International Journal of Engineering Trends and Applications (IJETA) – Volume 7 Issue 4, Aug-Jul 2020

RESEARCH ARTICLE OPEN ACCESS

Multi Medical Data Classification with Feature Reduction Technique

Raja Seaker R
Annai College of arts and Science-kovilacheri

ABSTRACT
One of the challenging tasks in the field of medical informatics is medical data classification. From medical datasets and intends to upgrade the design of human services with the medical data classification entails to taking in classification designs. The medical data classification is the major aim of our research. The medical data features are selected with the help of the hybrid feature selection method. The high dimensional features are reduced by Discriminant Independent Component Analysis (DICA). The dataset details are collected from the UCI machine learning repository.
***Keywords*: -** medical informatics, data classification, hybrid feature selection, DICA, UCI machine learning repository.

I. INTRODUCTION

In disease diagnosis, the medical data classification can effectively assist physicians and that treat many diseases with proper prediction. The medical data classification performance is improved with the many efforts that have been made. Practically, the class imbalance issues are quite widespread [1-5]. The development of computing technologies with the medical field plays an important role. Leads to acquire the digital storage equipment and various medical activities such as prognosis, screening and diagnosis are the great amount of knowledge [6,7]. Different kinds of medical data mining or medical data are developed for knowledge discovery [59-62]. Here, the improvement of medical data accuracy by potentially useful and novel information [8-10]. The diagnosis and prognosis purposes present in medical data classification. Moreover, the noise is included in medical data that exhibit unique features by missing values and systematic errors [11-15].

The highly dependent on physicians experience the diagnosis accuracy. The larger amounts of information are stored and it is relatively easy to acquire. The medical decision support systems computerize the deployment. The prediction is the goal of the classification task [16-20]. One of the challenging tasks in medical informatics is classification tasks. Apply the medical data classification with statistical techniques. The difficult task to know the properties of the dataset that is not possible. Each parameter defines the types of medical databases contains a larger collection of medical data [21,22,56].

Two conflicts criterion such as intensification of the optimal solution and the diversification of the search space are defined during the design of metaheuristics [23-26]. The search space algorithm performance is improved with the help of a proper balance between these criteria [58]. The local and global search algorithm fusion is the memetic algorithm [27,57]. To ensure the diversification employs the global search approaches [28-30]. The optimal feature set is generated in these algorithms' different kinds of a searching algorithm. While compared to filter methods, the features generated with the help of wrapper methods [31,53,54,55]. To imbalances the diagnosis of medical classification samples that exist manor and major classes.

II. RELATED WORKS

The medical data classification based techniques were proposed in the past years. Few of these techniques are discussed in this section.

The feature ranking based method was proposed by Alam et al. [32] to classify the medical data. Few suitable ranking algorithms were used to dataset feature ranking and the high ranked features are predicted using Random Forest classifier. Ten-fold cross-validations with many other feature ranking algorithms are applied to evaluate the performance of feature ranking algorithm. Khanmohammadi et al. [33] proposed the model of Gaussian Mixture based Discretization (GMBD) algorithm for medical data classification, which preserve the most relevant features from the original dataset. The six various publicly available datasets are used to verify the efficiency of the GMBD algorithm. Gorzalczany et al. [34] proposed Multi-Objective Genetic Fuzzy Optimization (MOGFO) for fuzzy rule-based classification system (FRBCS) design from medical data. The different stage of accuracy interpretability characterizes the collection of medical FRBCS solutions. For medical data processing, they introduced a rule base representation of special coding-free and original genetic operators.

Bania et al. [35] proposed a selection method of parameter-free greedy ensemble attribute (R-Ensembler). The attribute-attribute relevance measure, attribute significant and

attribute class adopts the rough set theory concept. The various rough set produces multiple subset combination, which is combined by Ensembler method. During the attribute selection process time, the new *n* number of intersection method was proposed for the reduction of biasness and the dataset is preprocessed with kNN imputation. From the different subsets of the attribute pool, the Ensembler method highly effective is to select the high relevant features from the UCI medical dataset. The emperor penguin and social engineering optimization (memetic algorithm) were proposed by Baliarsingh et al. [36] for the classification of medical data. SVM is the faster classification method, but the selection of regularization and kernel parameters are the challenging issues of SVM. Hence, the author used a memetic algorithm for the tuning purpose of both regularization and kernel parameters (Memetic based SVM). The Memetic based SVM provided optimal medical data classification results than other fifteen existing algorithms but its computational complexity is high. Hybridized harmony search and Pareto optimization method were proposed by Dash [37] for the selection of high dimensional features from the medical data. In both feature subset prediction and sample classification, the hybridized method provided high potentiality results for high dimensional databases. For medical data classification, Fan et al. [38] proposed a Hybridized model of fuzzy decision tree and integrating case based clustering method. The authors collected both breast cancer Wisconsin and liver disorder datasets from the UCI machine learning repository. The dataset pre-processed by case based clustering method and fuzzy decision tree is applied for disease identification [39,40]. These methods provided the average forecasting accuracy of 94.4% for breast cancer and 81.6% for liver disorders when compared to existing methods.

III. PROPOSED METHODOLOGY:

In this section, we design the proposed algorithm for medical data classification. For this work, we have chosen four disease datasets such as heart disease, liver disease, cancer disease and lung disease from UIC machine learning repository [41-45]. Initially, the medical dataset is pre-processed and the high dimensionality features are reduced using method. The proposed framework of medical data classification is explained in Fig 1 and the step by step process of proposed work is explained as follows:

3.2 Feature reduction using Discriminant Independent Component Analysis (DICA):

The Negentropy maximization obtains the independent features and lower dimensions with multivariate data in DICA (Discriminant Independent Component Analysis) method. Simultaneously maximize the sum of the marginal Nagentrophy of independent extracted features and Fisher criterion in DICA. In order to develop a better classification, the DICA combines the properties of Independent Component Analysis (ICA) [46,47].

4.3.1. Nagentropy maximization for independent feature extraction:

Thereafter feature selection, the high dimensional features of medical data are reduced by means of DICA method. For non-Gaussian random variables, Nagentrophy is the better statistical scale and the marginal Nagentropy approximation is given as below:

$$R(x_i) \approx k_1(H(M^1(x_i)))^2 + k_2(H(M^1(x_j)))^2 - H(M^2(\sigma)))^2 \quad (1)$$

From the above equation, x_i is the standard deviation and the same mean with the univariate Gaussian distribution is σ. For random vector x_i, the below equation is used to prove the random vectors M^1 and M^2. Where, $k_1 = 36/(8\sqrt{3}-9)$ and $k_2 = 16\sqrt{3}-27)$. The Lagrange formula in equation (5) explains the unit covariance with maximization sum of the marginal Negentropy.

$$\hat{L}(N) = \sum_{i=1}^{R}\left[H(M(n_i^T z)) - H(M(\sigma))\right]^2 + \sum_{i=1}^{R}\alpha_i(n_j^T n_j - 1) \quad (2)$$

From equation (5), the target function maximization is received by the features. The Lagrange formula in equation defines the independent extracted features Nagentropy and the functional criterion of classification performance is maximized by an optimization problem [48-52].

IV. RESULT

The proposed work performance for medical data classification is evaluated in this section. Moreover, MATLAB 2016a with an i5 processor and 4GB RAM was used to actualize the performance of proposed work.

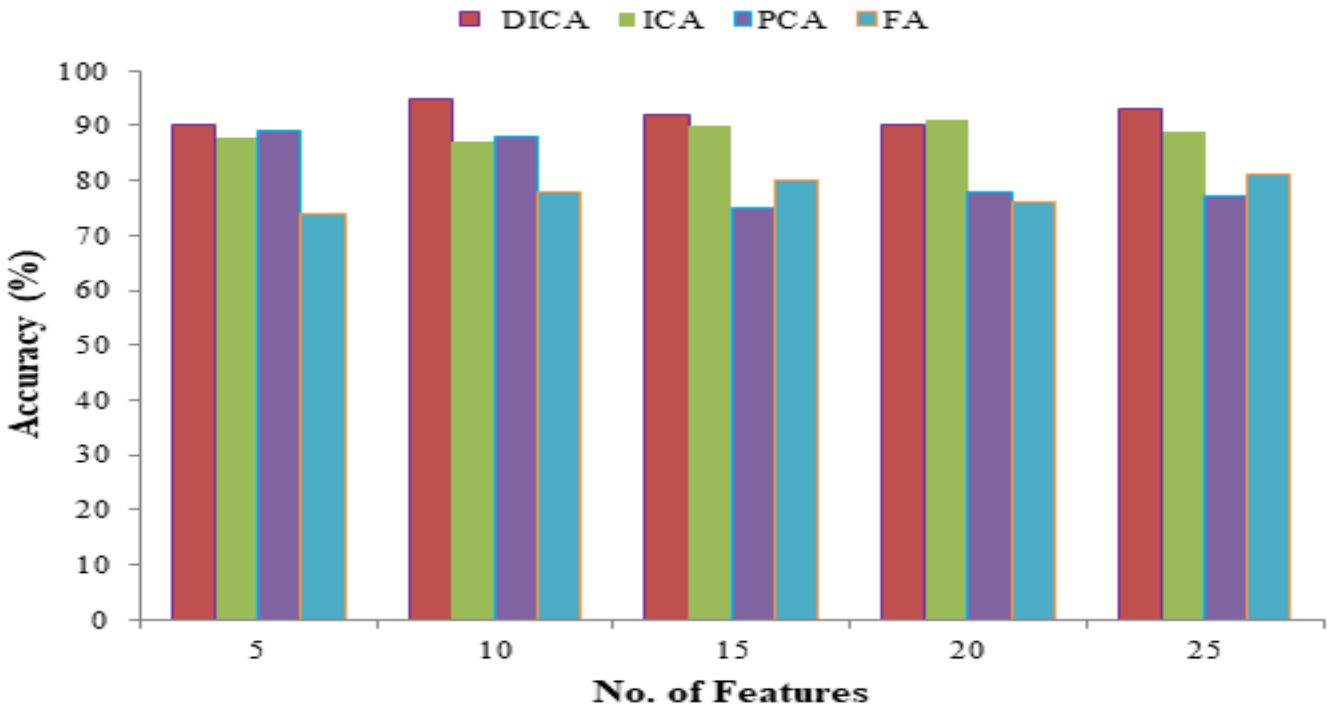

Fig 1: Number of selected features with state-of-art results

Fig 1 explains the convergence performance of the proposed optimization algorithm. Hence, the convergence performance is one of the important methods to validate the performance of optimization algorithms. The convergence performances increase the accuracy level and performance. So, we use state-of-art convergence performance analysis.

V. CONCLUSION

The medical data details are collected from the UCI machine learning repository. The convergence performance of proposed DT-SWO provides better convergence performances than other algorithms such as PSO, SWO, MFO and DT. The feature reduction performances of DICA is higher than other methods such as FA and ICA.

REFERENCES

[1] Chen, Y. T., Chen, C. H., Wu, S., & Lo, C. C. (2019). A two-step approach for classifying music genre on the strength of AHP weighted musical features. Mathematics, 7(1), 19.

[2] Elhoseny, M., Shankar, K., &Uthayakumar, J. (2019). Intelligent diagnostic prediction and classification system for chronic kidney disease. Scientific reports, 9(1), 1-14.

[3] Sivakumar, P., Velmurugan, S. P., & Sampson, J. (2020). Implementation of differential evolution algorithm to perform image fusion for identifying brain tumor.

[4] Khamparia, A., Gupta, D., Nguyen, N. G., Khanna, A., Pandey, B., & Tiwari, P. (2019). Sound classification using convolutional neural network and tensor deep stacking network. IEEE Access, 7, 7717-7727.

[5] Jansirani, A., Rajesh, R., Balasubramanian, R., & Eswaran, P. (2011). Hi-tech authentication for pslette images using digital signature and data hiding. Int. Arab J. Inf. Technol., 8(2), 117-123.

[6] Jain, R., Gupta, D., & Khanna, A. (2019). Usability feature optimization using MWOA. In International conference on innovative computing and communications (pp. 453-462). Springer, Singapore.

[7] Shankar, K., & Lakshmanaprabu, S. K. (2018). Optimal key based homomorphic encryption for color image security aid of ant lion optimization algorithm. International Journal of Engineering & Technology, 7(9), 22-27.

[8] Lyu, L., & Chen, C. H. (2020, July). Differentially Private Knowledge Distillation for Mobile Analytics. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1809-1812).

[9] Poonkuntran, S., Rajesh, R. S., & Eswaran, P. (2011). Analysis of difference expanding method for medical image watermarking. In International Symposium on Computing, Communication, and Control (ISCCC 2009) (Vol. 1, pp. 31-34).

[10] Sampson, J., & Velmurugan, S. P. (2020, March). Analysis of GAA SNTFT with Different Dielectric Materials. In 2020 5th International Conference on Devices, Circuits and Systems (ICDCS) (pp. 283-285). IEEE.

[11] Elhoseny, M., Bian, G. B., Lakshmanaprabu, S. K., Shankar, K., Singh, A. K., & Wu, W. (2019). Effective features to classify ovarian cancer data in internet of medical things. Computer Networks, 159, 147-156.

[12] Gochhayat, S. P., Kaliyar, P., Conti, M., Tiwari, P., Prasath, V. B. S., Gupta, D., & Khanna, A. (2019). LISA: Lightweight context-aware IoT service architecture. Journal of cleaner production, 212, 1345-1356.

[13] Dutta, A. K., Elhoseny, M., Dahiya, V., & Shankar, K. (2020). An efficient hierarchical clustering protocol for multihop Internet of vehicles communication. Transactions on Emerging Telecommunications Technologies, 31(5), e3690.

[14] Anand Nayyar, Vikram Puri, Nhu Gia Nguyen, BioSenHealth 1.0: A Novel Internet of Medical Things (IoMT) Based Patient Health Monitoring System, Lecture Notes in Networks and Systems. Springer, 2019

[15] Shankar, K., Lakshmanaprabu, S. K., Khanna, A., Tanwar, S., Rodrigues, J. J., & Roy, N. R. (2019). Alzheimer detection using Group Grey Wolf Optimization based features with convolutional classifier. Computers & Electrical Engineering, 77, 230-243.

[16] Paramathma, M. K., Pravin, A. C., Rajarajan, R., & Velmurugan, S. P. (2019, April). Development and Implementation of Efficient Water and Energy Management System for Indian Villages. In 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) (pp. 1-4). IEEE.

[17] Chen, C. H., Song, F., Hwang, F. J., & Wu, L. (2020). A probability density function generator based on neural networks. Physica A: Statistical Mechanics and its Applications, 541, 123344.

[18] Kathiresan, S., Sait, A. R. W., Gupta, D., Lakshmanaprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognition Letters.

[19] Gupta, D., & Ahlawat, A. K. (2016). Usability determination using multistage fuzzy system. Procedia Comput Sci, 78, 263-270.

[20] Amira S. Ashour, Samsad Beagum, Nilanjan Dey, Ahmed S. Ashour, Dimitra Sifaki Pistolla, Gia Nhu Nguyen, Dac-Nhuong Le, Fuqian Shi (2018), Light Microscopy Image De-noising using Optimized LPA-ICI Filter, Neural Computing and Applications, Vol.29(12), pp 1517–1533, Springer, ISSN: 0941-0643. (SCIE IF 4.664, Q1)

[21] Pan, M., Liu, Y., Cao, J., Li, Y., Li, C., & Chen, C. H. (2020). Visual Recognition Based on Deep Learning for Navigation Mark Classification. IEEE Access, 8, 32767-32775.

[22] Chen, C. H., Hwang, F. J., & Kung, H. Y. (2019). Travel time prediction system based on data clustering for waste collection vehicles. IEICE TRANSACTIONS on Information and Systems, 102(7), 1374-1383.

[23] Shankar, K., & Eswaran, P. (2015). ECC based image encryption scheme with aid of optimization technique using differential evolution algorithm. Int J Appl Eng Res, 10(55), 1841-5.

[24] Anand Nayyar, Vikram Puri, Nhu Gia Nguyen, Dac Nhuong Le, Smart Surveillance Robot for the Real Time Monitoring and Control System in Environment and Industrial Applications, Advances in Intelligent System and Computing, pp 229-243, Springer

[25] Le Nguyen Bao, Dac-Nhuong Le, Gia Nhu Nguyen, Vikrant Bhateja, Suresh Chandra Satapathy (2017), Optimizing Feature Selection in Video-based Recognition using Max-Min Ant System for the Online Video Contextual Advertisement User-Oriented System, Journal of Computational Science, Elsevier ISSN: 1877-7503. Vol.21, pp.361-370. (SCIE IF 2.502, Q1)

[26] Chakchai So-In, Tri Gia Nguyen, Gia Nhu Nguyen: Barrier Coverage Deployment Algorithms for Mobile Sensor Networks. Journal of Internet Technology 12/2017; 18(7):1689-1699.

[27] Le, D.-N.a, Kumar, R.b, Nguyen, G.N., Chatterjee, J.M.d, Cloud Computing and Virtualization, DOI: 10.1002/9781119488149, Wiley.

[28] Bhateja, V., Gautam, A., Tiwari, A., Nhu, N.G., Le, D.-N, Haralick features-based classification of mammograms using SVM, Advances in Intelligent Systems and Computing, Volume 672, 2018, Pages 787-795.

[29] Khamparia, A., Saini, G., Gupta, D., Khanna, A., Tiwari, S., & de Albuquerque, V. H. C. (2020). Seasonal crops disease prediction and classification using deep convolutional encoder network. Circuits, Systems, and Signal Processing, 39(2), 818-836.

[30] Uthayakumar, J., Elhoseny, M., & Shankar, K. (2020). Highly Reliable and Low-Complexity Image Compression Scheme Using Neighborhood Correlation Sequence Algorithm in WSN. IEEE Transactions on Reliability.

[31] Huyen, D.T.T., Binh, N.T., Tuan, T.M., Nguyen, G.N, Dey, N., Son, L.H, Analyzing trends in hospital-cost payments of patients using ARIMA and GIS: Case

study at the Hanoi Medical University Hospital, Vietnam, Journal of Medical Imaging and Health Informatics, 7(2), pp. 421-429.

[32] Alam, Rahman and Rahman, "A Random Forest based predictor for medical data classification using feature ranking," Informatics in Medicine Unlocked, vol.15, pp.100180, 2019.

[33] Khanmohammadi and Chou, "A Gaussian mixture model based discretization algorithm for associative classification of medical data," Expert Systems with Applications, vol.58, pp.119-129, 2016.

[34] Gorzalczany and Rudziński, "Interpretable and accurate medical data classification–a multi-objective genetic-fuzzy optimization approach," Expert Systems with Applications, vol.71, pp.26-39, 2017.

[35] Bania and Halder, "R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data," Computer Methods and Programs in Biomedicine, vol.184, pp.105122, 2020.

[36] Baliarsingh and Ding, Vipsita and Bakshi, "A memetic algorithm using emperor penguin and social engineering optimization for medical data classification," Applied Soft Computing, vol.85, pp.105773, 2019.

[37] Dash, ""n adaptive harmony search approach for gene selection and classification of high dimensional medical data," Journal of King Saud University-Computer and Information Sciences, 2018.

[38] Fan, Chang, Lin and Hsieh, "A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification," Applied Soft Computing, vol.11, no.1, pp.632-644, 2011.

[39] Van, V.N., Chi, L.M., Long, N.Q., Nguyen, G.N., Le, D.-N, A performance analysis of openstack open-source solution for IaaS cloud computing, Advances in Intelligent Systems and Computing, 380, pp. 141-150.

[40] Shankar, K., & Eswaran, P. (2016, January). A new k out of n secret image sharing scheme in visual cryptography. In 2016 10th International Conference on Intelligent Systems and Control (ISCO) (pp. 1-6). IEEE.

[41] Dey, N., Ashour, A.S., Chakraborty, S., Le, D.-N., Nguyen, G.N, Healthy and unhealthy rat hippocampus cells classification: A neural based automated system for Alzheimer disease classification, Journal of Advanced Microscopy Research, 11(1), pp. 1-10

[42] Velmurugan, S. P., & Rajasekaran, P. S. M. P. (2017). CLASSIFICATION OF BRAIN TUMOR USING MULTIMODAL FUSED IMAGES AND PNN. International Journal of Pure and Applied Mathematics, 115(6), 447-457.

[43] Shankar, K., Elhoseny, M., Perumal, E., Ilayaraja, M., & Kumar, K. S. (2019). An Efficient Image Encryption Scheme Based on Signcryption Technique with Adaptive Elephant Herding Optimization. In Cybersecurity and Secure Information Systems (pp. 31-42). Springer, Cham.

[44] Wu, L., Chen, C. H., & Zhang, Q. (2019). A mobile positioning method based on deep learning techniques. Electronics, 8(1), 59.

[45] Lydia, E. L., Kumar, P. K., Shankar, K., Lakshmanaprabu, S. K., Vidhyavathi, R. M., & Maseleno, A. (2020). Charismatic document clustering through novel K-Means non-negative matrix factorization (KNMF) algorithm using key phrase extraction. International Journal of Parallel Programming, 48(3), 496-514.

[46] Sujitha, B., Parvathy, V. S., Lydia, E. L., Rani, P., Polkowski, Z., & Shankar, K. (2020). Optimal deep learning based image compression technique for data transmission on industrial Internet of things applications. Transactions on Emerging Telecommunications Technologies, e3976.

[47] Lo, C. L., Chen, C. H., Hu, J. L., Lo, K. R., & Cho, H. J. (2019). A fuel-efficient route plan method based on game theory. Journal of Internet Technology, 20(3), 925-932.

[48] Kung, H. Y., Chen, C. H., Lin, M. H., & Wu, T. Y. (2019). Design of Seamless Handoff Control Based on Vehicular Streaming Communications. Journal of Internet Technology, 20(7), 2083-2097.

[49] Elhoseny, M., & Shankar, K. (2019). Reliable data transmission model for mobile ad hoc network using signcryption technique. IEEE Transactions on Reliability.

[50] Shanmugam, P., Rajesh, R. S., & Perumal, E. (2008, May). A reversible watermarking with low warping: an application to digital fundus image. In 2008 International Conference on Computer and Communication Engineering (pp. 472-477). IEEE.

[51] Shankar, K., & Elhoseny, M. (2019). Trust Based Cluster Head Election of Secure Message Transmission in MANET Using Multi Secure Protocol with TDES. J. UCS, 25(10), 1221-1239.

[52] Parvathy, V. S., Pothiraj, S., & Sampson, J. (2020). Optimal Deep Neural Network model based multimodality fused medical image classification. Physical Communication, 101119.

[53] Subbiah Parvathy, V., Pothiraj, S., & Sampson, J. (2020). A novel approach in multimodality medical image fusion using optimal shearlet and deep learning. International Journal of Imaging Systems and Technology.

[54] Parvathy, V. S., & Pothiraj, S. (2019). Multi-modality medical image fusion using hybridization of binary crow search optimization. Health Care Management Science, 1-9.

[55] Velmurugan, S. P., Sivakumar, P., & Rajasekaran, M. P. (2018). Multimodality image fusion using centre-based genetic algorithm and fuzzy logic. International Journal of Biomedical Engineering and Technology, 28(4), 322-348.

[56] Chen, C. H. (2018). An arrival time prediction method for bus system. IEEE Internet of Things Journal, 5(5), 4231-4232.

[57] Shankar, K., Perumal, E., & Vidhyavathi, R. M. (2020). Deep neural network with moth search optimization algorithm based detection and classification of diabetic retinopathy images. SN Applied Sciences, 2(4), 1-10.

[58] Mohanty, S. N., Ramya, K. C., Rani, S. S., Gupta, D., Shankar, K., Lakshmanaprabu, S. K., & Khanna, A. (2020). An efficient Lightweight integrated Blockchain (ELIB) model for IoT security and privacy. Future Generation Computer Systems, 102, 1027-1037.

[59] Elhoseny, M., & Shankar, K. (2020). Energy efficient optimal routing for communication in VANETs via clustering model. In Emerging Technologies for Connected Internet of Vehicles and Intelligent Transportation System Networks (pp. 1-14). Springer, Cham.

[60] Chen, C. H. (2020). A cell probe-based method for vehicle speed estimation. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 103(1), 265-267.

[61] Khamparia, A., Singh, A., Anand, D., Gupta, D., Khanna, A., Kumar, N. A., & Tan, J. (2018). A novel deep learning-based multi-model ensemble method for the prediction of neuromuscular disorders. Neural computing and applications, 1-13.

[62] Shankar, K., Zhang, Y., Liu, Y., Wu, L., & Chen, C. H. (2020). Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification. IEEE Access, 8, 118164-118173.