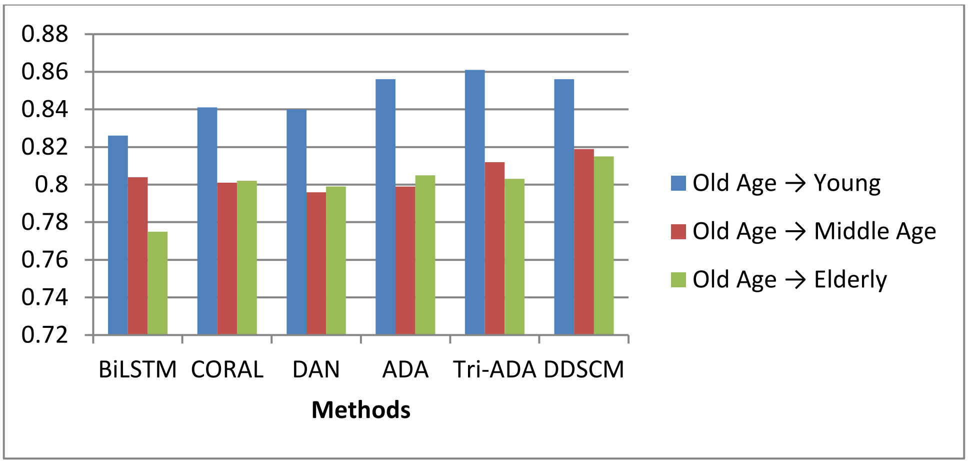


**Fig 7:**Comparisons Old Age to other Aged Group with Different Methods

Therefore, the model should explicitly increase the decision boundary to maintain intra-class compactness and inter-class separability. The Tri-ADA method increases the inter-class distance in the form of triple loss in the source domain training, so the separability between different categories of features is strengthened in the target domain test, thereby reducing the error rate.The AM-Softmax loss function is introduced into the DDSCM method, which can further increase the width of the decision boundary in the angle space. The overlap between the features of the surviving patients and the features of the dead patients is further reduced, and good inter-class separability and intra-class compactness are achieved. Benefiting from the wider decision boundary, the classifier trained on the source domain is less sensitive to class shifts, so it can achieve better accuracy when tested on the destination domain.

## 5 Conclusion

In this paper, DDSCM based on domain adversarial and additive cosine margin loss is proposed to deal with these problems. This paper reduces the data distribution shift between the overall data in the form of domain adversarial. To further improve the effect of unsupervised domain adaptation, the idea of metric learning is introduced to reduce the data distribution shift between classes in the form of minimizing the additive cosine margin loss. The proposed method is validated on the task of predicting death risk in intensive care patients, and the experimental results and visual analysis results on the MIMIC-III dataset demonstrate the effectiveness of the method. Future work will try to extend the proposed method to other tasks in the medical field, such as disease prediction and length of stay prediction tasks.

## References

[1] S. Hussein, P. Kandel, C. W. Bolan, M. B. Wallace and U. Bagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches," in IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1777-1787, Aug. 2019, doi: 10.1109/TMI.2019.2894349.

[2] N. Abid et al., "Burnt Forest Estimation from Sentinel-2 Imagery of Australia using Unsupervised Deep Learning," 2021 Digital Image Computing: Techniques and Applications (DICTA), 2021, pp. 01-08, doi: 10.1109/DICTA52665.2021.9647174.

[3] N. Rajpurohit, A. Agarwalla and J. S. Bhatt, "A Sure-Based Unsupervised Deep Learning Method for SarDespeckling," 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 5119-5122, doi: 10.1109/IGARSS47720.2021.9553918.

[4] Y. Hui, L. Du, S. Lin, Y. Qu and D. Cao, "Extraction and Classification of TCM Medical Records Based on BERT and Bi-LSTM With Attention Mechanism," 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2020, pp. 1626-1631, doi: 10.1109/BIBM49941.2020.9313359.

[5] T. Jain, R. Sharma and R. Malhotra, "Handwriting Recognition for Medical Prescriptions using a CNN-Bi-LSTM Model," 2021 6th International Conference for

Convergence in Technology (I2CT), 2021, pp. 1-4, doi: 10.1109/I2CT51068.2021.9418153.

[6] Q. -J. Lv et al., "A Multi-Task Group Bi-LSTM Networks Application on Electrocardiogram Classification," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 8, pp. 1-11, 2020, Art no. 1900111, doi: 10.1109/JTEHM.2019.2952610.

[7] Y. Ren, T. Zhang, X. Liu and H. Lin, "End-to-end Answer Selection via Attention-Based Bi-LSTM Network," 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), 2018, pp. 264-265, doi: 10.1109/HOTICN.2018.8606015.

[8] H. Zhao, L. Deng and Y. Xie, "A Training Strategy for Enhancing the Accuracy of Real-Time Tumor Tracking Based on Deep Bi-LSTM Learning," 2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE), 2019, pp. 1-4, doi: 10.1109/ICMIPE47306.2019.9098226.

[9] E. Dong, Y. Qiao and Z. Zhang, "Application of Difficult Sample Mining based on Cosine Loss in Face Recognition," 2020 IEEE International Conference on Mechatronics and Automation (ICMA), 2020, pp. 1052-1057, doi: 10.1109/ICMA49215.2020.9233852.

[10] R. Li, N. Li, D. Tuo, M. Yu, D. Su and D. Yu, "Boundary Discriminative Large Margin Cosine Loss for Text-independent Speaker Verification," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 6321-6325, doi: 10.1109/ICASSP.2019.8682749.

[11] R. Ji, J. Cao, X. Cai and B. Xu, "Max Margin Cosine Loss for Speaker Identification on Short Utterances," 2018 11th International Symposium on Chinese Spoken Language Processing (ISCSLP), 2018, pp. 304-308, doi: 10.1109/ISCSLP.2018.8706654.

[12] D. Zhong and J. Zhu, "Centralized Large Margin Cosine Loss for Open-Set Deep Palmprint Recognition," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 6, pp. 1559-1568, June 2020, doi: 10.1109/TCSVT.2019.2904283.

[13] H. Wang et al., "CosFace: Large Margin Cosine Loss for Deep Face Recognition," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5265-5274, doi: 10.1109/CVPR.2018.00552.

[14] C. Gui and J. Hu, "Unsupervised Domain Adaptation by regularizing Softmax Activation," 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 397-402, doi: 10.1109/ICPR.2018.8545224.

[15] Z. -H. Liu, B. -L. Lu, H. -L. Wei, L. Chen, X. -H. Li and M. Rätsch, "Deep Adversarial Domain Adaptation Model for Bearing Fault Diagnosis," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 7, pp. 4217-4226, July 2021, doi: 10.1109/TSMC.2019.2932000.

[16] B. Gholami, O. Rudovic and V. Pavlovic, "PUnDA: Probabilistic Unsupervised Domain Adaptation for Knowledge Transfer Across Visual Categories," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3601-3610, doi: 10.1109/ICCV.2017.387.

[17] J. Wu, S. Liao, Z. lei, X. Wang, Y. Yang and S. Z. Li, "Clustering and Dynamic Sampling Based Unsupervised Domain Adaptation for Person Re-Identification," 2019 IEEE International Conference on Multimedia and Expo (ICME), 2019, pp. 886-891, doi: 10.1109/ICME.2019.00157.

[18] M. A. Kabir and X. Luo, "Unsupervised Learning for Network Flow Based Anomaly Detection in the Era of Deep Learning," 2020 IEEE Sixth International Conference on Big Data Computing Service and Applications (BigDataService), 2020, pp. 165-168, doi: 10.1109/BigDataService49289.2020.00032.

[19] H. Yan, P. Yu and D. Long, "Study on Deep Unsupervised Learning Optimization Algorithm Based on Cloud Computing," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2019, pp. 679-681, doi: 10.1109/ICITBS.2019.00168.

[20] H. -X. Yu, A. Wu and W. -S. Zheng, "Unsupervised Person Re-Identification by Deep Asymmetric Metric Embedding," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 956-973, 1 April 2020, doi: 10.1109/TPAMI.2018.2886878.

[21] B. Zhou, D. Su and Z. Qu, "Medical Text Classification System Based on Deep

Learning," 2021 International Conference on Intelligent Computing, Automation and Applications (ICAA), 2021, pp. 388-392, doi: 10.1109/ICAA53760.2021.00077.

[22] X. Hu and Q. Yuan, "Epileptic EEG Identification Based on Deep Bi-LSTM Network," 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT), 2019, pp. 63-66, doi: 10.1109/ICAIT.2019.8935899.

[23] B. Liu et al., "Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 10051-10060, doi: 10.1109/ICCV.2019.01015.

[24] N. I. Kajla, M. M. S. Missen, M. M. Luqman, M. Coustaty, A. Mehmood and G. S. Choi, "Additive Angular Margin Loss in Deep Graph Neural Network Classifier for Learning Graph Edit Distance," in IEEE Access, vol. 8, pp. 201752-201761, 2020, doi: 10.1109/ACCESS.2020.3035886.

[25] H. Ren, M. El-Khamy and J. Lee, "Stereo Disparity Estimation via Joint Supervised, Unsupervised, and Weakly Supervised Learning," 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 2760-2764, doi: 10.1109/ICIP40778.2020.9191126.

[26] P. Du, X. lin, X. Pi and X. wang, "An Unsupervised Learning Algorithm for Deep Recurrent Spiking Neural Networks," 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2020, pp. 0603-0607, doi: 10.1109/UEMCON51285.2020.9298074.

[27] K. Yu, "Deep Learning for Unsupervised Neural Machine Translation," 2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE), 2021, pp. 614-617, doi: 10.1109/ICBASE53849.2021.00121.

[28] Y. Xu, B. Liu, Y. Quan and H. Ji, "Unsupervised Deep Background Matting Using Deep Matte Prior," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 7, pp. 4324-4337, July 2022, doi: 10.1109/TCSVT.2021.3132461.

[29] Q. Liu, G. Luan, F. Wang, B. Peng and D. Hu, "Multi-focus Image Fusion Algorithm Based on Unsupervised Deep Learning," 2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI), 2021, pp. 362-366, doi: 10.1109/CISAI54367.2021.00076.

[30] Jaruwatcharaset, C. . (2023). Effects of Using a Temperature Control System in Bandicota indica Stalls with Internet of Things Technology. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 166–170. https://doi.org/10.17762/ijritcc.v11i4s.6324

[31] Yathiraju, D. . (2022). Blockchain Based 5g Heterogeneous Networks Using Privacy Federated Learning with Internet of Things. Research Journal of Computer Systems and Engineering, 3(1), 21–28. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/37

[32] Sai Pandraju, T.K., Samal, S., Saravanakumar, R., Yaseen, S.M., Nandal, R., Dhabliya, D. Advanced metering infrastructure for low voltage distribution system in smart grid based monitoring applications (2022) Sustainable Computing: Informatics and Systems, 35, art. no. 100691, .