DIMENSIONALITY REDUCTION AND FILTERING APPROACH FOR IDENTIFYING PTP1-B DESCRIPTORS FOR TYPE-2 DIABETES

R.BHAGYASRI
Assistant Professor
Department of Information Technology
Sir C.R.Reddy College of Engineering, Eluru, Andhra Pradesh

ABSTRACT - In signaling pathway of insuling and leption Protein-TyrosinePhosphatase1B(PTP1B) exhibits critical negative regulation. This exhibits attractive target activity obesity and Type-2 Diabetes (T2D). In this paper proposed classification model with inclusion of PTP-1B dataset to identify the inhibators of PTP-1B through high dimensionality reduction and filtering stated as High Dimensional Correlation Filtering (HDCF). Through integration of high dimensioanlity reduction technique molecular structure of PTP-1B is estimated. The data for PTP-1B is collected from genecards database, those vast range of noisy and redundant information. Those reducdant data leads to reduced accuracy and increase the processing complexity. To reduce processing complexity high correlation filtering is adopted for elimination of noisy filter. The analysis of results expressed proposed HDCF exhibits significant performance characteristics in identification of descriptors of PTP-1B.

Keywords: Protein-tyrosinephosphatase1B, Type-2 Diabetes, High Dimensionality Reduction, High Correlation Filtering

I. INTRODUCTION

A huge volume of digital data has been developed on an ongoing basis in various application areas over the past few years. In addition, data size, heterogeneity, sophistication, and dimensionality are increasing exponentially[1]. High Dimensional Data (HDD) applications have been established in various fields, including biomedical, internet, education, medicine, industry, and social media. The vast number of new HDDs in various formats is constantly evolving (e.g., text, digital images, speech signals, and videos). High data dimensionality can pose many problems for precise classification, pattern recognition, and visualization for Machine Learning (ML) models. Due to high computational complexity, learning in high-dimensional spaces or with a large number of features can become hard [2]. The term dimensionality curse implies that if the amount of data for which a model is trained is set, then overfitting will result from increasing dimensionality. This problem can be considered in the deep learning process. The Dimensionality Reduction (DR) can be performed through feature extraction (feature transformation) and feature selection. Feature extraction transforms the original HDD sets to new reduced data sets by removing redundant and irrelevant features. The new feature set retains the full amount of original data set knowledge [3]. Feature selection selects from the input data a subset of features that are most important to the problem [4]. Manually extracting the necessary and specific characteristics for various applications becomes a tiresome activity. Choosing a suitable methodology for DR will reduce the time and effort needed to pick and extract appropriate research features.

Dimensionality Reduction Techniques (DRTs) provide an effective way to reduce the number of input variables (dimensions). There are several DRTs available that can be used to decrease processing time and allow better use of computing resources. DRTs can be applied prior to data analysis and ML model creation at the pre-processing level. For various data types, a broad range of DRTs [5] are available. Each DRT has been designed to retain some aspects of the original data, which is the major difficulty in applying DRTs. Therefore, For some kind of data or program, a specific DRT may be acceptable and may be unacceptable for others. In addition, under some restrictions that are restricted in scope and implementation, some DRTs have been designed. In addition, the main functionality of the DRT can be altered by modifying internal parameters and functions. A short description of the DRTs and their available variants is provided by this report. In addition, we identified different domain areas and data sets that have been successfully applied to DRTs. In addition, numerous concerns have been highlighted that may occur during the DR process.

According to the International Diabetes Federation (IDF), in 2015, 415 million people worldwide (1 in 11 adults aged 20-79 years) had diabetes mellitus[7]. In addition, this incidence of diabetes is expected to grow to 642 million by 2040, and this will inevitably pose a major challenge for healthcare systems[8], which are currently small, costly and some are actually unproductive due to drug resistance, and thus there is an urgent need for new

approaches to more potent drugs with different modes of action to handle this life-threatening diseases.

1.1. Protein tyrosine phosphatase 1B

One of the most important mechanisms for the control of cell growth, differentiation and regulation of cell functions is tyrosine phosphorylation of proteins. It is a reversible mechanism regulated by the opposing effects of dephosphorylation-responsible protein tyrosine phosphatases (PTPs) and protein tyrosine kinases (PTKs) [9]. In addition, skewed and faulty effects of PTKs and PTPs contribute to irregular phosphorylation processes of tyrosine, leading to a variety of diseases such as diabetes, inflammatory disorders, and cancer[10].PTP1B belongs to intracellular PTP, which is involved in the signaling mechanism of negative insulin regulation and leptin. The dephosphorylation of tyrosine residues (pY1162/pY1163) of the activated insulin receptor b subunit (IRb) and the insulin receptor substrate-1 (IRS-1) is actually catalyzed by PTP1B, thereby significantly affecting the length and amplitude of the cellular response to insulin[11]. The catalytic domain of PTP1B revealed that nearly 40 percent of its sequence is like that of another family.

Due to its ability to attenuate insulin signaling, PTP1B has gained a great deal of attention in recent years and is widely considered as a possible therapeutic target against metabolic syndrome, obesity and diabetes. The role of PTP1B in the negative regulation of insulin sensitivity and metabolism using whole-body and tissue-specific PTP1B knockout mouse models has been demonstrated in several studies [15]. In addition, PTP1B was found to contribute to the impairment of cardiovascular functions directly or indirectly (through insulin resistance), especially endothelial function Dictionary Learning (DL) is wildly applied to classification because of its superiority in data representation. In different tasks for classification, such as visual tracking, face recognition, fault diagnosis and event detection many examples can be discovered where DL is explicitly beneficial and useful. Classic dictionary learning methods, such as K-Singular Value Decomposition (KSVD), Label Consistent K-SVD (LC-KSVD), Fisher Discrimination Dictionary Learning (FDDL), and Support Vector Guided Dictionary Learning (SVGDL) have been presented recently. However, these dictionary learning methods become challenging when applied to high-dimensional datasets, since the memory usage and computational complexity in dictionary learning increase dramatically along with the increase in dimensionality [16]. Accordingly, the classification performance will also be hindered by the curse of dimensionality.

A dimension reduction technique is also implemented to project the original high-dimensional data to a low-dimensional latent space to overcome the problems caused by high dimensionality, and then a desirable dictionary is learned in the low-dimensional space. The most representative algorithms employed for dimension reduction are Principal Component Analysis (PCA) and its extensions. They may not completely maintain optimal features for dictionary learning despite their simple implementation, as most of these methods merely perform dimension reduction and dictionary learning as two individual phases. Recent studies show that the dimension reduction and dictionary learning processes are required to be carried out jointly, so that a more efficient dictionary can be obtained for better classification performance[17]

II. RELATED WORKS

Many researchers have focused on conducting dimension reduction and dictionary learning jointly. As an exploratory work, in [18] simultaneously learned a task driven dictionary and a linear transform matrix, to reduce the dimension of the feature space. Similarly, [19] jointly learned a projection matrix and a discriminative dictionary for face representation. In [20] proposed a unified framework to integrate dimension reduction with dictionary learning. More recently, [21] carried out dimension reduction and dictionary learning jointly. Dimension reduction and dictionary learning procedures were also formulated into a single method.

In [22] developed a novel method for simultaneously studying a projection matrix and a low-rank dictionary. To find the most acceptable projection matrix and dictionary. In [23] suggested a joint projection and dictionary learning process. These techniques that jointly perform dimension reduction and dictionary learning (JDRDL) are more efficient than traditional techniques that individually perform dimension reduction and dictionary learning. Although JDRDL approaches have some encouraging contributions on high dimensional data classification tasks, a common criticism is that most of these methods may fail to consider the intrinsic nonlinear structure within high dimensional data. Specifically, at the stage of dimension reduction, the high-dimensional data is transformed through a linear transformation matrix into a low-dimensional space. Each data point is modeled as a linear combination of the dictionary atoms in the dictionary learning stage. However, in order to model the nonlinear associations between the original data dimensions, such a linear model is not effective enough. This inevitably results in limited performance for the high-dimensional task of data classification, even when the property of discrimination is taken into account during the training phase.

III. PROBLEM FORMULATION

Vectors are denoted by bold lower case letters and matrices by upper case ones. Let $X = [x_1, \cdots, x_N] \in R^{D\times N}$ denote the training set that contains N training samples belonging to C classes. Each training sample xi is a column vector with D dimensions. Let $D \in R^{D\times K}$ be a dictionary that contains K atoms. A general dictionary learning model can be determined as stated in equation (1):

$$\min_{D,A} \sum_{i=1}^{N} \left(\|x_i - Da_i\|_2^2 + \lambda \|a_i\|_P^P \right) \quad (1)$$

where $A = [a_1, \cdots, a_N] \in R^{K \times N}$ is the coding coefficient matrix of X over D. λ is a regularization parameter which balances the effects of data reconstruction and prior of coefficient. $\|\cdot\|_P^P$ denotes the p-order power of the l_p norm and p ∈ {1, 2}. Different settings of p can lead to different computational complexity. Having p = 1 leads to l_1-norm minimization problem which is time consuming. Setting p to 2 instead of 1 can reduce the computational complexity and achieve competitive classification accuracy.

Consequently, the basic dictionary learning model in (1) can be rewritten in an equivalent matrix form as follows in equation (2):

$$\min_{D,A} \sum_{i=1}^{N} \left(\|X - DA\|_F^2 + \lambda \|A\|_F^2 \right) \quad (2)$$

Where, $\|\cdot\|_F$ is the Frobenius norm of a matrix

Let $f : R^D \rightarrow R^d$, d < D denote the mapping that transforms the high-dimensional data to a low-dimensional latent space. In general, the mapping f is characterized using a transformation matrix P, whose dimension is d×D and d < D. By requiring P is orthonormal, one can jointly learn a transformation matrix P and a dictionary D through minimizing the following function in equation (3):

$$\min_{P,D,A} \|PX - DA\|_F^2 + \|X - P^T PX\|_F^2 + \lambda \|A\|_F^2 \quad (3)$$

then the dictionary is learned in the low-dimensional space, i.e. $D \in R^{d \times K}$. Specially, the second term of (3) represents the difference between the reconstruction data and the original high-dimensional data.

3.1 Proposed HDCF

The proposed joint dimension reduction and dictionary learning framework is introduced in details. Fig. 1 illustrates the basic idea of the proposed method, which is composed of the following parts. First, an autoencoder is used to maintain the nonlinear structure at the dimension stage to learn a nonlinear mapping that maps the high-dimensional data to a reduced space. The local structure words are integrated into the learning process to maintain the nonlinear structure at the dictionary learning level. The label information is further incorporated into the local restriction words in order to improve the ability to discriminate. In addition, simultaneous optimization of the nonlinear mapping and dictionary. Consequently, two elements of the nonlinearity of high-dimensional data are retained. In the following subsection, these two sections, called Nonlinear Mapping and Locality Restriction with Label Embedding, are introduced in detail.

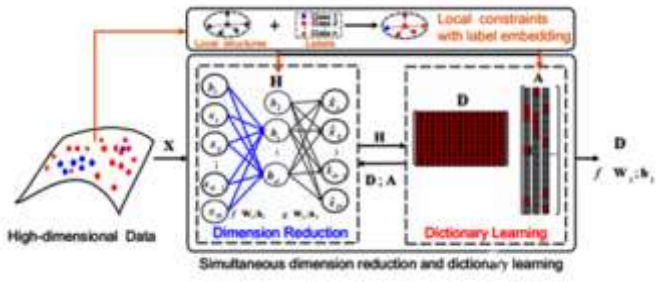

Figure 1: Architecture of HDCF

The proposed method employs an autoencoder to learn a nonlinear mapping $F : R^D \rightarrow R^d$, d < D. As the name suggests, an autoencoder consists of encoding and decoding stages. At encoding stage, the encoder maps the high-dimensional data x_i into a latent representation h_i as follows in equation (4):

$$h_i = \sigma(W_1 x_i + b_1) \quad (4)$$

Where W_1 and b_1 denote the weight matrix and the bias vector of the encoder, respectively. σ (·) is a nonlinear activation function applied component-wise tovectors. One of the commonly used activation functions is the logistic sigmoid given as $\sigma(t) = 1/(1 + \exp(-t))$

At decoding stage, the decoder reconstructs the high-dimensional data xi from the encoder output h_i as in equation (5) follows:

$$\hat{X}_i = \sigma(W_2 h_i + b_2) \quad (5)$$

where W_2 and b_2 denote the weight matrix and the bias vector of the decoder, respectively.

Considering the nonlinear mapping $F : R^D \rightarrow R^d$, d < D, the dictionary learning problem can be cast in the low-dimensional space as in equation (6)

$$\min_{\Omega,D,A} J_1 = l(\Omega, D, A) + \lambda_1 w(\Omega, A) = \frac{1}{2N} \left(\|H(\Omega - DA)\|_F^2 + \left\|\hat{X}(\Omega - X)\right\|_F^2 \right) + \lambda_1 (\|W_1\|_F^2 + \|W_2\|_F^2 + \|A\|_F^2) \quad (6)$$

where $\Omega = [W_1, W_2, b_1, b_2]$ denotes the parameter set of the autoencoder, D is the dictionary learned in the reduced space, A is the coding coefficients matrix of H (Ω) over D . $H(\Omega) = [h1\ (\Omega), \cdots, hN\ (\Omega)]$ and $h_i(\Omega) = \sigma(W_1 x_i + b_1)$ for $i = 1, \cdots N$. To improve the generalization ability, the weight decay penalty is added to restrict the magnitudes of weights W_1 and W_2. Based on these points above, the high-dimensional data are nonlinearly mapped to a low-dimensional space. The dictionary is learned in the low-dimensional space and simultaneously optimized with the nonlinear mapping illustrated in figure 2.

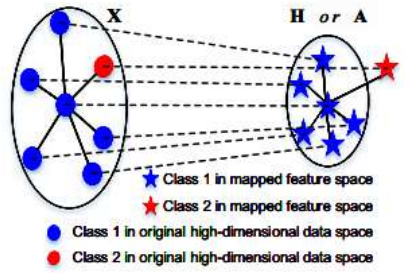

Figure 2: Constraints in label modelling

From the perspective of classification, it is desired to minimize intra-class variance and maximize inter-class separability. When considering the problem at dictionary learning stage, the following objective function can be determined in equation (7):

$$\min_A r(A) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|a_i - a_j\|_2^2 (S_{ij}^{intra} - S_{ij}^{inter}) = Tr(AL^{intra}A^T - Tr(AL^{intra}A^T) = Tr(ALA^T) \quad (7)$$

where T r(·) is the trace operator, L = Lintra − Linter, L inter = Binter − S inter , L intra = Bintra − S intra . Binter and Bintra are diagonal matrices whose elements are Binter ii = P N j=1 S inter ij and Bintra ii = P N j=1 S intra ij , separately.

3.2 HDCF for PTP-1B data processing

High dimensions increases computational complexity of the underlying algorithm and it reduces quality of the results and sometimes misleads the algorithm. The real-world large dataset consists of irrelevant, redundant and noisy dimensions. Therefore in our proposed approach we consider the dimension reduction for high dimensional data. Feature selection is a process that selects a subset of original features by rejecting irrelevant and redundant features according to certain criteria.

3.2.1 Dimensionality Reduction with Correlation Filtering Algorithm

In filter method, the variables are selected based on the model used. They are based only on general features like the correlation with the variable to predict [7]. Filter method contains the minimum interesting variables. Other variables will be used for either classification or regression models. These methods work efficiently under time consumption. It is only used as a preprocessing method and they choose the redundant variable because they do not matter the relationship between the variable. The flow of this class is proposed in Figure4. In this all the features are taken and from that the best subsets are chosen and the algorithms for further tasks are applied to it to find the performance shown in figure 3.

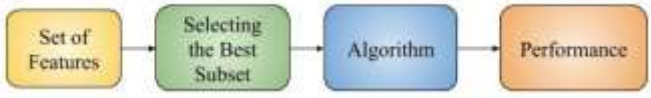

Figure 3: Process of Feature Estimation

The correlation based algorithm effectively deal with irrelevant and redundant attributes or features. The algorithm works in two steps. In the first step Pearson's correlation measure is used to remove the irrelevant attributes and helps to retain the subset of relevant attributes. In the second step we get the optimal subset by removing the redundant attributes with the help of graph based clustering.

Algorithm 1: HDCF for PTP-1B
Input: Training data $X = [x1, \cdots, xN] \in R^{D \times N}$,Training labels, Neighborhood size k, hidden layer size d, regularization parameters λ1, λ2, λ3, max iteration number T and convergence error ε. **Output:** Autoencoder parameter Ω, dictionary D. 1. Initialization 2. Initialize D and A by using the K-SVD algorithm. 3. Initialize W as random matrixes and b as zero vectors. 4. Calculate Laplacian matrix L 5. Learn Ω with Fixed D and A via (7)-(9); 6. Learn A with Fixed Ω and D via (21);

7. Learn D with Fixed Ω and A via (26);

8. Go to Step 2 until the convergence is met or maximum number of iteration T is reached.

9. Compute the d-dimensional mean vectors for the different classes from the dataset.

10. Compute the scatter matrices (between-class and withinclass scatter matrix).

11. Compute the eigenvectors (e_1, e_2... e_d) and corresponding eigenvalues (λ_1, λ_2... λ_d) for the scatter matrices.

12. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k-dimensional matrix W (where every column represents an eigenvector).

13. Use this d×k eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the equation Y = X × W (where X is an n×d-dimensional matrix; the i^{th} row represents the i^{th} sample, and Y is the transformed n×k-dimensional matrix with the n samples projected into the new subspace

14. Output the autoencoder parameters Ω and the learned dictionary D.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

This section presents the experimental results achieved by the proposed HDCF for the PTP-1B inhibators estimation. For the experimentation, the work considers the images from the genecards database.

4.1 Experimental setup

The entire implementation of the proposed HDCF is done in the MATLAB tool and the configurations considered for the experimentation are: PC with Windows 10 OS, 4GB RAM, and Intel I3 processor.

4.2 Database description

The single cell blood smear samples for the experimentation of the proposed HDCF are taken from genecards database. The genecards database has the large collection of cropped section of normal and blast cells. The gray level properties of the genecards database is almost similar with the genecards database, but with the large dimension.

4.3 Performance metrics

The performance of the proposed genecards is evaluated based on metrics, such as True Negative Rate (TNR), True Positive Rate (TPR) and accuracy.

The proposed method was evaluated on four public benchmark datastes from different areas, including one face image dataset, one scene image dataset, one audio dataset and one industrial process dataset. These datasets containseveral hundreds or thousands of dimensions and considerably differ in types, making them a suitable choice for high-dimensional data classification task. In table 1 parameters and characteristics are presented.

Table 1: Parameters Characteristics

The effects of reduced dimensionality and dictionary size were first evaluated on the classification performance. Different reduced dimensions and dictionary sizes were discussed on the four benchmark datasets. The average classification accuracy was obtained by repeating the experiment ten times. In table 2. the different iteration simulation values estimated are presented.

Table 2: Simulation Time

Iteration	Values
0	2.4
5	0.7
10	0.6
15	0.5
20	0.4
25	0.4
30	0.4

As Figure 4, demonstrated, both the reduced dimensionality and dictionary size have great impacts on the classification performance.

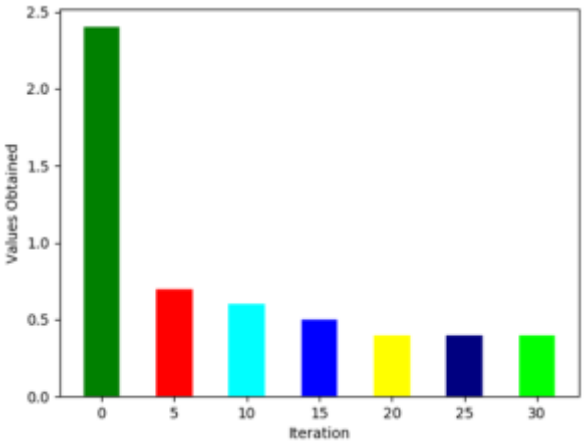

Figure 4: Simulation Time for different iteration

Fortunately, the proposed method achieves the best results across all dimensions and dictionary sizes. Specially, when the reduced dimension is low, or the number of dictionary size is small, the proposed method is significantly better than the other methods. Both the nonlinear preserving and joint learning strategies contribute the superior classification performance of the proposed method, which further shows the effectiveness of exploring nonlinear structure at both dimension reduction and dictionary learning stages in the proposed method.

4.4 Comparative Analysis

To highlight the advantages of the proposed method, the performance is further compared with several approaches which combines dictionary learning with multilayer neural networks, including Deep Micro-Dictionary Learning and Coding Network (DDLCN), Multi-layer Discriminative Dictionary Learning with two layers (MDDL-2), Multi-layer Discriminative Dictionary Learning with three layers (MDDL-3), Dictionary Learning Enhancement (DLE) framework, Supervised Nonlinear Dictionary Learning (SNDL) and Structured Dictionary Layer (SDL).

Since the DDLCN method relies on scale-invariant feature transform (SIFT) descriptor, illustrative experiments are only performed on the genecards datasets in this section. The detailed classification results of the comparison methods are summarized in Table 3 and graphically illustrated in figure 5.

Method	Mean %	Best %	Classification Accuracy %	Time (Sec)
PCA+CRC	93.28	93.42	94.23	30.69
PCA+FDDL	92.00	92.14	94.52	66.72
ETDDL	90.71	90.85	94.90	16.43
SEDL	94.29	94.71	89.00	11.76
KDLDP	95.14	95.43	91.37	12.86
HDCF	96.35	96.74	95.73	11.28

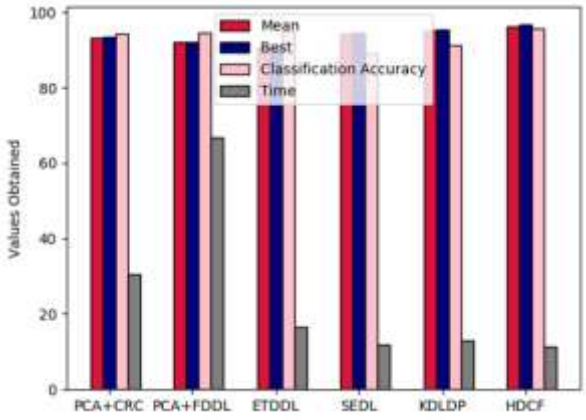

Figure 5: Comparative analysis

Specifically, all the results were from the original papers except those of DDLCN, which were obtained by re-implementations. Moreover, the results of DLE was obtained when genecards dataset is corrupted by Gaussian noise of standard deviation 20. To have a fair comparison, the proposed method was also implemented by following the settings in when compared with DLE.

An observation is that the proposed method performs better than the compared methods. Compared with multilayer dictionary learning methods, such as MDDL and DDLCN, which learn multiple levels of dictionaries layer by layer, the proposed method achieves higher classification results. The experimental results demonstrate that joint optimization the nonlinear mapping and dictionary can exploit more discriminative information for classification performance improvement. Similarly, the classification performance of the proposed method is better than that of DLE, SNDL and SDL. The main reason is that the DLE, SNDL and SDL algorithms only exploit nonlinear structure at nonlinear mapping stage. In contrast, the proposed method not only explores the nonlinear structure via an autoencoder at dimension reduction stage, but also uses the local structure to model the nonlinearity at dictionary learning stage. As a consequence, the learned dictionary of the proposed method is more discriminative than the learned dictionaries of DLE, SNDL and SDL.

V. CONCLUSION

A novel joint dimension reduction and dictionary learning framework for high-dimensional data classification was proposed in this paper. The main advantage of the proposed method is that it can efficiently preserve nonlinearity within the high-dimensional data, at both dimension reduction and dictionary learning stages. The proposed method automatically learns a nonlinear mapping from the original data via an autoencoder, so that the nonlinear structure of high-dimensional data can be well exploited at dimension reduction stage. Besides, by incorporating locality and category label information into the learning process, the proposed method can better preserve the nonlinear local structure at dictionary learning stage and further enhance discrimination capability. Experimental results demonstrate that considering the nonlinear structure among the high-dimensional data at both dimension reduction and dictionary learning stages can significantly promote classification performance. However, the proposed method still has some limitations. First, the proposed method may be a bit sensitive to outliers since the l2 regularization is less robust to outliers than the l1 regularization. When the training data is contaminated heavily because of noise and outliers, the proposed method may result in degenerated classification performance. Also, the proposed method assumes that the sizes of different classes are similar. However, in real-word applications the collected data often exhibits imbalanced class distribution.

Therefore, the future work will focus on designing a more robust classification framework and considering imbalanced data distribution to obtain enhanced classification performance.

VI. REFERENCES

[1] Anagaw, A., & Chang, Y. L. (2019). A new complement naïve Bayesian approach for biomedical data classification. *Journal of Ambient Intelligence and Humanized Computing*, *10*(10), 3889-3897.

[2] Zhang, L., Yang, H., & Jiang, Z. (2018). Imbalanced biomedical data classification using self-adaptive multilayer ELM combined with dynamic GAN. *Biomedical engineering online*, *17*(1), 181.

[3] Bressan, R. S., Camargo, G., Bugatti, P. H., & Saito, P. T. M. (2018). Exploring active learning based on representativeness and uncertainty for biomedical data classification. *IEEE journal of biomedical and health informatics*, *23*(6), 2238-2244.

[4] Liu, S., Zhang, J., Xiang, Y., Zhou, W., & Xiang, D. (2019). A study of data pre-processing techniques for imbalanced biomedical data classification. *arXiv preprint arXiv:1911.00996*.

[5] Wang, T., Shao, W., Huang, Z., Tang, H., Ding, Z., & Huang, K. (2020). MORONET: Multi-omics Integration via Graph Convolutional Networks for Biomedical Data Classification. *bioRxiv*.

[6] Desai, H. (2020). Biomedical Data Classification with Improvised Deep Learning Architectures.

[7] Ceylan, R., & Barstugan, M. (2018). Feature selection using ffs and pca in biomedical data classification with adaboost-svm. *International Journal of Intelligent Systems and Applications in Engineering*, *6*(1), 33-39.

[8] Ceylan, R., & Barstugan, M. (2018). Feature selection using ffs and pca in biomedical data classification with adaboost-svm. *International Journal of Intelligent Systems and Applications in Engineering*, *6*(1), 33-39.

[9] Zhang, B., Cao, P., Zhang, Y., Zhang, C., Li, Z., Qu, Z., ... & Hu, B. (2018, July). Feature selection algorithm for high dimensional biomedical data classification based on redundant removal. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32* (pp. 1-5).

[10] Das, H., Naik, B., & Behera, H. S. (2020). An experimental analysis of machine learning classification algorithms on biomedical data. In *Proceedings of the 2nd International Conference on Communication, Devices and Computing* (pp. 525-539). Springer, Singapore.

[11] Gautam, C., Mishra, P. K., Tiwari, A., Richhariya, B., Pandey, H. M., Wang, S., ... & Alzheimer's Disease Neuroimaging Initiative. (2020). Minimum variance-embedded deep kernel regularized least squares method for one-class classification and its applications to biomedical data. *Neural Networks*, *123*, 191-216.

[12] Student, S., Łakomiec, K., Płuciennik, A., Bensz, W., & Fujarewicz, K. (2019, June). Classification System for Multi-class Biomedical Data that Allows Different Data Fusion Strategies. In *International Conference on Information Technologies in Biomedicine* (pp. 593-602). Springer, Cham.

[13] Bidi, N., & Elberrichi, Z. (2017). Using Penguins search optimization algorithm for best features selection for biomedical data classification. *International Journal of Organizational and Collective Intelligence (IJOCI)*, *7*(4), 51-62.

[14] Shakeel, P. M., & Manogaran, G. (2020). Prostate cancer classification from prostate biomedical data using ant rough set algorithm with radial trained extreme learning neural network. *Health and Technology*, *10*(1), 157-165.

[15] Nagpal, S., Arora, S., & Dey, S. (2017). Feature selection using gravitational search algorithm for biomedical data. *Procedia Computer Science*, *115*, 258-265.

[16] Bodur, E. K., & Atsa'am, D. D. (2019). Filter variable selection algorithm using risk ratios for dimensionality reduction of healthcare data for classification. *Processes*, *7*(4), 222.

[17] Reddy, G. T., Reddy, M. P. K., Lakshmanna, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis of dimensionality reduction techniques on big data. *IEEE Access*, *8*, 54776-54788.

[18] Lee, C., Luo, Z., Ngiam, K. Y., Zhang, M., Zheng, K., Chen, G., ... & Yip, W. L. J. (2017). Big healthcare data analytics: Challenges and applications. In *Handbook of Large-Scale Distributed Computing in Smart Healthcare* (pp. 11-41). Springer, Cham.

[19] Dinov, I. D. (2016). Volume and value of big healthcare data. *Journal of medical statistics and informatics*, *4*.

[20] Wang, F., & Sun, J. (2015). Survey on distance metric learning and dimensionality reduction in data mining. *Data mining and knowledge discovery*, *29*(2), 534-564.

[21] Avci, A., Bosch, S., Marin-Perianu, M., Marin-Perianu, R., & Havinga, P. (2010, February). Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey. In *23th International conference on architecture of computing systems 2010* (pp. 1-10). VDE.

[22] Dhillon, A., & Singh, A. (2019). Machine learning in healthcare data analysis: a survey. *Journal of Biology and Today's World*, *8*(6), 1-10.

[23] Fujiwara, T., Sakamoto, N., Nonaka, J., Yamamoto, K., & Ma, K. L. (2020). A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction. *IEEE Transactions on Visualization and Computer Graphics*.

[24] Chhikara, P., Jain, N., Tekchandani, R., & Kumar, N. (2020). Data dimensionality reduction techniques for Industry 4.0: Research results, challenges, and future research directions. *Software: Practice and Experience*.

[25] Pasha, A., & Latha, P. H. (2020). Bio-inspired dimensionality reduction for Parkinson's disease (PD) classification. *Health Information Science and Systems*, *8*(1), 1-22.

[26] Karami, A. (2019). Application of fuzzy clustering for text data dimensionality reduction. *International Journal of Knowledge Engineering and Data Mining*, *6*(3), 289-306.

[27] Xu, N., Huang, S. H., & Wulsin, L. (2007). Dependency analysis for dimensionality reduction in healthcare data analysis. In *IIE Annual Conference. Proceedings* (p. 1398). Institute of Industrial and Systems Engineers (IISE).

[28] Modaresnezhad, M., Vahdati, A., Nemati, H., Ardestani, A., & Sadri, F. (2019). A rule-based semantic approach for data integration, standardization and dimensionality reduction utilizing the UMLS: application to predicting

bariatric surgery outcomes. *Computers in biology and medicine*, *106*, 84-90.

[29] Nguyen, T., Khosravi, A., Creighton, D., & Nahavandi, S. (2015). Classification of healthcare data using genetic fuzzy logic system and wavelets. *Expert Systems with Applications*, *42*(4), 2184-2197.

[30] BIRNBAUM, E. B. D. (2004). Application of data mining techniques to healthcare data. *Infection control and hospital epidemiology*.