IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DETECTION OF INHERITED RETINAL DISEASES IN INFANTS USING MACHINE LEARNING

[1]Dr.P.Nagendra Kumar, [2]P.Rishitha

Abstract: Inherited retinal diseases cause several visual deficits in children. They often cause blindness in childhood. Machine learning is one of the most widely used concepts around the world. It will be essential in the healthcare sector as it is helpful for doctors to fasten the diagnosis.

The objective of this paper is to build a Machine learning model for detection of genetic eye diseases. Based on the data obtained from pupillometer, the model predicts the presence of an eye disease. The proposed work deals with neural networks, a machine learning concept. We are implementing two algorithms-SVM (an existing model) and the other is a new way using ELM (Extreme Learning Machine) algorithm. We analyse and predict the result whether the patient has a retinal disease or no disease using this algorithm. ELM algorithm gives more accuracy for the prediction. This prediction will make the detection of disease faster. This model can be helpful to the medical practitioners at their clinic as a decision support system.

***Index Terms* -Inherited disease, visual defects, infants, SVM, ELM,machine learning.**

I. INTRODUCTION

Inherited retinal diseases (IRDs) refer to a group of genetic disorders that affect the retina, leading to vision impairment and potentially causing blindness, particularly in infants. Early detection and accurate diagnosis of IRDs are crucial for timely intervention and management of these conditions. Machine learning techniques have shown promise in various medical applications, including the field of ophthalmology, and offer a potential solution for the automated detection of IRDs in infants.

This research paper aims to develop a machine learning-based approach for the detection of IRDs in infants using Support Vector Machines (SVM) and Extreme Learning Machines (ELM). SVM and ELM are popular algorithms for classification tasks and have been successfully applied in various medical image analysis tasks. By leveraging these algorithms, we aim to analyze retinal images and accurately classify them into different IRD categories

The proposed methodology involves several stages, starting with the collection of a comprehensive dataset consisting of retinal images from infants diagnosed with different IRDs. Preprocessing techniques will be applied to enhance the quality of the images and extract relevant features. The extracted features will then be used as inputs to train and optimize the SVM and ELM models. The models will undergo rigorous evaluation using cross-validation and performance metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).

The outcomes of this research project are expected to contribute to the early detection and diagnosis of IRDs in infants, enabling timely intervention and personalized treatment plans. The use of machine learning algorithms like SVM and ELM can potentially provide a reliable and efficient tool for ophthalmologists and healthcare professionals to screen infants for IRDs, facilitating prompt referral for specialized care. Furthermore, this study may pave the way for the development of automated systems that can aid in large-scale screening programs, reducing the burden on healthcare resources and improving the overall management of IRDs.

II. LITERATURE REVIEW

"Pupil response derived from outer and inner retinal photoreception are normal in patients with hereditary optic neuropathy": The study compared pupil responses in patients with hereditary optic neuropathy (HON) and healthy controls. The results showed that the pupil responses derived from outer and inner retinal photoreception were not quantitatively different between HON patients and controls, indicating that mild-to-moderate visual dysfunction does not affect these responses. However, a correlation was found between visual field loss and the intensity of cone response, suggesting impaired pupil light reflexes in advanced stages of the disease.

"Toward a clinical protocol for assessing rod, cone, and melanopsin contributions to the human pupil response": This paper aimed to establish a clinical protocol for evaluating the contributions of different photoreceptors to the pupillary light reflex (PLR). By conducting experiments with specific stimuli and conditions, the researchers determined optimal durations and intensity levels for assessing the health of rod, cone, and melanopsin pathways. The developed protocol was tested on both healthy individuals and patients with retinal diseases, providing valuable insights into evaluating PLR in clinical settings.

"Learning-based approach to segment pigment signs in fundus images for Retinitis Pigmentosa analysis": This study focused on automating the segmentation of pigment signs in retinal fundus images, which is crucial for diagnosing and monitoring Retinitis Pigmentosa. The researchers proposed a supervised segmentation technique that utilized ensemble classifiers trained on pre-processed images. They collected a large dataset of retinal images with manually segmented pigment signs, making it publicly available. The performance of the classifiers was evaluated on this dataset, demonstrating the effectiveness of the machine learning approach for automated analysis of Retinitis Pigmentosa.

"Do we need hundreds of classifiers to solve real-world classification problems?": This paper evaluated 179 classifiers from different families to determine the best performers in solving classification problems. The researchers assessed the classifiers on a diverse set of 121 datasets, including real-world problems. The results showed that random forest (RF) versions and support vector machines (SVM) with Gaussian and polynomial kernels were among the top-performing classifiers, with random forest being the most successful overall.

"A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble": This study focused on developing a clinical decision support system (CDSS) for predicting diabetic retinopathy (DR). The researchers analyzed data from over 1.4 million diabetic patients and used demographic and lab data to detect the susceptibility to retinopathy. By incorporating multiple data preparation and modeling steps, including a novel "confidence margin" ensemble technique, the CDSS achieved high accuracy in predicting DR. The study provided insights into the risk factors for DR and highlighted the potential of AI in improving early diagnosis and compliance with retinopathy screenings.

"Current state and future prospects of artificial intelligence in ophthalmology: A review": This review article discussed the current status and future potential of artificial intelligence (AI) in ophthalmology. It emphasized the utility of AI in diagnosing and treating various eye conditions, including corneal ectasias, glaucoma, age-related macular degeneration, and diabetic retinopathy. The paper aimed to provide an overview of AI concepts and developments relevant to the field of ophthalmology, addressing the need for bridging the gap between computer science and medical professionals.

III. PROPOSED METHODOLOGY

In this paper we propose a novel machine learning model to achieve this task.Machine Learning algorithms are totally subject to data since it is the most vital perspective that makes model training possible. On the other hand, if won't be able to make sense out of that data, before feeding it to ML algorithms, a machine will be useless. In straightforward words, we generally need to take care of the right data for example the data in the right scale, group, and containing important features, for the problem we need a machine to solve.This makes data preparation the most important step in the ML process. Data preparation defined as the procedure that makes our dataset more appropriate to work with in the ML process.

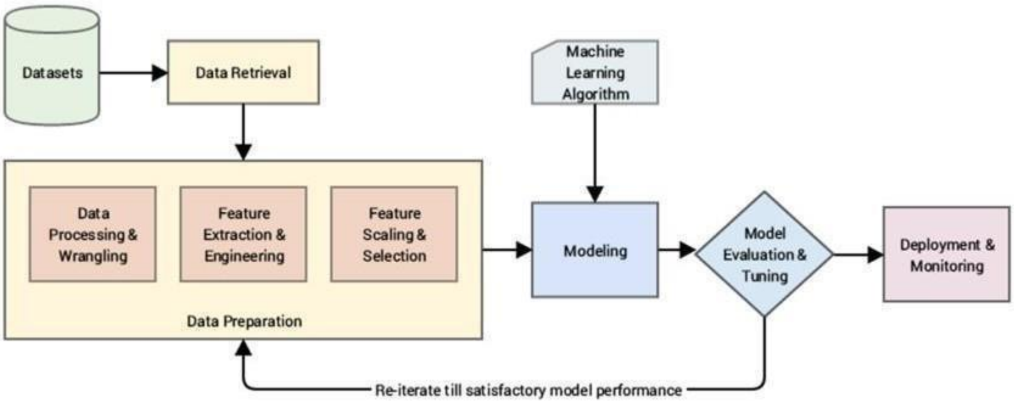

Fig 1: Workflow of the system

A dataset can be viewed as a collection of data objects, which are often also called as records, points, vectors, patterns, events, cases, samples, observations or entities.
Data objects are described by a number of features that capture the basic characteristics of an object, such as the time at which an event occurred, etc... Features are often called as variables, characteristics, fields, attributes or dimensions.
A feature is an individual measurable property or characteristic of a phenomenon beingobserved features can be categorical (nominal, Ordinal), Numerical (Interval, ratio).

DATASET:

Machine learning heavily depends on data and dataset makes machine learning training feasible. A dataset is used to train the model for performing various actions, to work automatically. The training dataset is a dataset in which machine learning algorithms have been trained and the dataset we use to validate the accuracy of our model is called testing dataset.
The raw data for detection of inherited retinal diseases is the collection of historical data that includes a variety of important attributes like max, min, delta, latency, ch etc. These attributes correspond to maximum diameter at baseline, minimum diameter corresponding to the peak constriction, difference between max and min, delay between stimulus and onset of the pupillary constriction, percentage maximum constriction and so on.
The dataset was collected from Kaggle. Our dataset is named "Pupillometric data" which has files in 'asc' format. This file will contain text containing ASCII characteristics. Our dataset contains 593 records of data. Kaggle provides you with preprocessed data. The dataset won't provide you any background information of the data. By using this dataset, the preprocessing of data, training and testing is done.

DATA PREPROCESSING

Data preprocessing is an information mining strategy that is used to change the unrefined information in an accommodating and appreciable format. After selecting the raw data for ML training, the important task is data preprocessing. In broader sense, data preprocessing will convert the selected data into a form we can work with or can feed to ML algorithms. We always need to preprocess data so that it can be as per the exception of the machine learning algorithm. In other words, features of the data can now be easily interpreted by the algorithm. Kaggle provides you preprocessed dataset. But how this data is preprocessed is discussed below.

Steps involved in Data Preprocessing:

Data Filtering
Feature Extraction
Feature Reduction

Data Filtering:

Because data is collected from multiple sources which are in different formats, more than half out time is fed in dealing with the data quality issues when working on a machine learning problem. It is simply unrealistic to expect that the data will be perfect. There may be problems due to human error, limitations of measuring devices, or flaws in the data collection process.
Let's cover over some of them and techniques to deal with them

Missing Values:

It is common to have missing values in your dataset. It is ordinary to have missing values on your dataset. It may have occurred at some stage in data collection or perhaps due to some data validation rule, however regardless missing values need to be taken into consideration.

Eliminate rows with a missing data: Simple and sometimes effective strategy fails if many objects have missing values. If a feature has mostly missing values, then that feature itself can also be eliminated.
Estimate Missing Values: If only a reasonable percentage of values are missing, then we can also run simple interpolation methods to fill in those values. However, the most common method we have used to deal with missing values is by filling them in with the mean, median or mode value of the respective feature.
Inconsistent values:

We know that data can contain inconsistent values. Most probably we have faced this issue at some point. For instance, the 'size of code' field contains the numerical value'. It may be due to human error or maybe the information was misread while being scanned from a handwritten form.
It is therefore always advised to perform data assessment like knowing what the data type of the features should be and whether it is the same for allthe data objects.
Duplicate Values:

A dataset may include data objects which are duplicates of one another. It may happen when the same person submits a form more than once. The term deduplication is often used to refer to the process of dealing with duplicates. In most cases, the duplicates are removed so as to not give that particular data object an advantage or bias, when running machine learning algorithms.
Feature Extraction

In this step, we extract the features and its data from the pupillometric data. All the features along with their corresponding values are extracted.
Feature Reduction
Feature reduction is a very common method for selecting a subset of the dataset that we are analyzing. In most cases, working with the complete dataset can turn out to be too expensive considering the memory and time constraints. So, by determining few useful attributes, we can reduce the size of the dataset to a point where we can use a better, but more expensive, machine learning algorithm. The key principle here is that the reduction should be done in such a manner that the sample generated should have approximately the same properties as the original dataset, meaning that the sample is representative. This involves choosing the correct sample size and sampling strategy.
Train / Validation / Test Split
After the above steps are done, our dataset is ready for the exciting machine learning algorithms. But before we start deciding the algorithm which should be used, it is always advised to split the dataset into 2 or sometimes 3 parts. Machine Learning algorithms, or any algorithm for that matter, has to be first trained on the data distribution available and then validated and tested, before it can be deployed to deal with real-world data.
Training data: This is the part on which your machine learning algorithms are actually trained to build a model. The model tries to learn the dataset and its various characteristics.
Validation data: This is the part of the dataset which is used to validate our various model fits. In simpler words, we use validation data to choose and improve our model hyper parameters. The model does not learn the validation set but uses it to get to a better state of hyper parameters.
Test data: This part of the dataset is used to test our model hypothesis. It is left untouched and unseen until the model and hyper parameters are decided, and only after that the model is applied on the test data to get an accurate measure of how it would perform when deployed on real-world data.

ALGORITHMS:
SVM algorithm:
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.
The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

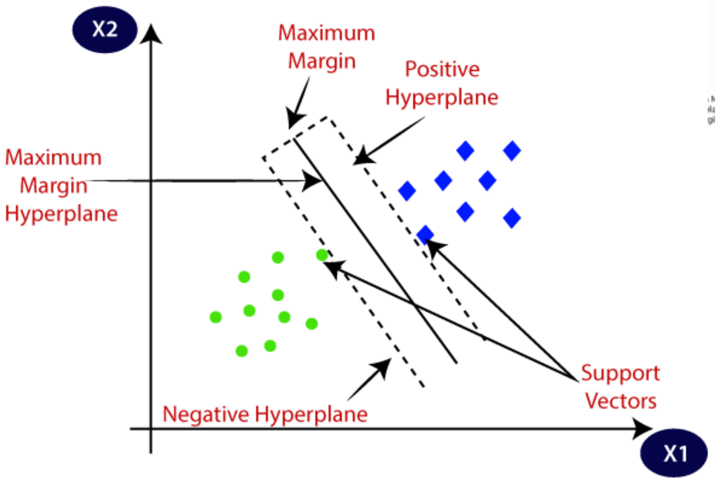

Fig: 3.3 Classification using SVM

The followings are important concepts in SVM −

Support Vectors: Data points that are closest to the hyperplane is called support vectors. Separating line will be defined with the help of these data points.
Hyperplane: We can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
Margin: It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

Advantages of SVM:
SVM works relatively well when there is a clear margin of separation between classes.SVM is more effective in high dimensional spaces.SVM is effective in cases where the number of dimensions is greater than the number of samples.
SVM is relatively memory efficient.

ENSEMBLE MODEL:
Ensemble vote classifier is a type of machine learning model that combines the predictions of multiple individual models to make a final prediction. This approach is based on the idea that combining the predictions of multiple models can lead to better overall performance than relying on a single model.
In an ensemble vote classifier, each individual model produces a prediction, and the final prediction is made by aggregating the individual predictions using a voting scheme. The most common voting schemes are:
Majority voting: The final prediction is the class that receives the most votes from the individual models.
Weighted voting: Each individual model is assigned a weight based on its performance on a validation set, and the final prediction is the weighted sum of the individual predictions.

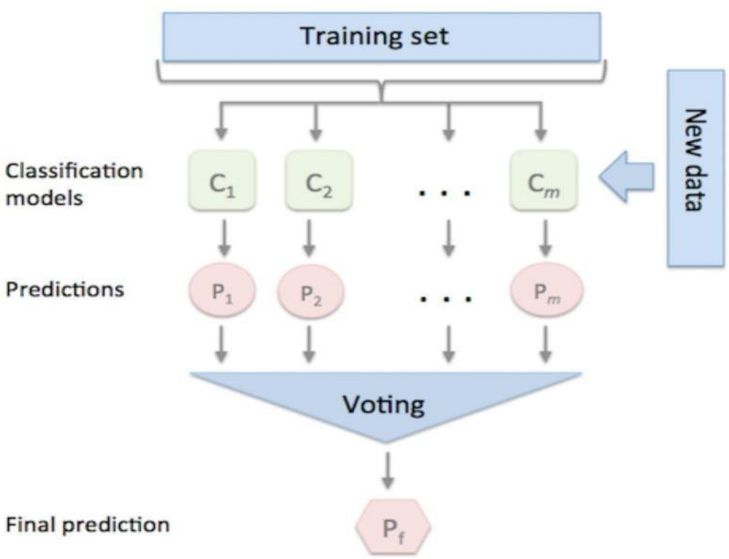

Fig: 3.6 Working of Ensemble Model

EXTREME LEARNING MACHINE (ELM) ALGORITHM:

Extreme Learning Machine (ELM) is a type of neural network algorithm that is known for its fast-training speed and good generalization performance. It is a single-layer feed forward neural network.An ELM is a single hidden layer neural network, which has better training performance than traditional algorithms. ELM is particularly suitable for large scale datasets and has been used for various machine learning tasks such as classification, regression, feature extraction. ELM is implemented completely automatically without any iterative tuning and no intervention is required from users.

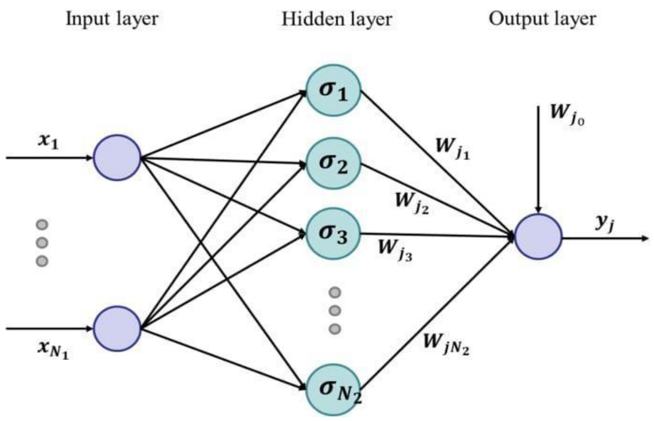

Fig: 3.7 Extreme Learning Machine

Existing System

The clinical evaluation of IRDs is being done using processes which involve many clinical tests.eg: electrophysiological testing.

A machine learning approach that detects IRDs using SVM algorithm has been proposed in the past.

Disadvantages of Existing System

The traditional clinical methods include invasive techniques which are not suitable for infants.These often require sedation of children, which effects their mental health.The existing machine learning approach using SVM takes more computational time for training and testing.It is not effective with larger datasets.It cannot handle over-fitting i.e, model cannot work well if number of attributes are greater than the number of records in data.

Proposed System

Machine Learning basically provides the system with the “ability to learn”.The machine is able to use the training data set and learn and based on the knowledge it takes the best decision.Our proposed system is a machine learning model which uses the concept of neural networks for detecting inherited retinal diseases.The model is developed using ELM (Extreme Learning Machine) algorithm.

Advantages of Proposed System

Predicts the presence of an IRD.The performance of the predicting model is improved.It is easy and takes less time for training and testing.Effectively handles large datasets.

After preprocessing the dataset, the model is trained on the dataset after proper datasplit. A tkinter based GUI is being developed to interact with the machine learning model and to predict the IRDs.

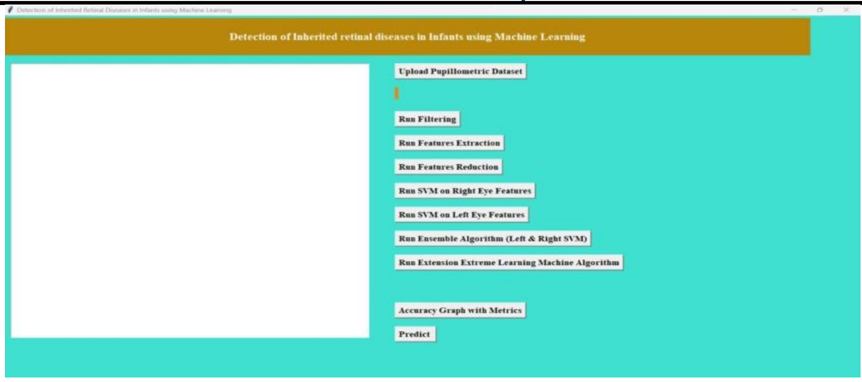

Fig 1. Tkinter Frame of the ML Application

IV. CONCLUSION

In this paper, we detected the presence of inherited retinal disease using ELM algorithm. We performed the entire process on pupillometric data set. We used accuracy, specificity and sensitivity as deciding parameters to come to come to a conclusion. There is a system existing for detection of IRDs which used SVM ensemble model with an accuracy of 92.8. Our developed Machine learning model using ELM algorithm obtained an accuracy of 98.7. Based on the results, we can conclude that ELM algorithm provides an easy and efficient model to detect inherited retinal diseases with pupillometric data.

REFERENCES

1) https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithm_support_vector_machine.html

2) X.-F. Huang, F. Huang, K.-C. Wu, J. Wu, J. Chen, C.-P. Pang, F. Lu, J. Qu, and Z.-B. Jin, ‘‘Genotype–phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing,’’ Genet. Med., vol. 17, no. 4, pp. 271–278, Apr. 2015.

3) A. Kawasaki, S. Collomb, L. Léon, and M. Münch, ‘‘Pupil responses derived from outer and inner retinal photoreception are normal in patients with hereditary optic neuropathy,’’ Exp. Eye Res., vol. 120, pp. 161–166, Mar. 2014.

4) P. Melillo, A. de Benedictis, E. Villani, M. C. Ferraro, E. Iadanza, M. Gherardelli, F. Testa, S. Banfi, P. Nucci, and F. Simonelli, ‘‘Toward a novel medical device based on chromatic pupillometry for screening and monitoring of inherited ocular disease: A pilot study,’’ in Proc. IFMBE, vol. 68, 2019, pp. 387–390

5) https://ieeexplore.ieee.org/document/8998205

6) R. Kardon, S. C. Anderson, T. G. Damarjian, E. M. Grace, E. Stone, and A. Kawasaki, ‘‘Chromatic pupil responses. Preferential activation of the melanopsin-mediated versus outer photoreceptor- mediated pupil light reflex,’’ Ophthalmology, vol. 116, no. 8, pp. 1564–1573, 2009.

7) S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, ‘‘Supervised machine learning: A review of classification techniques,’’ Emerg. Artif. Intell. Appl. Comput. Eng., vol. 160, pp. 3–24, Jun. 2007.

8) A. J. Roman, S. B. Schwartz, T. S. Aleman, A. V. Cideciyan, J. D. Chico, E. A. M. Windsor, L.

9) M. Gardner, G.-S. Ying, E. E. Smilko, M. G. Maguire, and S. G. Jacobson, ‘‘Quantifying rod photoreceptormediated vision in retinal degenerations: Dark-adapted thresholds as outcome measures,’’ Exp. Eye Res., vol. 80, no. 2, pp. 259–272, Feb. 2005

10) Y. J. Yoo, H. K. Yang, and J.-M. Hwang, ‘‘Efficacy of digital pupillometry for diagnosis of Horner syndrome,’’ PLoS ONE, vol. 12, no. 6, Jun. 2017, Art. no. e0178361.

11) B. Lorenz, E. Strohmayr, S. Zahn, C. Friedburg, M. Kramer, M. Preising, and K. Stieger, ‘‘Chromatic pupillometry dissects function of the three different light-sensitive retinal cell populations in RPE65 deficiency,’’ Invest. Ophthalmol. Vis. Sci., vol. 53, no. 9, p. 5641, Aug. 2012.

12) R. Kapoor, S. P. Walters, and L. A. Al-Aswad, ‘‘The current state of artificial intelligence in ophthalmology,’’ Surv. Ophthalmol., vol. 64, no. 2, pp. 233–240, Mar. 2019.

13) E. Iadanza, V. Mudura, P. Melillo, and M. Gherardelli, ‘‘An automatic system supporting clinical decision for chronic obstructive pulmonary disease,’’ Health Technol., pp. 1–12, Mar. 2019, doi: 10.1007/s12553-019- 00312-9

14) A. Hall, ‘‘Correlation-based feature selection for machine learning,’’ Ph.D. dissertation, Dept.

15) Comput. Sci., Waikato Univ., Hamilton, New Zealand, 1999

16) https://www.analyticsvidhya.com/blog/2021/12/time-series-forecasting-with-extreme-learning-machines/
17) J. A. ALzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, and C. Thaventhiran, ''Boosted neural network ensemble classification for lung cancer disease diagnosis,'' Appl. Soft Comput., vol. 80, pp. 579–591, Jul. 2019
18) https://www.frontiersin.org/articles/10.3389/fneur.2019.00076/full

19) N. Brancati, M. Frucci, D. Gragnaniello, D. Riccio, V. Di Iorio, L. Di Perna, and F. Simonelli, ''Learning-based approach to segment pigment signs in fundus images for Retinitis Pigmentosa analysis,'' Neurocomputing, vol. 308, pp. 159–171, Sep. 2018.
20) https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/