HHS Public Access

Author manuscript

Annu Int Conf IEEE Eng Med Biol Soc. Author manuscript; available in PMC 2022 March 21.

Published in final edited form as:

Annu Int Conf IEEE Eng Med Biol Soc. 2021 November ; 2021: 1810–1813. doi:10.1109/EMBC46164.2021.9629990.

Temporomandibular Joint Osteoarthritis Diagnosis Using Privileged Learning of Protein Markers

Winston Zhang[1], **Jonas Bianchi**[2,3,4], **Najla Al Turkestani**[2], **Celia Le**[2], **Romain Deleat-Besson**[2], **Antonio Ruellas**[2], **Lucia Cevidanes**[2], **Marilia Yatabe**[2], **Joao Gonçalves**[3], **Erika Benavides**[5], **Fabiana Soki**[5], **Juan Prieto**[6], **Beatriz Paniagua**[7], **Kayvan Najarian**[1], **Jonathan Gryak**[1], **Reza Soroushmehr**[1]

[1] Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA.

[2] Department of Orthodontics and Pediatric Dentistry, University of Michigan, Ann Arbor, MI 48109, USA

[3] Pediatric Dentistry and Orthodontics, São Paulo State University, São Paulo, BRAZIL

[4] Department of Orthodontics, University of the Pacific, Arthur A. Dugoni School of Dentistry, San Francisco, CA, United States

[5] Department of Periodontics and Oral Medicine, University of Michigan, Ann Arbor, MI 48109, USA

[6] Psychiatry, University of North Carolina, Chapel Hill, NC, USA

[7] Departments of Psychiatry, Orthodontics and Computer Science, University of North Carolina, Chapel Hill, NC, USA

Abstract

Diagnosis of temporomandibular joint (TMJ) Osteoarthritis (OA) before serious degradation of cartilage and subchondral bone occurs can help prevent chronic pain and disability. Clinical, radiomic, and protein markers collected from TMJ OA patients have been shown to be predictive of OA onset. Since protein data can often be unavailable for clinical diagnosis, we harnessed the learning using privileged information (LUPI) paradigm to make use of protein markers only during classifier training. Three different LUPI algorithms are compared with traditional machine learning models on a dataset extracted from 92 unique OA patients and controls. The best classifier performance of 0.80 AUC and 75.6 accuracy was obtained from the KRVFL+ model using privileged protein features. Results show that LUPI-based algorithms using privileged protein data can improve final diagnostic performance of TMJ OA classifiers without needing protein microarray data during classifier diagnosis.

wwzhang@umich.edu .

I. Introduction

Temporomandibular Joint (TMJ) disorders are clinical conditions affecting over 10 milllion Americans [1]. TMJ degeneration breaks down cartilage and alters the bone shape of the mandibular condyle and articular fossa [2]. These morphological changes result in chronic pain and decreased quality of life [3]. Diagnosis of osteoarthritic (OA) changes in the TMJ is not straightforward due to the frequent absence of symptoms before significant degeneration of the joint has occurred [4]. No disease-modifying therapy exists.

Clinical, imaging, and biomolecular protein features have been studied and found to help with early diagnosis of TMJ OA [5]. Cone-beam computed tomography (CBCT) imaging has been added as diagnostic criteria to detect presence of bone morphology alteration since 2014 [6]-[7]. Protein levels in synovial fluid, serum, and saliva were found to be correlated with bone resorption [8] and may be potential therapeutic targets.

Unfortunately, clinical centers are often not equipped with the necessary microarray kits to measure protein levels in patient serum and saliva. Furthermore, collection of protein biomarkers from every patient may become cost-prohibitive. Thus, reducing the dependence of TMJ OA diagnosis on collecting protein biomarkers will improve clinical access to developed OA classifiers.

Learning using Privileged Information (LUPI) aims to improve the generalization ability of machine learning models by using a set of additional information named privileged information [9]. Intuitively, privileged information helps the diagnostic model determine between easy and hard samples in the training set [10]. Privileged information acts as a teacher that guides the learning model during training, but is not available during testing stage prediction.

LUPI is suitable for clinical applications when multimodal data can be used to construct a training dataset, but is difficult to collect during diagnosis. Li et al. found that the LUPI framework was effective when using separate neuro-imaging modalities as privileged information to diagnose Alzheimer's disease [11]. Ye et al. also used multimodal MRI imaging data to improve glioma classification [12]. LUPI is also useful for incorporating information from widely differing data sources. Duan et al. augmented glaucoma detection from images with privileged information from single nucleic polymorphisms [13].

We propose a new LUPI-based framework for TMJ OA diagnosis incorporating protein marker levels as privileged information. The main advantage is an increased accessibility of TMJ OA diagnosis, as final diagnostic models will only rely on routinely collected clinical and radiographic data to make a classification. Furthermore, the multimodal nature of data used for TMJ OA classification is well suited for the LUPI paradigm. The dataset used is introduced in Section II. The proposed LUPI algorithms are described in III. Experiment results are shown in IV and finally, conclusions are drawn in Section V.

II. Dataset

The cross-sectional dataset consists of data from 46 early-stage TMJ OA patients and 46 age and gender-matched controls. Patient diagnosis was confirmed by a TMJ specialist at the University of Michigan Medicine Oral Surgery Clinic using the Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) [14]. Data was collected with informed consent and following guidelines of the Institutional Review Board (IRB) at the University of Michigan (number HUM00113199).

6 clinical features were collected by a specialist for every patient based on the DC/TMD criteria: age of patient, headaches in last 6 months, muscle soreness in last 6 months, vertical range unassisted without pain (mm), vertical range unassisted maximum (mm), vertical range assisted maximum (mm).

CBCT scans of each patient's condyles were obtained using the 3D Accuitomo machine (J. Morita MFG. CORP Tokyo acquisition, Japan) and exported to DICOM format using manufacturer software. Details of scanning protocol can be found in [5]. An example scan is in Figure 1. Radiomic features were collected using the BoneTexture module from the 3D-Slicer software using optimal parameters found in [15]. 23 texture and bone morphology features were extracted from the lateral condyle and an additional 23 features from the mandibular fossa.

13 protein levels previously found correlated with arthritis progression, inflammation, and bone morphology changes [8] were measured in each patient's saliva and serum samples using custom protein microarrays from RayBiotech, Inc. Norcross, GA. Sample acquisition protocols and raw values are described in [5]. The protein MMP-3 was not expressed in saliva, hence 25 protein features were collected.

In total, the dataset consists of 77 features (6 clinical, 46 imaging, and 25 protein) collected from 92 patients.

III. Methods

In the general LUPI-framework, the training dataset can be represented as a set of triplets $\Gamma = \{(x_i, x_i^*, y_i) \mid x_i \in X, x_i^* \in X^*, y_i \in Y, i = 1 \ldots n\}$, where n is the number of training samples, X and X^* represent the original and privileged feature sets, d and d^* are the number of features in X and X^*, and Y is the set of training labels.

In the TMJ OA dataset, there are 92 patients ($n = 92$) and we treat the imaging and clinical features as the original feature set X, where $d = 52$ features (6 clinical + 46 imaging). We treat the protein features as the privileged feature set X^*, where $d^* = 25$ features (i.e. 25 protein). $y_i \in \{-1, 1\}$ which denotes control and OA patients respectively.

Figure 2 summarizes the steps for training and applying the LUPI-based TMJ OA classifier. During training, the imaging and clinical feature set, X, and the protein feature set, X^*, undergo feature selection to find the optimal set of features for classifier training. The LUPI classifier will then attempt to differentiate between control and OA patients in the training

set with X, while using X^* as auxiliary information to distinguish between easy and hard training samples [10].

In the testing stage, protein features X^* are unavailable. The features previously selected from X during training are used for classifier inference on the unseen test set patients, after which a classification of control or OA is obtained for each test patient.

A. Feature Selection

Traditional feature selection methods choose features based on relevance (i.e. mutual information) with the target label Y [16]. Following the intuition that privileged features transfer information to help a model assess the difficulty of each training sample, features chosen from X^* should contain information about Y to ensure chosen features will help final performance. However, chosen privileged features should not contain information from X to prevent transfer of redundant information [17].

We calculate feature relevance with the Area Under Receiver Operating Curve (AUC) value from a two-sample Mann-Whitney U test between feature and target variable. We then evaluate two feature selection methods: (M1) selection of the top k and k^* features by highest AUC, where k and k^* are hyperparameters representing the number of features chosen from X and X^* respectively; and (M2), selection of the same features as in (M1), except for the replacement of all features from X^* that have an absolute Pearson correlation with any feature from X exceeding 0.5.

B. LUPI Algorithms

We develop and compare 3 LUPI classifiers with non-LUPI analogues in order to evaluate the effectiveness of the LUPI paradigm for TMJ OA diagnosis.

The Support Vector Machine (SVM) is a commonly used classifier. In this work we use the gaussian kernel SVM for all experiments. SVM+ proposed by Vapnik et al. [9] uses X^* to model the slack variables for training samples found in soft-margin SVM by solving for the following primal problem:

$$\begin{aligned} \min_{w, w^*, b, b^*} \frac{1}{2}\left(\|w\|_2^2 + \gamma\|w^*\|_2^2\right) + C\sum_{i=1}^{n}(w^* \cdot x_i^* + b^*) \\ s.t.\ y_i(w \cdot x_i + b) \geq 1 - (w^* \cdot x_i^* + b^*), i = 1, \ldots, n \\ (w^* \cdot x_i^* + b^*) \geq 0, i = 1, \ldots, n \end{aligned} \tag{1}$$

where w, b, w^*, and b^* are the weight and bias variables for the decision rule and correcting function respectively and γ and C are hyperparameters > 0. In our experiments we use the alternating SMO introduced by Pechyony et al. for optimization of SVM+ [18].

The random vector functional link network (RVFL) [19] is a simple feedforward neural network.

In Figure 3, w and b are randomly initialized and fixed during training, so that input data fed into the network is transformed into two sets of inputs: H_1 containing the original input

and H_2 containing the inputs transformed by w and b. The weights for the H_1 and H_2 inputs, β, are then calculated. To incorporate privileged input X^*, the model RVFL+ optimizes a similar problem as SVM+:

$$\min_{\beta,\beta^*} \frac{1}{2}\left(\|\beta\|_2^2 + \gamma\|\beta^*\|_2^2\right) + C\sum_{i=1}^{n}\left(\tilde{h}(x_i^*)\beta^*\right) \quad s.t.\ h(x_i)\beta = y_i - \tilde{h}(x_i^*)\beta^*, i = 1, \ldots, n \tag{2}$$

Here, X^* is fed into a separate RVFL network and used to estimate the correcting function. $h(x_i)$ and $\tilde{h}(x_i^*)$ represent the concatenation of H_1 and H_2 in the original and privileged networks respectively. A kernel-based RVFL+ algorithm, named KRVFL+, approximates RVFL+ while improving generalization. A detailed formulation for RVFL and KRVFL+ can be found in [20].

Finally, the iterated privileged learning model (IPL) [21] applies privileged information to weak learner boosting. We compare this LUPI model with the traditional boosting method Adaboost [22]. Classification and regression tree (CART) method is used as the weak learner in both algorithms.

IV. Experiments and Results

We evaluate both non-LUPI and LUPI-based algorithms on the TMJ OA dataset using an 80%−20% training testing split. We conduct a grid search and 5-fold cross-validation (CV) on the training set to determine the optimal hyperparameters for each algorithm. Hyperparameters k and k^* are also grid searched over the range [5, d] and [5, d^*] respectively. Test results are obtained by taking the average of test set predictions from all 5 models trained during the best 5-fold CV run. The overall procedure is repeated 10 times to avoid sampling bias from random train-test partitioning, and final reported results are the mean±standard deviation (SD) test results across all 10 repetitions.

A. Feature Selection Comparison

We first evaluate the 2 proposed feature selection methods by comparing the SVM and SVM+ algorithms on different feature sets. In table I, Cl are the clinical features, Im the imaging, and Pr the protein. *Pr indicates privileged protein features, meaning that the LUPI-based SVM+ algorithm is used. (M1) and (M2) indicate the feature selection method used.

Experiments using LUPI-based methods perform better than non-LUPI methods. Experiment A outperforms B despite B containing protein features in the original feature set. This suggests protein information is better incorporated into diagnosis using the LUPI framework. Feature selection (M2) results in higher performance than (M1), so the following experiments will use method (M2) for privileged selection.

B. LUPI and non-LUPI Comparison

We evaluate the performance of 3 non-LUPI algorithms (Gaussian SVM, RVFL, and Adaboost) using $X = \{\text{Cl} \cup \text{Im}\}$. We compare them to the performance of 3 LUPI-based algorithms (SVM+, KRVFL+, and IPL) using X = {Cl ∪ Im} and $X^* = \{\text{Pr}\}$. Results are in Table II.

In all comparisons, LUPI outperforms non-LUPI methods. Accuracy increased by at least 2.2% for all 3 comparisons. AUC and Accuracy SD decreased for IPL and KRVFL+ models.

C. Protein Interactions

Finally, we evaluate the effect of interaction features when added to X^*. Protein interactions were found to be important in TMJ OA diagnosis [5]. We create feature set PrX by taking the pairwise product of every feature in Pr, generating $\sum_{i=1}^{d^*} i = 325$ new interaction features when $d^* = 25$. Here, we use $X = \{\text{Cl} \cup \text{Im}\}$ and $X^* = \{\text{Pr} \cup \text{PrX}\}$. All experiments use feature selection method (M2). Results are in Table III.

All comparisons had an increase in performance after including protein interaction features. Comparing PrX LUPI with non-LUPI performance from the previous experiment, SVM+ improved AUC by 0.027 and accuracy by 2.7% when compared to SVM, IPL improved AUC by 0.018 and accuracy by 7.3% when compared to Adaboost, and KRVFL+ improved AUC by 0.056 and accuracy by 5.6%. The SD of AUC and accuracy also decreased for the IPL and KRVFL+ models after the introduction of protein interactions. The corrected two-tailed t-test between the AUC performances of KRVFL+ with PrX features and non-LUPI RVFL found a significant p-value of 0.035.

V. Conclusions

We developed a novel LUPI framework for TMJ OA diagnosis, combining clinical, imaging, and protein data. While clinical and imaging markers are the current criteria for disease classification, this study's experimental results show that all LUPI algorithms improved the performance when compared with their analogous baseline classifiers that do not use privileged protein information. Correlation-based privileged feature selection also showed improvement compared to naive feature ranking for privileged features. Furthermore, LUPI outperforms non-LUPI performance even when protein features are included in the original feature set and available during testing. This agrees with previous studies by Li et al. [11] and Duan et al. [13] suggesting that incorporating multimodal data as privileged information helps train better classifiers. overall, the LUPI framework shows promise for improving the cost accessibility and performance of TMJ OA diagnosis using protein markers.

Acknowledgments

*Grant supported by NIDCR DEO24450

References

[1]. "Tmj (temporomandibular joint and muscle disorders)." [Online]. Available: https://www.nidcr.nih.gov/health-info/tmj

[2]. Wang X, Zhang J, Gan Y, and Zhou Y, “Current understanding of pathogenesis and treatment of tmj osteoarthritis,” Journal of dental research, vol. 94, no. 5, pp. 666–673, 2015. [PubMed: 25744069]

[3]. Tanaka E, Detamore M, and Mercuri L, “Degenerative disorders of the temporomandibular joint: Etiology, diagnosis, and treatment,” Journal of Dental Research, vol. 87, no. 4, pp. 296–307, 2008. [PubMed: 18362309]

[4]. Kalladka M, Quek S, Heir G, Eliav E, Mupparapu M, and Viswanath A, “Temporomandibular joint osteoarthritis: diagnosis and long-term conservative management: a topic review,” The Journal of Indian Prosthodontic Society, vol. 14, no. 1, pp. 6–15, 2014. [PubMed: 24604992]

[5]. Bianchi J, de Oliveira Ruellas AC, Gonçalves JR, Paniagua B, Prieto JC, Styner M, Li T, Zhu H, Sugai J, Giannobile W et al. , “Osteoarthritis of the temporomandibular joint can be diagnosed earlier using biomarkers and machine learning,” Scientific reports, vol. 10, no. 1, pp. 1–14, 2020. [PubMed: 31913322]

[6]. dos Anjos Pontual M, Freire J, Barbosa J, Frazão M, dos Anjos Pontual A, and Fonseca da Silveira M, “Evaluation of bone changes in the temporomandibular joint using cone beam ct,” Dentomaxillofacial Radiology, vol. 41, no. 1, pp. 24–29, 2012. [PubMed: 22184625]

[7]. Ebrahim FH, Ruellas AC, Paniagua B, Benavides E, Jepsen K, Wolford L, Goncalves JR, and Cevidanes LH, “Accuracy of biomarkers obtained from cone beam computed tomography in assessing the internal trabecular structure of the mandibular condyle,” Oral surgery, oral medicine, oral pathology and oral radiology, vol. 124, no. 6, pp. 588–599, 2017.

[8]. Cevidanes LH, Walker D, Schilling J, Sugai J, Giannobile W, Paniagua B, Benavides E, Zhu H, Marron JS, Jung BT et al. , “3d osteoarthritic changes in tmj condylar morphology correlates with specific systemic and local biomarkers of disease,” Osteoarthritis and Cartilage, vol. 22, no. 10, pp. 1657–1667, 2014. [PubMed: 25278075]

[9]. Vapnik V and Vashist A, “A new learning paradigm: Learning using privileged information,” Neural networks, vol. 22, no. 5–6, pp. 544–557, 2009. [PubMed: 19632812]

[10]. Sharmanska V, Quadrianto N, and Lampert CH, “Learning to rank using privileged information,” in Proceedings of the IEEE international conference on computer vision, 2013, pp. 825–832.

[11]. Li Y, Meng F, and Shi J, “Learning using privileged information improves neuroimaging-based cad of alzheimer’s disease: a comparative study,” Medical & biological engineering & computing, vol. 57, no. 7, pp. 1605–1616, 2019. [PubMed: 31028606]

[12]. Ye F, Pu J, Wang J, Li Y, and Zha H, “Glioma grading based on 3d multimodal convolutional neural network and privileged learning,” in 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017, pp. 759–763.

[13]. Duan L, Xu Y, Li W, Chen L, Wong DWK, Wong TY, and Liu J, “Incorporating privileged genetic information for fundus image based glaucoma detection,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2014, pp. 204–211.

[14]. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet J-P, List T, Svensson P et al. , “Diagnostic criteria for temporomandibular disorders (dc/tmd) for clinical and research applications: recommendations of the international rdc/tmd consortium network and orofacial pain special interest group,” Journal of oral & facial pain and headache, vol. 28, no. 1, p. 6, 2014. [PubMed: 24482784]

[15]. Bianchi J, Gonçalves JR, Ruellas A. C. d. O., Vimort J-B, Yatabe M, Paniagua B, Hernandez P, Benavides E, Soki FN, and Cevidanes LHS, “Software comparison to analyze bone radiomics from high resolution cbct scans of mandibular condyles,” Dentomaxillofacial Radiology, vol. 48, no. 6, p. 20190049, 2019. [PubMed: 31075043]

[16]. Liu H and Motoda H, Computational methods of feature selection. CRC Press, 2007.

[17]. Izmailov R, Lindqvist B, and Lin P, “Feature selection in learning using privileged information,” in 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, 2017, pp. 957–963.

[18]. Pechyony D, Izmailov R, Vashist A, and Vapnik V, “Smo-style algorithms for learning using privileged information.” in Dmin. Citeseer, 2010, pp. 235–241.

[19]. Pao Y-H, Park G-H, and Sobajic DJ, “Learning and generalization characteristics of the random vector functional-link net,” Neurocomputing, vol. 6, no. 2, pp. 163–180, 1994.

[20]. Zhang P-B and Yang Z-X, “A new learning paradigm for random vector functional-link network: Rvfl+,” Neural Networks, vol. 122, pp. 94–105, 2020. [PubMed: 31677442]

[21]. Li X, Du B, Zhang Y, Xu C, and Tao D, “Iterative privileged learning,” IEEE transactions on neural networks and learning systems, vol. 31, no. 8, pp. 2805–2817, 2019. [PubMed: 30843851]

[22]. Freund Y and Schapire RE, “A decision-theoretic generalization of on-line learning and an application to boosting,” Journal of computer and system sciences, vol. 55, no. 1, pp. 119–139, 1997.

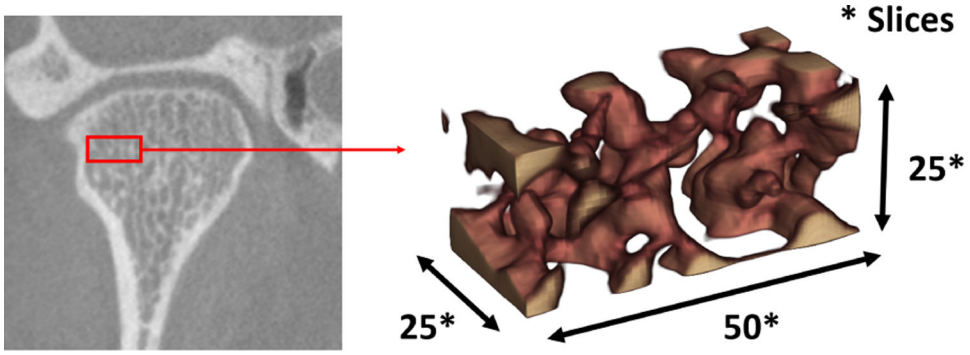

Fig. 1.

Example of raw CBCT slice with condyle volume of interest (VOI) selected in red (left). Trabecular bone VOI visualization (right).

III. METHODS

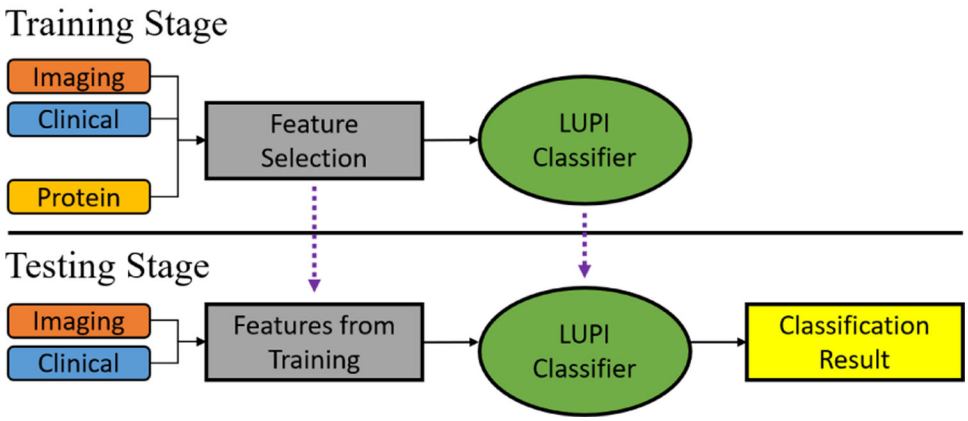

Fig. 2.
Proposed LUPI framework for TMJ OA diagnosis

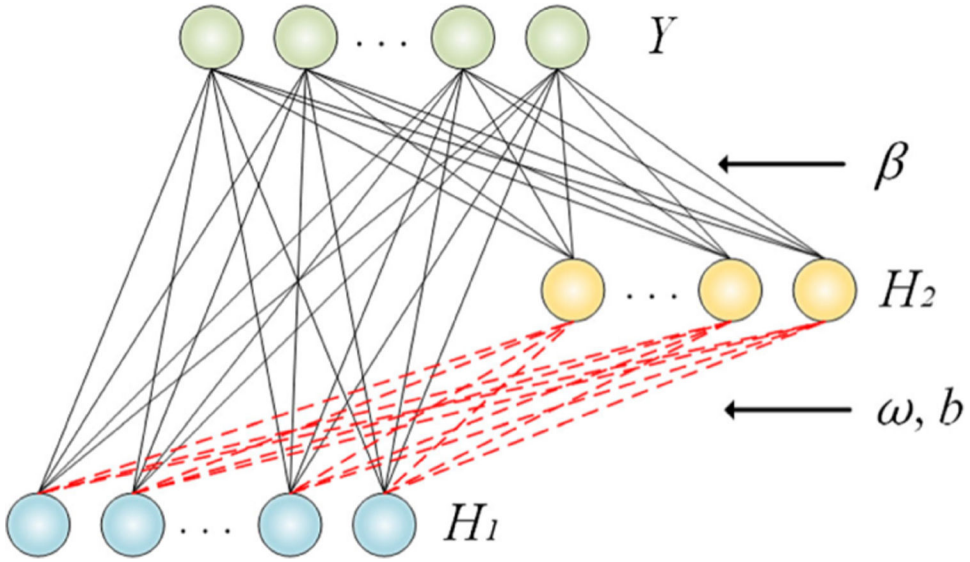

Fig. 3.
Random vector functional link network architecture [11]

TABLE I

SVM AND SVM+ COMPARISON ON FEATURE SETS

	Features	AUC	Accuracy
A	Cl+Im (M1)	0.759 ± 0.050	71.7 ± 4.9
B	Cl+Im+Pr (M1)	0.754 ± 0.048	70.0 ± 6.0
C	Cl+Im *Pr (M1)	0.764 ± 0.054	73.3 ± 5.7
D	Cl+Im *Pr (M2)	**0.773 ± 0.051**	**73.9 ± 6.4**

Author Manuscript

Author Manuscript

Author Manuscript

TABLE II

LUPI AND NON-LUPI COMPARISON

Algorithm	AUC	Accuracy
SVM	0.759 ± 0.050	71.7 ± 4.9
SVM+	0.773 ± 0.051	73.9 ± 6.4
Adaboost	0.767 ± 0.095	68.3 ± 8.7
IPL	0.775 ± 0.070	**75.0 ± 4.7**
RVFL	0.744 ± 0.072	70.0 ± 6.5
KRVFL+	**0.779 ± 0.053**	73.9 ± 5.3

TABLE III

EFFECT OF PROTEIN INTERACTION FEATURES ON PERFORMANCE

Algorithm	AUC	Accuracy
SVM+	0.773 ± 0.051	73.9 ± 6.4
SVM+ *PrX	0.786 ± 0.041	74.4 ± 6.0
IPL	0.775 ± 0.070	75.0 ± 4.7
IPL *PrX	0.785 ± 0.071	**75.6 ± 5.4**
KRVFL+	0.779 ± 0.053	73.9 ± 5.3
KRVFL+ *PrX	**0.800 ± 0.047**	**75.6 ± 6.0**