

Classifying Parkinson Disease using Mixed Weighted Mean

BALA BRAHMESWARA.KADARU 1*, SIVA CHINTAIAH.NARNI 2, Dr.B.RAJA SRINIVASA REDDY.3

¹Research Scholar, Acharya Nagarjuna Univeristy, Nambur Andhra Pradesh, India. baluanuphd@gmail.com.

² Research Scholar, Acharya Nagarjuna Univeristy, Nambur, Andhra Pradesh, India. narni.sivachintaiah@gmail.com.

³CSE DEPARTMENT, RamaChandra College of Engineering, Eluru, Andhra Pradesh, India. brs_121@yahoo.com

Abstract — Parkinson's disease (PD) could be a chronic unhealthiness whose fundamental basic indication embody inability, shake & Bradykinesia. Extra in the direction of a well known impacts metallic element will have an extreme accomplish to oration & vocalize? An oration signs determine ordinarily incontestable through for bearing to metallic element area unit summary voiced intensity, sensations, interruption a tone feature & peculiarly quick tempo to communication. Oration signs are commonly termed as mobility defect in speech. The unwellness will be tough to identify correctly, particularly in the advance moment, owing near to cause. We applied required preprocessing techniques and then, **mixed weight mean classification** technique, a feature selection subset technique on Parkinson's data. By this technique we overcome the problem of miss prediction using IG, GR, GI, correlation & CS. This technique gets the most accurate results for predicting the Parkinson's knowledge.

.Keywords— **monopitch, Dysarthria, knowledge, measure, methodology, Mixed Weight Mean Classification.**

I. INTRODUCTION

Parkinson's is a disease having speech disorder. Once the Parkinson's is attacked, gradually decreases the voice. By using Telemonitoring device we record the voice samples. System & Signal models are applied to detect Parkinson's disease.

NS

NS communicates with neurons of every part of human body. Nervous System will advices the parts of human body based on sensor input data. It consists of Central Nervous System and Peripheral Network System.

Parkinson's disease: Organization

The most typical symptoms of Parkinson's are shaking, Gait, Posture, Stiffness of arm and legs, Bradykinesia, Aphasia. By this disease about 80 to 90 % of people are suffering, but it is not considered as risky for regular life. Initially it may effect either right or left side of the brain. In its advanced stage brain cells will dead and Nervous System will not responds correctly.

Voice related Organs

Voice will be generated by connecting speech organs directly or indirectly.

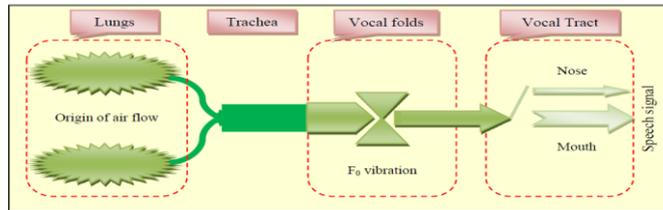


Figure 1: Schematic diagram of the key elements concerned within the production of speech

Respiratory Organ

The RO system includes the lungs and metabolic process. The breathe comprises countless tiny air sacs of lungs, and they helps in anabolism organization during **breathing**, i.e. conversion of O₂ to CO₂ and vice versa for the natural action of cells within the body it relies on air flow on the region.

Vocal Cords

Vocal Cord is placed in Trachea and Larynx plane. Brain stem is the process to monitors the functionality of brain by connecting nervous and muscle systems. It will be differed in size, age and vibrations are based on the gender. To define arena era to compute one fluctuation of the voice fold

$$F_0 \stackrel{\text{def}}{=} 1/T. \quad (1)$$

Vocal Zone (Tract)

Voice Zone is the collection of nose, mouth, tongue and lips. Based on the shape, it increases the frequencies in the swing of vocal cords resonance.

..sustained Vowels fluctuation

Utilization of utterance for observing the level of voice will helps to detect Parkinson's. On an average depending on the gender, health and physique a human voice can be maintained about 20 seconds.

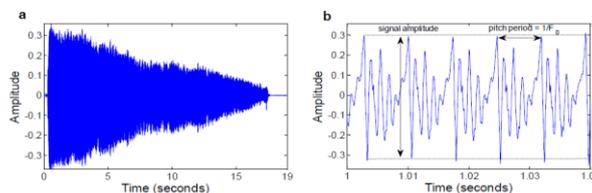


Figure2: Typical sustained vowel /a/ phonation

Life-span changes in physiology

Tone of voices are changed depending on the age because of alterations in larynx, vocal tract size and changes in hormones which tend to reduces the power of nervous system. The tendency is different with males & females with and increasing of the age, for women it will monotonically decreases and the depending on the age decreases in women well lean towards modify by period outstanding towards modify in voice box & tone region volume, physique which be inclined to fix a choice of hormonal transform & compact anxious structure manage. Most likely the mainly feature modify of audio changeable

by the period. The movement be similar into the gender with monotonic falling in women roughly awaiting period 40 & later raise intended for man.

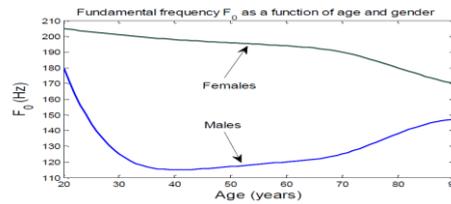


Figure3: Life-span changes of the fundamental frequency

Jitter

Such as outlined because the variation into otherwise dislocation of a few side of a rhythm during elevated-occurrence digital indicator. Variation will exist during conditions of amplitude, section temporal order, otherwise a breadth of the indication beat[7]. Additionally referred to as the tone anxiety it's principally unremarkably found within the sparkles of monitors, etc. It will be explained because a little reflex deviation within a occurrence of neighboring moving sequence of a voiced crinkle. Pathological voices usually exhibit a higher share of noise share of jitter.

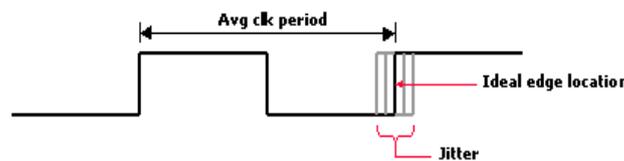


Figure4: Jitter

shimmer:

Jitter was a live to live the anxiety within the tone otherwise occurrence to a signal provide options of a abnormality inside the occurrence. flicker on top of a opposite hand is that the live to denote a abnormality inside a vibration other wise as well can exist say because the vibration anxiety. noise and shimmer square measure measures of the wave rotation difference of first harmonic and vibration that are for the most part used for the outline of pseudo logia fantastical and myth mania tone feature.

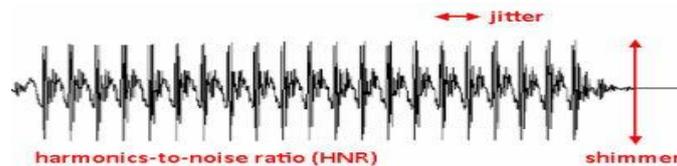


Figure5: Shimmer

Features:

jitter (local)

This constraint mean the common of the distinction among successive amounts separated with a common time. Multi-Dimensional **Voice** Program provides this parameter $1040 \times 10^{-3}\%$ since an entry designed for pathway. The chance of the right threshold is lower as this range was supported disturbance measurements prejudiced with sound

$$\text{Jitter (Seconds)} = \frac{\sum_{i=2}^N |T_i - T_{i-1}|}{(N-1)} \quad (2)$$

Jitter (local, absolute):

This characteristic means the typical of the distinction among successive time. Multi-Dimensional **Voice** Program offers this parameter $83200 \times 10^{-3} \mu s$ like to entry in favor of forensic purposes. like a preceding aspect the likelihood of the right threshold is lower as this variety was supported interference capacity prejudiced by sound.

$$\text{Jitter (Seconds)} = \frac{\sum_{i=2}^N |T_i - T_{i-1}|}{(N-1)} \tag{3}$$

jitter (Relative Average Perturbation):

This be a RAP, a common complete distinction among a amount & also a standard on as well as its 2 neighbors, separated with a common amount. Multi-Dimensional **Voice** Program offers this parameter $680 \times 10^{-3} \% \mu s$ as a entry in favor of pathology. Almost like the previous 2 options the likelihood of the proper entry is lesser as this range was supported noise capacity unfair by sound. RAP is outlined in terms of 3 successive time, as follows.

$$\text{RAP} = \text{absAP(seconds)} / \text{meanPeriod(seconds)} \tag{4}$$

Jitter (PPQ5):

It is just like a JITTER (RAP) characteristic however with five points. In this constraint we have a tendency to see the five- purpose amount anxiety measure, a common complete distinction among a amount as well as the standard of it 4 nearest *contiguous*, separated with a common amount. Multi-Dimensional **Voice** Program provides this constraint $840 \times 10^{-3} \% \mu s$ when an entry meant on behalf of pathway.

$$\text{PPQ5} = \frac{\text{PPQ5(Seconds)}}{\text{meanperiod(seconds)}} \tag{5}$$

jitter (ddp):

We can see during this characteristic of standard complete distinction among successive variations involving repeated time, separated with a typical amount. A worth equals 3 times RAP.

$$\text{DDP} = \frac{\text{absDDP(Seconds)}}{\text{meanperiod(seconds)}} \tag{6}$$

Shimmer (local):

In this feature we are able to see the distinction involving a oscillation a successive time, separated with typical oscillation. MSDP offers a limitation $3810 \times 10^{-3} \%$ while an entry used in favor of path way.

$$\text{Shimmer} = \frac{\text{Average absolute difference}}{\sum_{i=1}^{n-1} A_i} \tag{7}$$

shimmer (local , dB):

This be a typical complete base-10 index of a distinction among a oscillation a successive time, increased with twenty. MSDP provides this parameter 350×10^{-3} decibel while an entry used in favor of path way.

$$DB_{\text{Shimmer}} = \frac{\sum_{i=1}^{n-1} |20 * \log_{10}(A_i/A_{i+1})|}{n - 1} \tag{8}$$

shimmer (Amplitude Perturbation Quotient3):

Here we tend to see a 3 purpose APQ that be that a standard complete distinction among the oscillation a amount &also a standard of a oscillation to *contiguous*, separated with a standard vibrations.

$$APQ3 = \frac{\frac{1}{n-2} \sum_{i=2}^{n-1} | \frac{A_{i-1} + A_i + A_{i+1}}{3} - A_i |}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (9)$$

Shimmer (Amplitude Perturbation Quotient5):

related toward a preceding characteristic to once more we tend to see a 5 point vibration anxiety measure that is that a standard complete distinction among a vibration a amount & also a standard a oscillation to contiguous, separated with a typical vibrations.

$$APQ5 = \frac{\frac{1}{n-4} \sum_{i=5}^{n-2} | \frac{A_{i-2} + A_{i-1} + A_i + A_{i+1} + A_{i+2}}{5} - A_i |}{\frac{1}{n} \sum_{i=1}^n A_i} \quad (10)$$

shimmer (DDP):

A common complete distinction among successive variations among amplitudes to successive time. It can be often Pratt's novel obtain flicker. A worth be thrice Amplitude Perturbation Quotient3.

$$DDP = \frac{abDDP(Seconds)}{mean\ amplitude\ (sconds)} \quad (11)$$

DFA:

We should perceive a conception to a delimited statistic will be mapped to a discussed method. though, a new test study features while related to sort of pattern analysis to physiological information to these statistic area unit usually extremely non-stationary. A basic common meaning describe a statistic as stationary of mean, variance and better instant, similarly because the correlation functions area unit invariant beneath time translation. sign to don't conform these situation area unit non-stationary. The mixing process can any exaggerate the non-stationary of the first information.

$$DFA = 1 / (1 + \exp(-Y)) \quad (12)$$

Harmonic – to - Noise Ratio (HNR) & (NHR)

HNR & NHR are to be ordinarily utilize methods, planning towards specific a quantity a sound within the communication sign. Used force exist thought-about a element of a 3rd giant cluster of speech disorder method that map to characterize the sign exploitation sign toward sound quantitative relation Signal to Noise Ratio move towards

$$HNR\ (dB) = 10 \cdot \log_{10} [R_{xx}(l_{max}) / (1 - R_{xx}(l_{max}))] \quad (13)$$

$$NHR = 10 \cdot [(1 - R_{xx}(l_{max})) / R_{xx}(l_{max})] \quad (14)$$

Pitch Period Entropy (PPE):

Aspire towards measure cycle – to - cycle anxiety , though require a precise, obvious strict meaning that have crystal rectifier to growth of a event various interference modification. Interference are often calculate exploitation any a form, otherwise a reciprocally relative arena quantity.

$$T_0=1/F_0 \text{ contour;}$$

Represent complete distinction of F0 approximation connecting ordered cycles:

$$Jitter_{F_0,abs} = \frac{1}{N} \sum_{i=1}^{N-1} |F_{0,i} - F_{0,i+1}| \tag{15}$$

Shimmer variants

flicker be that a analogue of disturbance in favor of a oscillation to a communication sign, instead of F0 They has use a similar computation bestowed within a earlier part on behalf of a disturbance modification, however mistreatment a oscillation A0 form rather than a F0. Shimmer-variant acoustic live that we have a tendency to compute is flicker into volume (dB), from the time when.

$$Shimmer_{dB} = \frac{1}{N} \sum_{i=1}^{N-1} 20 \cdot \left| \log_{10} \frac{A_{0,i}}{A_{0,i+1}} \right|. \tag{16}$$

Modern Dysphonia Measures

The Modern Dysphonia measure is

R P D E

Recurrence Period Density Entropy deal with flexibility to a voiced covered a continue steady voiced crinkle fluctuation, measure a variation starting precise regularity. A fundamental idea is to the communication sign consists of a (nonlinear) deterministic21 & a random element plus also the technique try in the direction of carry away a final. This live be predicated under the view of return which can exist see like a overview of time.

$$RPDE \equiv \frac{-\sum_i^{T_{max}} p(i) \cdot \ln(p(i))}{\ln(T_{max})}. \tag{17}$$

Fo:

Average fundamental frequency (F0) (i.e.) the common worth of every one calculate amount to amount original values.

Flo:

The lowest harmonic (Flo) (i.e.) very cheap of all extracted amount to amount harmonic values **Highest**

Fhi:

Period : a quantity is that a instance in use for single absolute cycle of a continuation signal figure.

Frequency: the amount of cycles completed per second. The evaluate component in favor of occurrence

$$F_0=1/Period \tag{18}$$

Correlation Dimension (D2):

The D2 that is that the live of the complexness of a settled system offers the amount of freelance changeable essential to explain to categorization presentation. The correlation dimension (D2) is designed through early

down time implant a sign in the direction of rebuild a gap to a not sequential phase space of planned near come up with the communication sign.

Spread1:

Spread1 is the log of the variance of the whitened pitch periods, First I need to calculate Series then I will be using this to extract „F then this is used to get PDX then I used PDX in calculating Spread1 and Spread2. In order to find Series we have a formula.

$$12 * \log_2(x/127.09); \quad (19)$$

Spread2:

Spread2 is the entropy (estimated using histograms) of the whitened pitch periods; the function 'entropy' just calculates the Shannon entropy.

$$\text{spread2} = \text{entropy}(p_{dx}) / \log(\text{length}(p_{dx})) \quad (20)$$

III Related Work

Machine Learning Approaches:

1. ANN:

Artificial Neural Network square measure complex form to might expect, categorize estimated operate otherwise acknowledge model. in theory, artificial neural networks square measure capable in the direction of assess some operate & stipulation use correctly can be apply successfully into PD. Production starting Artificial Neural Netowrk models square measure produce since not sequential combos otherwise enter unpredictable, & such form is successfully develop to upset experimental knowledge habitually determined in PD and to seek out system principal a method as of rare enter records functions. ANNs is depicted as organized neuron, qualified to judge as of input/output. [8]

2. k - NN :

Nearest-Neighbour classifier square measure supported knowledge with similarity, i.e, with scrutiny a known take a look at row a coaching rows to square measure the same as it. A coaching rows square measure delineate near N elements every row correspond to some extent in AN N- measurement area. During this manner, everyone be a coaching row square measure keep in AN N- measurement sample area. Once known AN unidentified row, a k-NN classifier seeks the sample area in favor of a k coaching rows to square measure nearest to the strange row. These k coaching rows square measure the k-NN of a strange row. [4]

3. SVM:

SVM, a capable novel technique designed for the categorization of each linear and nonlinear information. In a very shell, a SVM is AN algorithmic rule to mechanism like pursue. It applies a nonlinear plot to remodel the initial coaching information into the next measurement. inside this novel measurement, it look for a linear optimum extrication hyperplane With AN acceptable nonlinear diagram towards a suitably elevated measurement, information from 2 categories will continually be separated

with a hyperplane. The Support Vector Machine discovers this hyperplane mistreatment maintain vector and margins: [8].

$$\left\{ \begin{array}{l} w^T x_i + b_0 \geq +1, \text{ for } y_i = +1 \\ w^T x_i + b_0 \leq -1, \text{ for } y_i = -1 \end{array} \right\} \Rightarrow y_i \cdot (w^T x_i + b_0) \geq 1, \forall i \in \{1 \dots N\} \quad (21)$$

4. C A R T

CART is the regression and classification method in the data mining. In these methods to construct model predicate data. In this method construct respectively the data can divided into no of parts. The se representation to display tree model. In this model may be considered continues variable and discrete variables. The CART algorithm divided into three categories gini index, towing and regression tree [5].

IV PARKINSON’S DISEASE DATA

The Parkinson’s Data Set [2] is description given below

Tabell: The various attributes for Parkinson’s disease

S.NO	Description	Features
1	voiced essential incidence	Fo, Fh, Flo
2	Measure of deviation to essential incidence	Jitter(%), Jitter(Abs), RAP, PPQ, DDP .
3	Measure of deviation to amplitude	shimmer, shimmer(Db), APQ3, APQ5, APQ, DDA
4	Measure a part of sound towards tonal component into the tone	(NHR) & (HNR)
5	Non linear Dynamic Complexity measures	RPDE ,D2
6	Scaling Exponent	DFA
7	Non linear measure of essential incidence deviation	spread1, spread2, PPE Pitch Period Dimension
8	Health Status 1-parkinson; 0-Healthy	Status

Phon_R01,151.989,163.736,144.148,0.00168,0.00001,0.000068,0.00092,0.00204,0.01064,0.097,0.00522,0.00632,0.00928,0.00928,0.01567,0.00233,29.746,0.334171,-0.67793,-6.9812,0.18455,2.129924,0.106802,1.

Phon_R01,152.848,57.339,132.857,0.00174,0.00001,0.000070,0.00096,0.00225,0.01024,0.093,0.00455,0.00576,0.00993,0.01364,0.00238,29.928,0.311369,0.676066,-6.73915,0.160686,2.296873,0.11513,1.

Phon_R01,151.884,160.267,128.62,0.00183,0.00001,0.00076,0.001,0.00229,0.0103,0.094,0.00469,0.00582,0.00871,0.01406,0.00243,28.409,0.263654,0.691483,-7.11158,0.14478,2.065477,0.093193,1.

V MIXED WEIGHTED MEAN-CLASSIFICATION-C USED method

Information Gain:

IG like a element choice live hire join n stand in favor of otherwise grasp a elements to division D. A element among the very best data get a selected because the cacophonous element in favor of join n. These element reduces the knowledge required towards categorize a rows within the ensuing separation & imitate the smallest amount chance or “degradation” into this division. Such AN move towards reduce a accepted scope of check required a categorize a known row & assurance to a simple hierarchy be establish.[4] The usual data required towards categorize a row into D be specified with

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (22)$$

a lot of additional info would we have a tendency to at rest would like so as to attain a precise categorization? In a quantity be calculated with

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} \times Info(D_j). \quad (23)$$

IG is outlined because the distinction among the first info demand and also the novel condition i.e

$$Gain(A) = Info(D) - Info_A(D). \quad (24)$$

.gain ratio

Data increase by separation happening this feature is greatest. Obviously, such a division is ineffective in favor of categorization. That tries to beat this prejudice. It affect a sort of standardization towards info increase employing a “split information” price outlined analogously by Info(D) as:[4].

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right). \quad (25)$$

This value stand for a possible info produce with cacophonous a coaching knowledge set, D, into v separation, corresponding to the v result of a take a look at on attribute.

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}. \quad (26)$$

chi – square Statistics

The Chi-Square method measure for the Parkinson’s data by evaluate chi-square statistic with respect to disease class distribution. In this method calculates difference between the actual non Parkinson’s disease distributions of classes in attribute with expected Parkinson’s disease distribution class attributes [1].

$$Chi-square = \sum (\text{Observed Parkinson’s Disease}_i - \text{Expected Parkinson’s Disease}_i)^2 / \text{Expected Parkinson’s Disease}_i \quad (27)$$

GI:

Gini index calculates the inequalities of Parkinson’s Data. It is calculates based on the uneven distribution of Data set. When Gini index the values lies between 0 and 1. Where 0 represent perfect Parkinson’s equality. 1 represents not perfect Parkinson’s. Gini index is used as a measure of impurity of a node and commonly when the dependent of the Parkinson’s. This method is considering only categories attributes. The minimum values is zero and maximum values is (1-1/k).Where k is no of categories attributes[1].

correlation:

A feature would like is powerfully associated with another element also are a disused ones. a number of redundant may be notice with association testing. Take 2 elements, such examination will live however powerfully one element imply the opposite, supported the offered knowledge. For statistical elements, we will value the relationship among 2 elements, Parkinson’s and non Parkinson’s, by calculate the correlation. It is used to methodology can exist useful just a information place by mathematical or binominal tag. It can’t exist practical on polynomial elements as a result of the polynomial categories give no info regarding their ordering; thus the weights square measure a lot of or less random looking on the interior numerical illustration of the categories. [6]

VI. Overview of MVM-C Method

For example to consider weather data set how to execute MWM-C process is given below

STEP1:

In this method first calculate the for each attribute gain, gain ratio, chi-square, gini index,

IG (outlook) =247 x10 ⁻³	IG(temperature)=29 x10 ⁻³	IG(humidity) = 152 x10 ⁻³	IG(windy)=48 x10 ⁻³
GR (outlook) =157 x10 ⁻³	GR(temperature)=19 x10 ⁻³	GR(humidity) = 152 x10 ⁻³	GR(windy)=49 x10 ⁻³
GI (outlook) =116 x10 ⁻³	GI(temperature)=19 x10 ⁻³	GI(humidity) = 92 x10 ⁻³	GI(windy)=21 x10 ⁻³
CS (outlook) =247 x10 ⁻³	CS(temperature)=29 x10 ⁻³	CS(humidity) = 152 x10 ⁻³	CS(windy)=48 x10 ⁻³

STEP2:

Take any two measure classifier mean values. In this example we take gain and gain ratio values.

Outlook=202 x10⁻³ temperature=34 x10⁻³ humidity=152 x10⁻³ and windy=485 x10⁻⁴

STEP3

Arrange the attributes in the sorting order and apply the classifier after the result is

Where W is weather, T is temperature, H is humidity, O is outlook

W T H O	4
T H O	3
H O	2
O	1

This process continues until the values are zero or impurity of attributes.

.Algorithm MWM-C

Input: Parkinson’s voice Data Set

Output: Predict the disease

Procedure:

- 1: Take Feature data set of Parkinson’s
- 2: Let P= {F_i |i=1, 2, 3....23} be the Set of Feature attribute for the Parkinson’s.
- 3: do

- 4: Compute all Feature attributes for the IG, GR, CS, GI, Correlation for the Parkinson's and non Parkinson's Classes.
- 5: for each Feature Attribute (Fi) €P Do
- 6: Compute (Fi) =mean of (IG, GR, CS, GI, Cor) for Parkinson's
- 7: end for.
- 8: construct tree
 - 8.1 Feature attribute are arrange sorting order for the Fi according to Parkinson's and Non Parkinson's
 - 8.2 select each Feature attribute is a root node for the majority voting of attribute
 - 8.3 Classify the tree next process continue, Up to impurity of tree.
 - 8.4 eliminate the Parkinson weight of element.
 - 8.5 The method is end until the values is zero.
- 9: Calculate Classifier performance for the model.
- 10: Display grapy for each classifier.
- 11: End While.

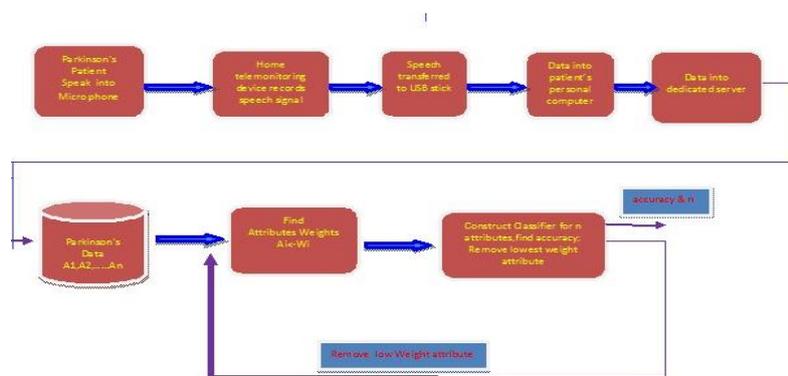


Figure6: Block diagram for **Mixed Weighting Mean-Classification** attribute subset selection

VII. EXPERIMENTAL RESULTS

Table2: Parkinson's disease Classifier measures

features	FN	IG	GR	CS	MWM-C
RAP	F6	397×10^{-3}	476×10^{-3}	172117×10^{-4}	60282×10^{-4}
spread1	F19	379×10^{-3}	381×10^{-3}	130932×10^{-4}	46177×10^{-4}
APQ	F13	379×10^{-3}	381×10^{-3}	130932×10^{-4}	46177×10^{-4}
PPE	F22	379×10^{-3}	381×10^{-3}	130932×10^{-4}	46177×10^{-4}
spread2	F20	298×10^{-3}	362×10^{-3}	136284×10^{-4}	476×10^{-2}
APQ5	F12	277×10^{-3}	279×10^{-3}	88798×10^{-4}	31452×10^{-4}
Shimmer	F9	277×10^{-3}	279×10^{-3}	88798×10^{-4}	31452×10^{-4}
APQ3	F11	277×10^{-3}	279×10^{-3}	88798×10^{-4}	31452×10^{-4}
shimmer(dB)	F10	277×10^{-3}	279×10^{-3}	88798×10^{-4}	31452×10^{-4}
DDA	F14	277×10^{-3}	279×10^{-3}	88798×10^{-4}	31452×10^{-4}

The Mixed weight Mean-Classification technique discuss below.

- 1: Compute Parkinson's each attributing any 3 classifier methods.
- 2: allocate weights are all feature.
- 3: construct attribute based on the weight coordinate in arrangement order.

4: compute classifier presentation.

5: remove the smallest feature weight.

In the above Parkinson's Voice data set is taken to calculate each Feature attribute numbers taken IG, GR, and CS. In this data set we apply MWM-C technique to compute each Feature attribute mean value. The attributes are arrange sorting order and remove the value zero. The attributes feature no are 1,2,3,4,5,7,8,15,16,17,18 and 21.

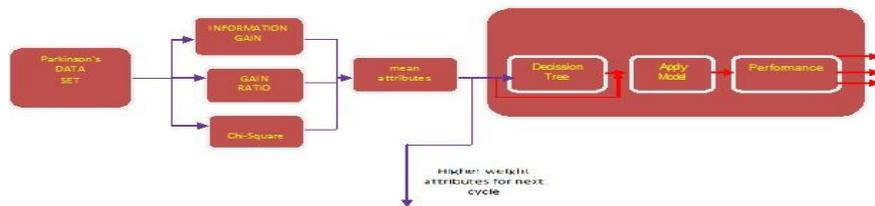


Figure7: MWM-C attributes subset selection

Classifier accuracy

. In the accuracy measure to defined as for the Parkinson's entitlement of accurate categorization of data[6].

In the define for the Parkinson's are

U is the acceptably categorization of who is well.

V is the acceptably categorization of who is patient among Parkinson's.

W is the inaccurately categorization of patient among Parkinson's.

X is the inaccurately categorization of who is well.

According above parameters the various classifier results are given below.

$$\text{Accuracy} = (U + X) / (U + V + W + X) \quad (28)$$

Accuracy of using ANN method for results is 0.740925, Accuracy of using ID3 method for results is 0.846153.

Accuracy of using J4.8 method for results is 0.89743 , Accuracy of using SVM method for results is 0.86153

Specificity:

In the specificity can be defined as the truth of identify of further long-suffering with Parkinson's. [3]

$$\text{Specificity} = U / (U + V) \quad (29)$$

Specificity of using ANN method for results are 0.8297, Specificity of using ID3 method for results are 0.8571

Specificity of using J4.8 method for results are 0.904, Specificity of using SVM method for results are 0.9251

Sensitivity

In the Sensitivity can be defined as the truth of detecting of wellbeing [3]

$$\text{Sensitivity} = X / (U + X) \quad (30)$$

Sensitivity of using ANN method for results are 0.2213, **Sensitivity** of using ID3 method for results are 0.2363

Sensitivity of using J4.8 method for results are 0.24, **Sensitivity** of using SVM method for results are 0.1904

Error-Rate: In the error- rate measure to defined as for the Parkinson's entitlement of inaccurate categorization of information. [7]

$$\text{Error-rate} = 1 - \text{Accuracy} (M) \quad (31)$$

Error-rate of using ANN method for results is 0.259075. Error-rate of using ID3 method for results is 0.153847

Error-rate of using J4.8 method for results is 0.10257, Error-rate of using SVM method for results is 0.13847

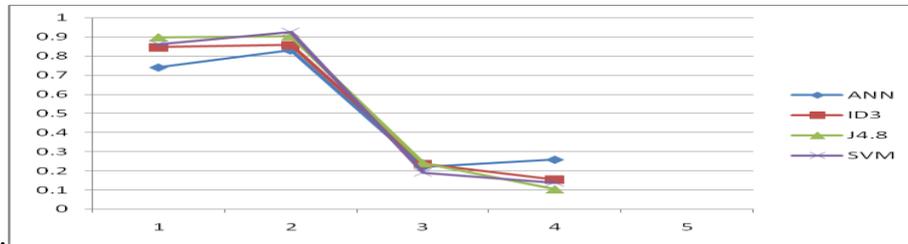


Figure 8: Performance measures various classifiers.

Table 3: Performance Measure of Gain

Performance Measure	Gain Value for Different Weights Threshold					
	100	90	80	70	60	50
Accuracy	87.69	88.71	89.23	88.20	88.20	88.20
Precision	0.93	0.94	0.94	0.89	0.89	0.89
Recall	0.89	0.90	0.91	0.88	0.88	0.88
Fmeasure	0.91	0.82	0.92	0.88	0.88	0.88

Table 5: Performance Measure of Gain Ratio

Performance Measure	Gain Ratio Value for Different Weights Threshold					
	100	90	80	70	60	50
Accuracy	90.76	89.23	89.23	88.20	88.20	88.20
Precision	0.91	0.90	0.90	0.89	0.89	0.89
Recall	0.90	0.89	0.89	0.88	0.88	0.88
F measure	0.91	0.89	0.89	0.88	0.88	0.88

VIII.CONCLUSION

The proposed Mixed Weight Mean-C methodology is estimated on the Parkinson's data after completion of preprocessing. The method produces the best results. In this paper, we elaborated the features of each attribute. We applied this MWM-C for finding best attribute to split by using the state-of-the-art methods. In future, this work can be extended by using IoT (Internet of Things) for collecting the data.

IX. REFERENCES

- [1] Hongbo Du, Data Mining Techniques And Applications An Introduction, Cengage Learning India PVT,(2010).
- [2] <https://archive.ics.uci.edu/ml/datasets/Parkinsons>
- [3] Ian.H.Witten, Eibe Frank, Mark A.Hall., Data Mining Practical Machine Learning Tools and Techniques, 3rd edition, MK,(2012).
- [4] Jiawei Han , Micheline Kamaber., Data Mining Concepts and Techniques, 3rd edition, MK, (2006).
- [5] K.P.Soman, Shyam Diwakar, V.Ajay., Insight into Data Mining Theory and Practice, 6th edition, PHI,(2012).
- [6] Pang-NingTan , Vipin Kumar, Michael Steinbach., Introduction to Data Mining, Pearson,(2013).
- [7] Ronald J.Baken, Robert F.OrliKoff, Clinical measurements of Speech and Voice, 2nd edition,(2000).
- [8] Sumeet Dua, Pradeep Chowriappa., Data mining for Bioinformatics., CRC Press,(2013).