A Feature-based Transfer Learning to Improve the Image Classification with Support Vector Machine

Nina Sevani[1], Kurniawati Azizah[2], Wisnu Jatmiko[3]
Faculty of Engineering and Computer Science, Krida Wacana Christian University, Jakarta, Indonesia[1]
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia[1, 2, 3]

***Abstract*—In the big data era there are some issues regarding real-world classification problems. Some of the important challenges that still need to be overcome to produce an accurate classification model are the data imbalance, difficulties in labeling process, and differences on data distribution. Most classification problems are related to the differences in the data distribution and the lack of labels on some datasets while other datasets have abundant labels. To address the problem, this paper proposes a weighted-based feature-transfer learning (WbFTL) method to transfer knowledge between different but related domains, called cross-domain. The knowledge transfer is done through making a new feature representations in order to reduce the cross-domain's distribution differences while maintaining the local structure of the domain. To make the new feature representation we implement a feature selection and inter-cluster class distance. We propose two stages of the feature selection process to capture the knowledge of the feature and its relation to the label. The first stage uses a threshold to select the feature. The second stage uses ANOVA (Analysis of Variance) to select features that are significant to the label. To enhance the accuracy, the selected features are weighted before being used for the training process using SVM. The proposed WbFTL are compared to 1-NN and PCA as baseline 1 and baseline 2. Both baseline models represent the traditional machine learning and dimensionality reduction method, without implementing transfer learning. It is also compared with TCA, the first feature-transfer learning work on this same task, as baseline 3. The experiment results of 12 cross-domain tasks on Office and Caltech dataset show that the proposed WbFTL can increase the average accuracy by 15.25%, 6.83%, and 3.59% compared to baseline 1, baseline 2, and baseline 3, respectively.**

Keywords*—*Feature-transfer learning; image; feature selection; weight; distance

I. Introduction

In this big data era, the use of machine learning is growing and expanding into various purposes and uses, including image classification. The success rate of machine learning in image classification is generally determined by the accuracy value. Although nowadays there are already many public datasets [1], some challenges associated with image classification still exist. The first challenge arises because there are some unlabeled or limited labels of datasets [2] [3] [4]. Meanwhile, on the other side there are many other datasets with a very abundant labels. The second challenge is related with the labeling process which needs much effort to make a label for the dataset [3] [4] [5] [6]. Whereas the availability of the labeled data training determines the success of the classification model [7]. The third challenge is associated with the limited ability of the traditional machine learning method, which requires the training and inference data come from the same dataset that has the same distribution. Even though this ability is needed to produce an accurate classification model [2] [3] [8]. Unfortunately, traditional machine learning methods such as: NN (Nearest Neighbor) and PCA (Principal Component Analysis) do not show good results on the third challenge [9].

To overcome the challenges, we can make a classification model by using knowledge from different but related datasets (domain) [4] [10] [11] [12]. The method is called transfer learning, which is an extension of traditional machine learning. The different but related dataset (domain) used in transfer learning is often called cross-domain. The use of cross-domain terms indicates the existence of the source domain (D_S) and the target domain (D_T). The original idea of transfer learning is to utilize the knowledge from the labeled domain often called the source domain, to predict the correct label for an unlabeled domain often called the target domain. Before transferring the knowledge, we need to conduct the similarity measurement between the cross-domain to avoid negative transfer. Negative transfer is when the classification model trained using a combination of the cross-domain gives poorer performance than the one trained using the source domain only. Generally, the similarity measurement in the cross-domain is performed using Maximum Mean Discrepancy (MMD).

There are many transfer learning approaches, such as feature-based, instance-based, parameter-based, and relational-based transfer learnings [3] [10] [11]. Our work focuses on feature-transfer learning, considering this approach is mostly used for image domain [12] [13] [14] [15] [16] [17]. Research on the feature-transfer learning began with the discovery of the Transfer Component Analysis (TCA) method which only focuses on overcoming marginal distribution differences in cross-domains [13]. After TCA, several other feature-transfer learning methods were found, such as Geodesic Flow Kernel (GFK) [16], Joint Distribution Adaptation (JDA) [14], Subspace Alignment (SA) [17], Transfer Joint Matching (TJM) [15], and Balanced Distribution Adaptation (BDA) [9]. Moreover, the success key for knowledge transfer in the cross-domain can be done by focusing on the instances in D_T and conducting the features matching to minimize data distribution difference [18]. Feature-based also can be improved the classification accuracy when it is used with the classifier like SVM [19] [20] [21]. However, one of the weakness of SVM is restricted the data precision and requires computational cost, so it needs an additional step, such as features reduction to minimize cost [22] [23]. Many techniques can be applied in feature-transfer learning, for example, the use of the Grassman manifold as the geometric property[16], adding balance unit parameter to overcome the data imbalance problem[9], or the formation of subspace features to minimize the distribution difference[17]. In general, the previous feature-transfer learning method works by adapting the dimensional reduction approach and generating an adaptation matrix which is a projection of the cross-domain's features. These techniques need an iteration process to get the best result and add some parameters in the model, where the parameters have to be tuned up to give optimal results. These previous feature-transfer learning methods need large computational requirements due to the complex process and the use of many parameters. CORAL [24] proposed simpler feature-transfer learning without using many parameters. It used the second-order statistical approach to overcome the distribution differences in the cross-domain. However, this method did not consider the label information contained in the source domain. Though some important information can also be obtained from the label.

Therefore, in this paper we propose a simple feature-transfer learning method without using many parameters while still adopting a dimension-reduction approach and utilize the label information from the source domain. We name this proposed approach as Weighted-based Feature-Transfer Learning (WbFTL). The dimension reduction in the method is more towards the formation of new features subsets through the implementation of features selection techniques. Implementing the features selection has several benefits. First, it can reduce the search space and generate significant features for the classification [25]. Second, the features selection also has a simpler way of working without the need to do features projection of the cross-domain, thus retaining the original form of the features and maintaining the explicit meaning of the selected features [26]. Lastly, the features selection method in the classification problem also can enhance the classification results [27] [28] [29] [30]. To optimize the model, the proposed feature-transfer learning adds some weight to the selected features and also use the closest distance between instances to the center of the class label. These combination techniques allow the proposed method to make a new features representation that can minimize the distribution differences between cross-domain while still maintaining the local structure of each domain to get an efficient and accurate classification model.

The experiment showed WbFTL increasing the accuracy by 15.25%, 6.83%, and 3.59%, respectively, against the 1NN, PCA, and TCA baseline models. WbFTL is also superior to the previous feature transfer method and provides a higher average accuracy than those of GFK, JDA, SA, and TJM. Compared to BDA, WbFTL has a competitive accuracy. However, our proposed WbFTL is more superior in the parameters used than BDA. Unlike BDA, WbFTL supports minimal use of parameters that can be run on limited machine resources.

Overall, the contribution of our work can be summarized as follows:

1) WbFTL is an easy feature-transfer learning approach that can overcome the distribution difference in the cross-domain without using many parameters in the training and inferencing process.

2) WbFTL is a novel feature-transfer learning method that uses feature selection as the strategy to make transformation of the features. Combined with the feature weighting, the experiment results show that the proposed method get better results compared to other feature-transfer learning methods.

3) WbFTL also the first feature-transfer learning method that utilize the label information in the source domains in the transformation process.

II. METHODS

Machine learning has two important components: domain and task. In transfer learning we have source domain, source task, target domain, and target task. The source domain or D_S contains source features space ($\mathcal{X}_S$) and marginal probability ($P(x_S)$), written as $D_S = \{\mathcal{X}_S, P(x_S)\}$, where $x_S \in \mathcal{X}_S$. The source task or $\mathcal{T}_S$, consists label space $\mathcal{Y}_S$ and classification function $f(\cdot)$ to determine the label for instances in D_T based on the knowledge from D_S and D_T. The source task can be written as $T_S = \{\mathcal{Y}_S, f(\cdot)\}$. Same as the source domain and task, the target domain, and target task can be written as $D_T = \{\mathcal{X}_T, P(X_T)\}$ and $T_T = \{\mathcal{Y}_T, f(\cdot)\}$.

The WbFTL method is included in the category of transductive transfer learning. In this category, only instances in D_S have the label, meanwhile D_T are unlabeled. There are 800 features in each D_S and D_T, and the number of instances on D_S and D_T can be different. The goal of the WbFTL is to do feature transformation and make a new feature representation that represents the cross-domain, D_S and D_T. The overview of WbFTL in carrying out the transformation can be seen in Fig. 1.

From Fig. 1 it can be seen that the WbFTL method emphasizes the problem of distribution equalization in the cross-domains. Distribution equalization is conducted by making a new features representation that reflects both domains. The process of equalizing this domain begins by measuring the similarity between domains, followed by features selection and features weighting to create a method that is cost effective while still being able to produce a good accuracy. WbFTL also overcomes the conditional distribution differences by utilizing the label information in the features transformation process.

As depicted in Fig. 1 below, the new features representation should minimize the distribution difference between D_S and D_T, while still preserving the local structure of the domain itself. The new features representation formed will be used for the training and inference process with SVM, as depicted in Fig. 2. Before being processed by SVM, the data used will be converted into a features vector using statistical calculations [31].

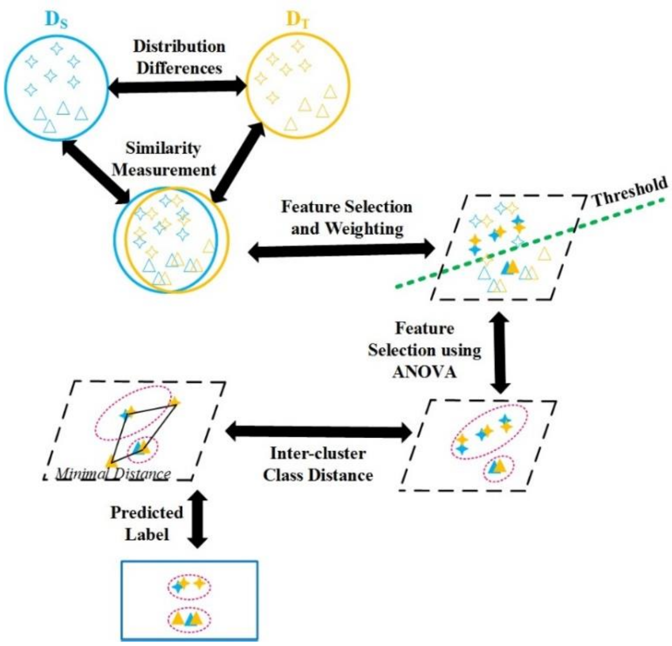

Fig. 1. The overview of WbFTL

There are three steps of features transformation in WbFTL as shown in Fig. 3. The first step is features selection using a threshold to select the features. The second transformation step is the features selection using ANOVA. The third step is only transforming the features in D_T, by calculating the minimum distance between instances in D_T to the class label of D_S.

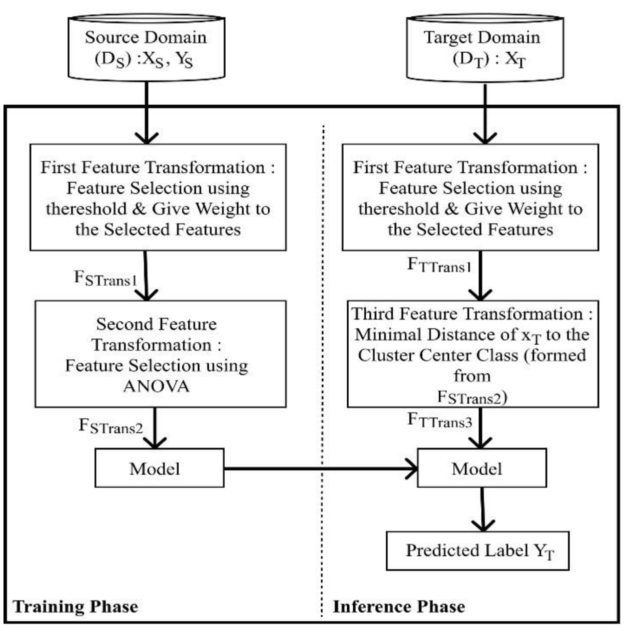

Fig. 2. The training and inference process in the WbFTL.

The features selection method is one of the strategies to form a new features representation besides features mapping, features clustering, features encoding, features alignment, and features augmentation [32]. This strategy can preserve the local and important structure of the domain, besides, also reducing the distribution difference of the cross-domain.

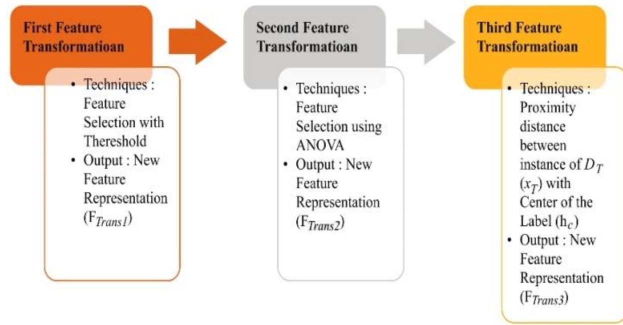

Fig. 3. Feature transformation steps in WbFTL.

Before doing the features transformation, there is a similarity measurement process between D_S and D_T using MMD, as a non-parametric method [9] [13] [14] [15] [16] [18] [32]. In the non-parametric method, the measurement is done using approximation distribution value, because it is difficult to get the real distribution value. In the proposed feature-transfer learning, MMD is used in conjunction with the kernel which will map the original value of the features space of each domain to a new features representation using a mapping function. Our proposed method uses a distance function as the mapping function, that is the Euclidean Distance and Reproducing Kernel Hilbert Space (RKHS) [13] [14] [15] [16] [18] [32]. The use of MMD and kernel makes it sufficient to calculate the similarity of the cross-domain using density estimation. The density estimation is done using average features values in the features space $\mathcal{X}_S, \mathcal{X}_T$. The formulation of the mapping function on feature-transfer learning can be seen in (1).

$$\boldsymbol{Sim}(D_S, D_T) = \|\boldsymbol{F}_{Si} - \boldsymbol{F}_{Ti}\|_{\mathcal{H}}, \quad (1)$$

where $\boldsymbol{Sim}(D_S, D_T)$ is the similarity measurement result using MMD and RKHS in the cross-domain, $\boldsymbol{F}_{Si}, \boldsymbol{F}_{Ti}$ is the vector of the mean value of the $i-$ features in D_S and D_T, sequentially.

The new features space formed from MMD and kernel will be the input for the first step of features transformation. The shape of the features space resulting from the application of MMD and the kernel can be seen in (2).

$$\boldsymbol{F} = \{f_1, f_2, \ldots, f_n\} \in R^n, \quad (2)$$

where $\boldsymbol{F}$, is the new features space which is the result of D_S and D_T features mapping, n is the amount of the original features, that is 800 features.

A. Features Selection and Weighting

Feature selection is the process of generating a feature subset based on its relevance and redundancy [33]. The purpose of feature selection is to select the right features in order to get a better understanding of the characteristics of the data. Therefore, selecting significant features will assist the model in studying the data and producing the right label [23]. Feature selection can also enhance classification accuracy [29] [30]. Moreover, feature selection can also be used to reduce the dimensionality while maintaining the local structure of the dataset and reducing the model complexity.

WbFTL applies the feature selection technique in the first and second steps of the feature transformation process. The two steps of feature selection in WbFTL use a statistical-based approach, i.e., average and varians.

1) Feature selection using thresholds and feature weighting: The first step in feature transformation is adopting the filter method. The filter method can work faster and simpler in the implementation and does not depend on the classifier used [34] [35]. The average value will be used as the threshold value, which serves as a stopping criteria. Features with a value under the threshold will be selected because they are considered similar or significant features between D_S and D_T. The feature subset that is formed can be written as in (3)

$$\boldsymbol{A} = \{a_1, a_2, \dots, a_k\} \in R^k \tag{3}$$

where $\boldsymbol{A}$ is the feature subset and k is the amount of the selected features, with $k < n$.

Previous feature-transfer learning also reveals that feature weighting can enhance classification accuracy [27] [36]. Feature weighting is the generalization of feature selection [37] [38]. Therefore, the selected features in the feature subset will be weighted according to their degree of similarity. The smaller the feature value, the greater the weight given. The weight value describes the level of similarity. The formula for the feature weight can be seen in (4). This weighting method is similar to Fisher's criteria, which are used in various cases.

$$w(a_i) = \frac{\frac{1}{k}\sum_{i=1}^{k} a_i}{a_i}, \tag{4}$$

where $w(a_i)$ is the weight for the $i-$features.

The first feature transformation is generated as a dot product between the selected features in (3) and the weight according to formula (4). The dot product can be written as in formula (5) and the representation of the first feature transformation can be written in (6).

$$b_i = w(a_i) \times f_i, \tag{5}$$

$$\boldsymbol{F_{Trans1}} = \{b_1, b_2, \dots, b_k\} \in R^k \tag{6}$$

where b_i is the first features transformation for the $i-$feature, $w(a_i)$ is the weight for the $i-$feature, f_i is the original value for the $i-$feature, and $\boldsymbol{F_{Trans1}}$ is the first feature transformation.

2) Feature selection using ANOVA: The disadvantage of the first-step feature transformation above is that it does not involve label information. Therefore, to select the features that are significant to the label, the second feature transformation is carried out using ANOVA. The ANOVA technique uses variants, a statistical property, to select the features. Previous research shows that the use of ANOVA with SVM gives good accuracy in image classification [28] [39] [40].

The second step of feature transformation in WbFTL applies ANOVA to select features. The first feature transformation ($\boldsymbol{F_{Trans1}}$) above became the input for the second feature transformation ($\boldsymbol{F_{Trans2}}$), which is used for the training process. The second feature transformation has the same value as the first feature transformation, only different in the number of features. The representation of the second feature transformation can be seen in (7).

$$\boldsymbol{F_{Trans2}} = \{b_1, b_2, \dots, b_a\} \in R^a \tag{7}$$

where $\boldsymbol{F_{Trans2}}$ is the second feature transformation that uses ANOVA, and a is the selected features from the implementation of ANOVA with $a < k$.

B. Inter-Cluster Class Distance

The third feature transformation is done by calculating the distance from instances of D_T to the center of the class label. The Euclidean distance will be used for the distance calculation. The third feature transformation adopts the gravity law, which is also similar to the general works of classification [41]. The use of distance has also been widely used in feature-transfer learning, such as for determining the weight proportion [8], determining the similarity between cross-domain [13] [14] [15], and the implementation of metric learning [32].

The input for the third feature transformation ($\boldsymbol{F_{Trans3}}$) comes from the first feature transformation ($\boldsymbol{F_{Trans1}}$). The formula to calculate the distance of each instances in D_T is written in (8). While the formula to get the third feature transformation can be seen in (9). The representation of the the third feature transformation written on (10).

$$d_i = \frac{x_T \times h_C}{Dist(x_T, h_C)^2} \tag{8}$$

$$z_i = d_i \times \boldsymbol{F_{Trans1}}, \tag{9}$$

$$\boldsymbol{F_{Trans3}} = \{z_1, z_2, \dots, z_k\} \in R^k \tag{10}$$

where d_i is the distance between the $i-$instance of $\boldsymbol{x_T}$ and the cluster center h_C. z_i is the third feature transformation for the $i-$ feature of $\boldsymbol{x_T}$. $\boldsymbol{F_{Trans3}}$ is the third feature transformation which is only applied to D_T, h_C is the center of the cluster class label calculated by average formula. $Dist(x_T, h_C)$ is the euclidean distance between $\boldsymbol{x_T}$ (instances in D_T) and th_C.

C. Dataset and Experimental Setup

The dataset for the proposed feature-transfer learning was taken from the image domain, which is the real-world object category. There are 10 class labels in each dataset: calculator, laptop, keyboard, mouse, monitor, video projector, headphones, backpack, mug, and bike [13] [14] [15] [16]. We use four datasets in the experiment. A detailed description of each dataset is shown in Table I. All the datasets used already implemented SURF as the feature descriptor. The features descriptor algorithm will extract the superior degree of the pixels in the original image so that it can capture stable features from each image [39]. Fig. 4 is an example of some images from the class label monitor, backpack, mug, and mouse in each dataset.

TABLE I. DESCRIPTION OF THE DATASET USED

Dataset	Instances #	Features #	Class #	Domain
Office-Amazon	958	800	10	A
Office-Webcam	295	800	10	W
Office-DSLR	157	800	10	D
Caltech-256	1123	800	10	C

Fig. 4. Example of dataset used.

Different from previous research on feature-transfer learning, which uses many parameters in its classification model, this proposed method does not use parameters. The only parameter needed in the proposed feature-transfer learning is the C parameter in the classifier SVM. The C value is set to 0.001 with a linear kernel. This value setting follows previous feature-transfer learning research [17]. Moreover, the proposed feature-transfer learning in this paper has a simpler feature transformation process. By using feature selection and feature weighting approaches and employing statistic properties like average and variance, the proposed method can be done with limited resources while still providing good accuracy results.

III. RESULT

The experiment result of the WbFTL will be compared with three baselines and five previous feature-transfer learning methods, namely: Geodesic Flow Kernel (GFK) [16], Joint Distribution Analysis (JDA) [14], Transfer Joint Matching (TJM) [15], Balanced Distribution Adaptation (BDA) [9], and Subspace Alignment (SA) [17]. All the datasets used in WbFTL were also used in the comparison methods, including the use of SURF as the feature descriptor.

A. *Comparison with the Baselines and Previous Feature-Transfer Learning Methods*

The difference between WbFTL and previous feature-transfer learning methods mainly lies in the feature transformation process performed, the optimization process, and the use of pseudolabels, as shown in Fig. 5. The red box indicates the feature transformation steps. Fig. 5(a) is the previous feature-transfer learning method. Meanwhile, Fig. 5(b) shows the steps of WbFTL. Even though both of the methods use a dimensionality reduction approach, WbFTL chooses feature selection rather than forming a projection matrix as used in previous methods. WbFTL also uses feature selection and feature weighting as optimalization processes. This approach makes WbFTL simpler and more cost-effective. Moreover, WbFTL does not need to go through an iterative process to find a stable pseudolabel that will be considered the predicted label.

All the methods will be compared based on their accuracy values. The accuracy values from the baselines and the previous feature-transfer learning methods will be taken from the original paper for each method. The comparison results for the accuracy of each pair of datasets can be seen in Table II. Meanwhile, Fig. 6 shows the comparison graph between WbFTL and the baselines model.

This research used three baselines, and the results of the baselines gained from the original paper [9]:

- Baseline 1: 1NN as the representation of the traditional machine learning method
- Baseline 2: PCA as the representation of the dimensional reduction method
- Baseline 3: TCA as the first feature-transfer learning method

There is no feature transfer or knowledge transfer in the implementation of baseline 1 and baseline 2.

In addition, the experiments also compare the average accuracy from 12 pairs of datasets between the WbFTL method, the baselines, and the comparison methods. The comparison results of the average accuracy are shown in Table III, and the visualization will be shown in graphical form in Fig. 7 and Fig. 8. From the experiment results in Table III, it can be seen that the WbFTL gets 46.62% for the average accuracy. This value overcomes the baselines and previous feature transfer learning methods, except for BDA. Although the average accuracy value of the WbFTL is comparable with the average accuracy of the BDA, the WbFTL uses fewer parameters, as shown in Table IV. So, it is possible to use it with limited resources.

Table II shows 12 pairs of datasets formed from four datasets: A, W, D, and C. These twelve pairs of datasets are formed by swapping the positions of the datasets that will become D_S and D_T. For example, the A→W means dataset A becomes D_S and dataset W becomes D_T. Another example is W→A which means dataset W becomes D_S and dataset A becomes D_T. Because the principle of transfer learning is that there are no definite rules on which datasets must become D_S or D_T. The best value for each dataset pair is written in bold, and the second best value is written in underline. From the value yield, we can see that WbFTL gives a better accuracy value on A→C, C→A, and C→W, where dataset C becomes one of the processed domains.

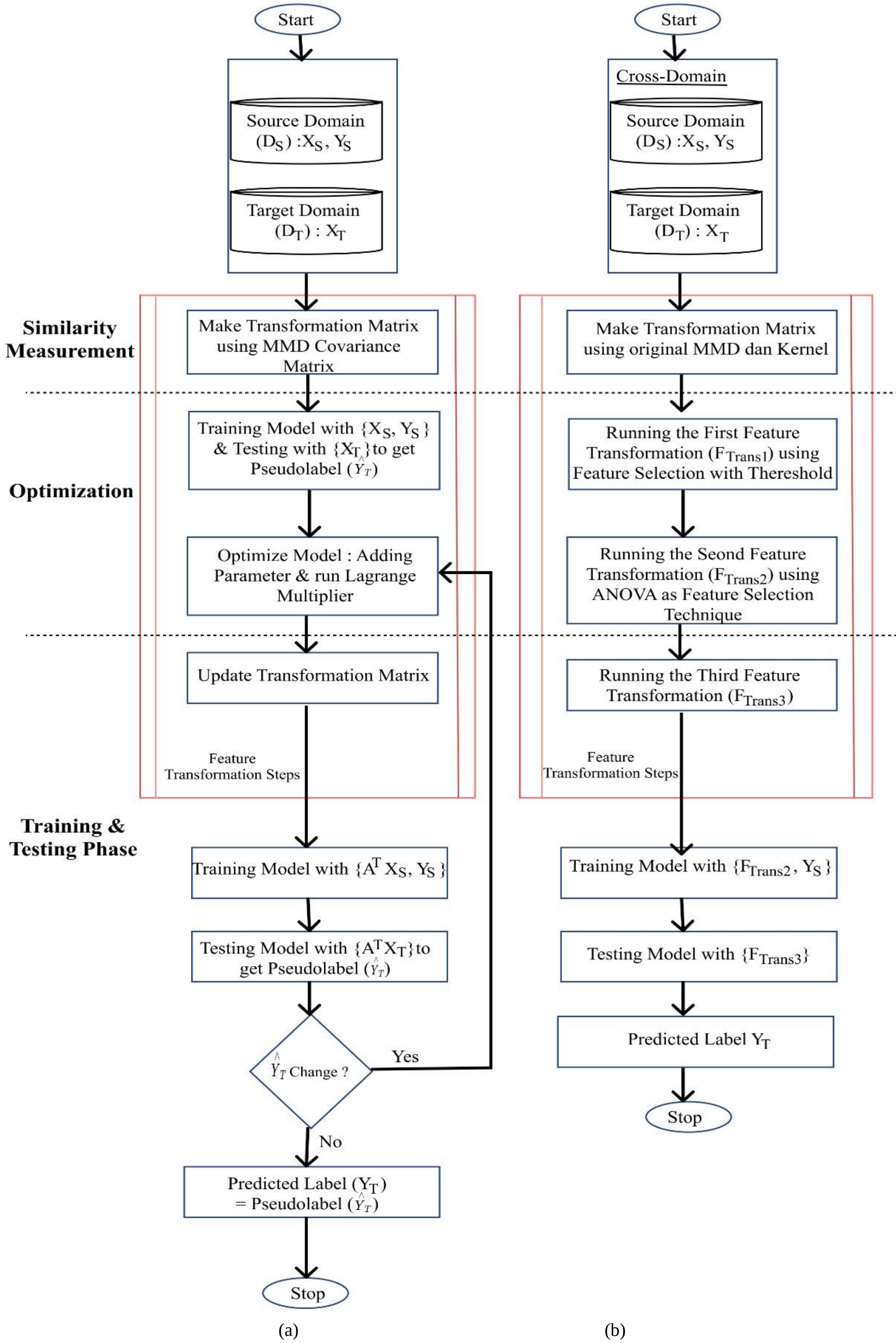

Fig. 5. Comparison method between previous feature-transfer learning and WbFTL.

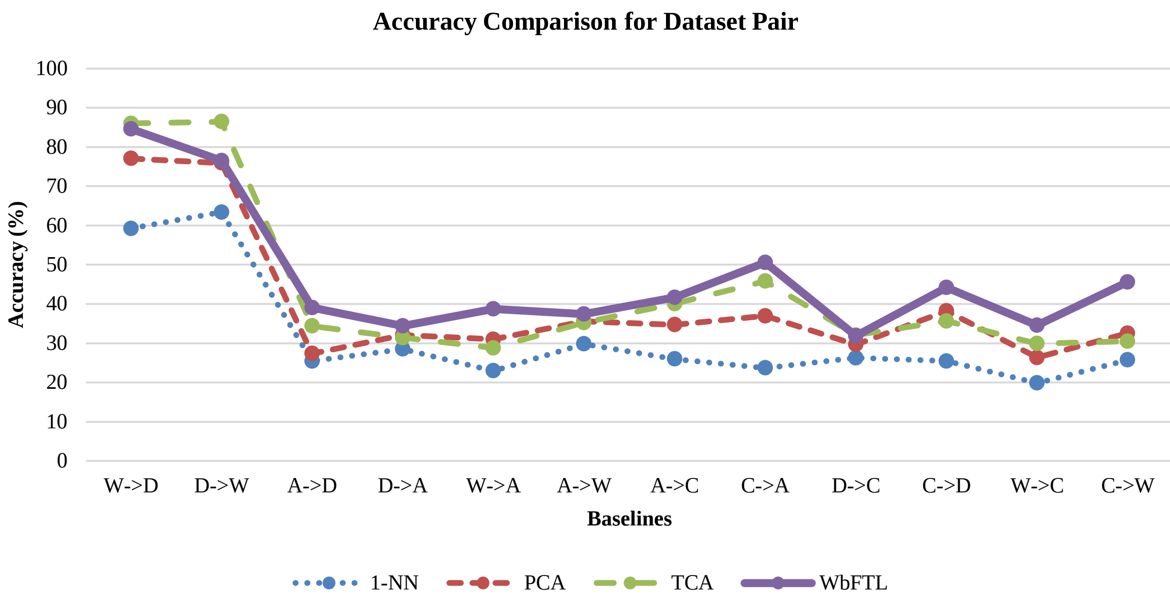

Fig. 6. Accuracy comparison between baselines and WbFTL for dataset pair.

TABLE II. ACCURACY COMPARISON PER DATASET PAIR

Dataset	**Baselines**			**Previous Feature-Transfer Learning Methods**					**Proposed Method**
	1NN	***PCA***	***TCA***	***GFK***	***SA***	***JDA***	***TJM***	***BDA***	***WbFTL***
W→D	59.24	77.07	85.99	80.89	75.16	89.17	89.17	**91.72**	84.62
D→W	63.39	75.93	86.44	75.59	76.95	89.49	85.42	**91.86**	76.53
A→D	25.48	27.39	34.39	36.31	33.76	39.49	**45.22**	43.31	39.1
D→A	28.50	32.05	31.42	32.05	**39.87**	33.09	32.78	33.09	34.38
W→A	22.96	31	28.81	29.75	**39.25**	32.78	29.96	32.99	38.77
A→W	29.83	35.59	35.25	38.98	33.22	37.97	**42.03**	32.99	37.41
A→C	26	34.73	40.07	40.25	39.98	39.36	39.45	40.78	**41.71**
C→A	23.7	36.95	45.82	41.02	49.27	44.78	46.76	44.89	**50.57**
D→C	26.27	29.65	32.06	30.28	**34.55**	31.52	31.43	32.5	31.91
C→D	25.48	38.22	35.67	38.85	39.49	45.22	44.59	**47.77**	44.23
W→C	19.86	26.36	29.92	30.72	**37.17**	31.17	30.19	28.94	34.58
C→W	25.76	32.54	30.51	40.68	40	41.69	39.98	38.64	**45.58**

TABLE III. AVERAGE ACCURACY COMPARISON FOR EACH METHOD

Methods	**Baselines**			**Previous Feature-Transfer Learning Methods**					**Proposed Method**
	1NN	***PCA***	***TCA***	***GFK***	***SA***	***JDA***	***TJM***	***BDA***	***WbFTL***
Average Accuracy	31.37	39.79	43.03	42.95	44.89	46.31	46.4	46.62	46.62

TABLE IV. PARAMETER COMPARISON

Methods	***TCA***	***GFK***	***SA***	***JDA***	***TJM***	***BDA***	***WbFTL***
Parameters	μ	d	β_1, β_2	λ	λ	λ, μ	None

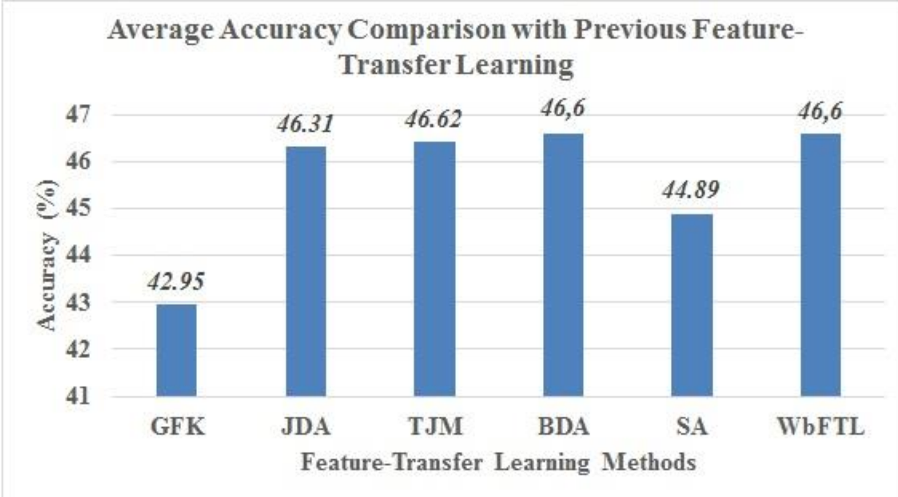

Fig. 7. Average accuracy comparison between WbFTL and previous feature-transfer learning.

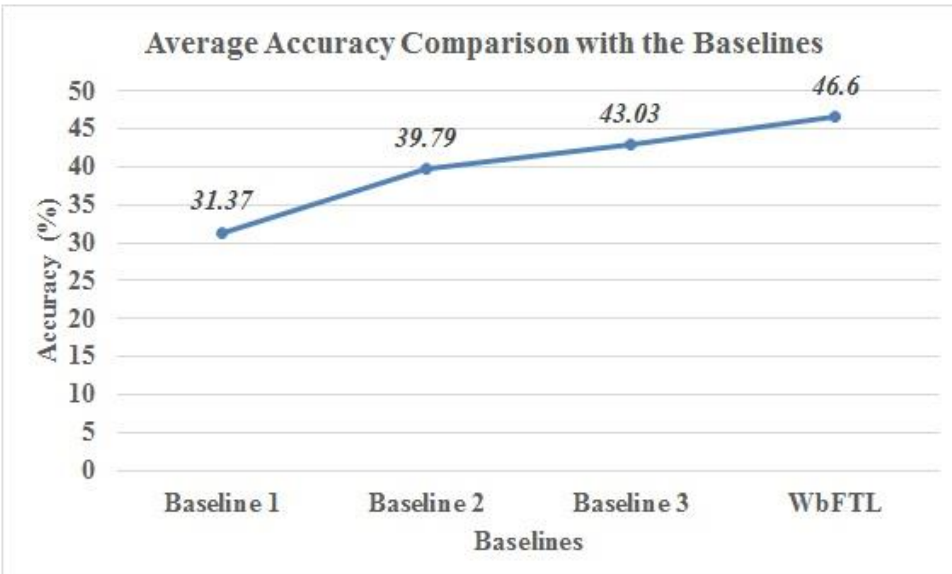

Fig. 8. Average accuracy comparison between WbFTL and the baselines.

Based on the accuracy value in Table III, we can see that the average accuracy of WbFTL exceeds the average value of all baselines. This value also shows an increase in accuracy of 15.25%, 6.83%, and 3.59% when compared to baseline 1, baseline 2, and baseline 3. This result also shows that the implementation of feature-transfer learning has proven to improve image classification accuracy across domains. Even when compared to using a simple classification model such as 1NN, which is baseline 1, WbFTL is far better at generating predictive labels. Compared to PCA, the commonly used dimensionality reduction method, WbFTL also provides better accuracy results. Although WbFTL also uses a dimensionality reduction approach, the process used in WbFTL is simpler than PCA. WbFTL does not take iterative steps to generate a stable transformation matrix, as can be seen in Fig. 5. This result also indicates that feature selection can be applied to form new feature representations to produce a better classification model without applying dimension reduction as in baseline 1. Feature selection as a form of dimensionality reduction has also proven to be applicable to transform features and produce good classification models at a cost-effective rate.

When compared with the previous transfer learning methods, such as TCA, which is baseline 3, it is also seen that WbFTL provides an increase in yield of 3.59%. This result indicates that the feature transformation process in WbFTL is able to minimize more differences in data distribution compared to that carried out in baseline 3. This is because TCA works to reduce marginal distribution differences without considering label information. While WbFTL reduces the difference distributions between cross-domains by utilizing label information through the second and third transformations.

The experiment results prove that the distribution difference in the cross-domain can be reduced by making a new feature representation that reflects both domains. The formation of this new feature representation can be done by using a dimensionality reduction approach to reduce the searching space. Concerning the implementation of dimensional reduction techniques, the feature selection method is proven to be implemented and gives good results for transfer learning cases.

The use of the feature selection method makes the WbFTL simpler compared to other feature-transfer learning methods, which have to make a projection matrix. Other than being complex, the use of the projection matrix in the previous feature-transfer learning method also requires optimization, which involves many parameters to be tuned, requiring more resources. A comparison of the number of parameters used in the previous feature-transfer learning method with the WbFTL can be seen in Table IV.

B. Ablation Study

The ablation study will show accuracy for each step of feature transformation, which is conducted in WbFTL. In this research, the ablation study will compare the accuracy of the second feature transformation with the third feature transformation. The first feature transformation will not be seen in particular because it uses the same approach as the second feature transformation, namely feature selection. By looking at the accuracy of the second feature transformation, it already includes the first transformation. The result of the ablation study can be seen in Table V. Where $\boldsymbol{F_{Trans2}}$ is the accuracy using only feature selection to transform features, and $\boldsymbol{F_{Trans3}}$ is the final accuracy value after carrying out all stages of feature transformation.

TABLE V. ABLATION STUDY FOR DATASET PAIR

Dataset	F_{Trans2}	F_{Trans3}
W→D	84.62	84.62
D→W	76.53	76.53
A→D	39.10	39.10
D→A	33.86	34.38
W→A	38.87	38.77
A→W	36.73	37.41
A→C	41.44	41.71
C→A	50.68	50.57
D→C	31.91	31.91
C→D	44.23	44.23
W→C	34.49	34.58
C→W	46.26	45.58

Based on the ablation study result in Table V above, we can see that the third feature transformation has a better result than the second feature transformation, although it is not showing a significant difference. When viewed from the 12 pairs of the existing datasets, the increase in accuracy values occurs in pairs D→A, A→W, A→C, and W→C. While several dataset pairs experience a very small decrease in accuracy after doing the third feature transformation, namely at W→A, C→A, and C→W.

By looking at these results, it can be seen that the majority of degradation was found when the Amazon dataset became D_T. Accuracy values also tend to decrease when the Caltech dataset becomes D_S. The biggest decrease was in C→W, which decreased by 0.68.

Meanwhile, when the Amazon became D_S, the accuracy value tended to improve at each step of the feature transformation performed, as shown in A→W and A→C. The increase in accuracy is also seen when the dataset that becomes D_T has more instances than the number of D_S, as shown in D→A, A→C, and W→C. The biggest increase in A→W was 0.68.

C. Discussion

The experimental results in Table V above show that the classification results are better when using large datasets as D_S. This can be seen from the accuracy value, which tends to be high when Amazon becomes D_S. Similar results are also shown in Table II. In Table II, the accuracy value of A-->D is 39.1%; this value is greater than the accuracy of D-->A, which is only 34.38%. As mentioned in Table I, the Office-Amazon dataset is larger than the Office-DSLR. Other dataset pairs, such as C-->A, also have 9% higher accuracy than A-->C because the Caltech-256 dataset is larger than Office-Amazon.

In addition, the imbalance of instances between D_S and D_T also affects the accuracy value of each pair of datasets. Better accuracy is obtained on pairs of datasets with an almost equal number of instances.As shown in Table II, where the highest accuracy is D-->W and W-->D of 76.53% and 84.62%, respectively. The pair with the next highest accuracy is C-->A at 41.71% and A-->C at 50.57%. In this case, the number of instances between the Office-DSLR and Office-Webcam datasets is more evenly matched than the number of instances between Office-Amazon and Caltech-256. Conversely, when the number of instances of the two domains is very different, the accuracy value will deteriorate, as shown by the D-->C pair, which only has an accuracy of 31.91%.

The condition of the original image in the dataset pair used also affects the accuracy value. In Fig. 4 above, it can be seen that the original images between the Office-Amazon and Caltech-256 datasets are quite different in terms of background color, objects in the image (original and cartoon objects), and lighting. This background color difference will affect the value of the resulting feature extraction results, so it can affect the level of image similarity. So the accuracy value between the Office-Amazon and Caltech-256 dataset pairs tends to be small, only in the range of 40%–50%. Meanwhile, the Office-DSLR and Office-Webcam datasets have more similar original image conditions, resulting in higher accuracy in the range of 75%–85%. Given that the original image conditions are almost the same, the feature extraction values will be more similar, so the resulting accuracy will also be better.

IV. CONCLUSION

In summary, we have proposed a simple feature-transfer learning method called WbFTL. The proposed method is more efficient than the previously reported feature-transfer learning methods because it employs a feature selection strategy for making a new feature representation. In addition, the WbFTL involved weighting on the selected features to improve the accuracy. The proposed method is also simple since it utilizes statistical properties such as averages and variance to transform the features. There are three steps of feature transformation in the proposed feature-transfer learning method. The first step was feature selection which used a threshold obtained from the feature averaging value as the stopping criteria. The second step was feature selection using the ANOVA technique. Finally, the third step was calculation of the distance between the target domain and the center of the class label employing the Euclidean distance.

This experiment was carried out using only one type of classifier, namely SVM, with SURF as a feature extraction technique. There are still many other types of classifiers that can be used. Another limitation in the experiment is the category of dataset used, which is limited to real-world objects only.

From the experiment result using 12 pairs of the dataset, we have shown that the WbFTL provides a better result than the previous method, except for the BDA. Although the WbFTL gives a comparable result to the BDA, it is superior in terms of simplicity because it does not use any parameters to run the model. Allowing it to use in the conditions of limited resources. We also showed that WbFTL get higher accuracy of 15.25%, 6.83%, and 3.59% when compared to 1-NN, PCA, and TCA model baselines.

We can improve the results and accuracy of the WbFTL in the future by combining it with CORAL as one of the feature transformation steps in WbFTL. Furthermore, the accuracy can also be increased by optimizing the feature transformation steps, such as optimizing the feature weighting process or applying the weighting to the distance.

REFERENCES

[1] Faculty of Informatics, International University of Rabat Technopolis parc, Sala el jadida 11100, Morocco, Y. Riahi, S. Riahi, and Department of Mathematics and Computer Science, Faculty of Sciences, University of ChouaibDoukkali Jabran Khalil JabranAvenu , El jadida 24000, Morocco, “Big Data and Big Data Analytics: concepts, types and technologies,” IJRE, vol. 5, no. 9, pp. 524–528, Nov. 2018, doi: 10.21276/ijre.2018.5.9.5.

[2] S. Niu, Y. Liu, J. Wang, and H. Song, “A Decade Survey of Transfer Learning (2010–2020),” vol. 1, no. 2, 2020.

[3] F. Zhuang et al., “A Comprehensive Survey on Transfer Learning.” arXiv, Jun. 23, 2020. Accessed: Jan. 12, 2023. [Online]. Available: http://arxiv.org/abs/1911.02685

[4] G. Csurka, “Domain Adaptation for Visual Applications: A Comprehensive Survey.” arXiv, Mar. 30, 2017. Accessed: Jan. 12, 2023. [Online]. Available: http://arxiv.org/abs/1702.05374

[5] X. Qian et al., “Generating and Sifting Pseudolabeled Samples for Improving the Performance of Remote Sensing Image Scene Classification,” IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, vol. 13, pp. 4925–4933, 2020, doi: 10.1109/JSTARS.2020.3019582.

[6] L. Song, Y. Xu, L. Zhang, B. Du, Q. Zhang, and X. Wang, “Learning From Synthetic Images via Active Pseudo-Labeling,” IEEE Trans. on Image Process., vol. 29, pp. 6452–6465, 2020, doi: 10.1109/TIP.2020.2989100.

[7] Y. Roh, G. Heo, and S. E. Whang, “A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective,” IEEE Trans. Knowl. Data Eng., vol. 33, no. 4, pp. 1328–1347, Apr. 2021, doi: 10.1109/TKDE.2019.2946162.

[8] V. Cheplygina, I. P. Pena, J. H. Pedersen, D. A. Lynch, L. Sorensen, and M. de Bruijne, “Transfer Learning for Multicenter Classification of Chronic Obstructive Pulmonary Disease,” IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1486–1496, Sep. 2018, doi: 10.1109/JBHI.2017.2769800.

[9] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, “Balanced Distribution Adaptation for Transfer Learning,” in 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA: IEEE, Nov. 2017, pp. 1129–1134. doi: 10.1109/ICDM.2017.150.

[10] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[11] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” J Big Data, vol. 3, no. 1, p. 9, Dec. 2016, doi: 10.1186/s40537-016-0043-6.

[12] O. Day and T. M. Khoshgoftaar, “A survey on heterogeneous transfer learning,” J Big Data, vol. 4, no. 1, p. 29, Dec. 2017, doi: 10.1186/s40537-017-0089-0.

[13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain Adaptation via Transfer Component Analysis,” IEEE Trans. Neural Netw., vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.

[14] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer Feature Learning with Joint Distribution Adaptation,” in 2013 IEEE International Conference on Computer Vision, Sydney, Australia: IEEE, Dec. 2013, pp. 2200–2207. doi: 10.1109/ICCV.2013.274.

[15] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, “Transfer Joint Matching for Unsupervised Domain Adaptation,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA: IEEE, Jun. 2014, pp. 1410–1417. doi: 10.1109/CVPR.2014.183.

[16] Boqing Gong, Yuan Shi, Fei Sha, and K. Grauman, “Geodesic flow kernel for unsupervised domain adaptation,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI: IEEE, Jun. 2012, pp. 2066–2073. doi: 10.1109/CVPR.2012.6247911.

[17] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, “Unsupervised Visual Domain Adaptation Using Subspace Alignment,” in 2013 IEEE International Conference on Computer Vision, Sydney, Australia: IEEE, Dec. 2013, pp. 2960–2967. doi: 10.1109/ICCV.2013.368.

[18] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, “Visual Domain Adaptation with Manifold Embedded Distribution Alignment,” in Proceedings of the 26th ACM international conference on Multimedia, Seoul Republic of Korea: ACM, Oct. 2018, pp. 402–410. doi: 10.1145/3240508.3240512.

[19] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, “Adaptation Regularization: A General Framework for Transfer Learning,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 5, pp. 1076–1089, May 2014, doi: 10.1109/TKDE.2013.111.

[20] A. Van Opbroek, H. C. Achterberg, M. W. Vernooij, and M. De Bruijne, “Transfer Learning for Image Segmentation by Combining Image Weighting and Kernel Learning,” IEEE Trans. Med. Imaging, vol. 38, no. 1, pp. 213–224, Jan. 2019, doi: 10.1109/TMI.2018.2859478.

[21] Y. Chen, J. Wang, M. Huang, and H. Yu, “Cross-position activity recognition with stratified transfer learning,” Pervasive and Mobile Computing, vol. 57, pp. 1–13, Jul. 2019, doi: 10.1016/j.pmcj.2019.04.004.

[22] Lixin Duan, I. W. Tsang, and Dong Xu, “Domain Transfer Multiple Kernel Learning,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 465–479, Mar. 2012, doi: 10.1109/TPAMI.2011.114.

[23] N. Chauhan and B.-J. Choi, “DNN-Based Brain MRI Classification Using Fuzzy Clustering and Autoencoder Features,” IJFIS, vol. 21, no. 4, pp. 349–357, Dec. 2021, doi: 10.5391/IJFIS.2021.21.4.349.

[24] B. Sun, J. Feng, and K. Saenko, “Return of Frustratingly Easy Domain Adaptation,” AAAI, vol. 30, no. 1, Mar. 2016, doi: 10.1609/aaai.v30i1.10306.

[25] W.-Y. Deng, A. Lendasse, Y.-S. Ong, I. W.-H. Tsang, L. Chen, and Q.-H. Zheng, “Domain Adaption via Feature Selection on Explicit Feature

Map," IEEE Trans. Neural Netw. Learning Syst., vol. 30, no. 4, pp. 1180–1190, Apr. 2019, doi: 10.1109/TNNLS.2018.2863240.

[26] Md. A. Hossain, X. Jia, and J. A. Benediktsson, "One-Class Oriented Feature Selection and Classification of Heterogeneous Remote Sensing Images," IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing, vol. 9, no. 4, pp. 1606–1612, Apr. 2016, doi: 10.1109/JSTARS.2015.2506268.

[27] X. Zhong, S. Guo, H. Shan, L. Gao, D. Xue, and N. Zhao, "Feature-Based Transfer Learning Based on Distribution Similarity," IEEE Access, vol. 6, pp. 35551–35557, 2018, doi: 10.1109/ACCESS.2018.2843773.

[28] Department of Statistics and Econometrics, Sofia University, Sofia, Bulgaria., B. Vrigazova*, I. Ivanov*, and Department of Statistics and Econometrics, Sofia University, Sofia, Bulgaria., "Optimization of the ANOVA Procedure for Support Vector Machines," IJRTE, vol. 8, no. 4, pp. 5160–5165, Nov. 2019, doi: 10.35940/ijrte.D7375.118419.

[29] N. A. M. Zaini and M. K. Awang, "Hybrid Feature Selection Algorithm and Ensemble Stacking for Heart Disease Prediction," International Journal of Advanced Computer Science and Applications, vol. 14, no. 2, pp. 158–165.

[30] S. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Enhanced Optimized Classification Model of Chronic Kidney Disease," International Journal of Advanced Computer Science and Applications, vol. 14, no. 2, pp. 321–331.

[31] W. Tangsuksant, M. Noda, K. Kitagawa, and C. Wada, "Viewpoint Classification for the Bus-Waiting Blinds in Congested Traffic Environment," IJFIS, vol. 19, no. 1, pp. 48–58, Mar. 2019, doi: 10.5391/IJFIS.2019.19.1.48.

[32] R. K. Sanodiya, A. Mathew, J. Mathew, and M. Khushi, "Statistical and Geometrical Alignment using Metric Learning in Domain Adaptation," in 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, United Kingdom: IEEE, Jul. 2020, pp. 1–8. doi: 10.1109/IJCNN48605.2020.9206877.

[33] C. Chen, Y. Tsai, F. Chang, and W. Lin, "Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results," Expert Systems, vol. 37, no. 5, Oct. 2020, doi: 10.1111/exsy.12553.

[34] S. Jain and A. O. Salau, "An image feature selection approach for dimensionality reduction based on kNN and SVM for AkT proteins," Cogent Engineering, vol. 6, no. 1, p. 1599537, Jan. 2019, doi: 10.1080/23311916.2019.1599537.

[35] B. Venkatesh and J. Anuradha, "A Review of Feature Selection and Its Methods," Cybernetics and Information Technologies, vol. 19, no. 1, pp. 3–26, Mar. 2019, doi: 10.2478/cait-2019-0001.

[36] S. Uguroglu and J. Carbonell, "Feature Selection for Transfer Learning," in Machine Learning and Knowledge Discovery in Databases, D. Gunopulos, T. Hofmann, D. Malerba, and M. Vazirgiannis, Eds., in Lecture Notes in Computer Science, vol. 6913. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 430–442. doi: 10.1007/978-3-642-23808-6_28.

[37] D. Singh and B. Singh, "Hybridization of feature selection and feature weighting for high dimensional data," Appl Intell, vol. 49, no. 4, pp. 1580–1596, Apr. 2019, doi: 10.1007/s10489-018-1348-2.

[38] D. Panday, R. Cordeiro de Amorim, and P. Lane, "Feature weighting as a tool for unsupervised feature selection," Information Processing Letters, vol. 129, pp. 44–52, Jan. 2018, doi: 10.1016/j.ipl.2017.09.005.

[39] S. Kim and J. Park, "Hybrid Feature Selection Method Based on Neural Networks and Cross-Validation for Liver Cancer With Microarray," IEEE Access, vol. 6, pp. 78214–78224, 2018, doi: 10.1109/ACCESS.2018.2884896.

[40] M. B. Desai, S. V. Patel, and B. Prajapati, "ANOVA and Fisher Criterion based Feature Selection for Lower Dimensional Universal Image Steganalysis," 2016.

[41] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative Transfer Joint Matching for Domain Adaptation in Hyperspectral Image Classification," IEEE Geosci. Remote Sensing Lett., vol. 16, no. 6, pp. 972–976, Jun. 2019, doi: 10.1109/LGRS.2018.2889789.