

---

# CHAPTER 1

## INTRODUCTION

### 1.1 BACKGROUND

Parkinson's disease is a neurological disorder that appears as both non-motor and motor symptoms. This neurological disorder is chronic and progressive, mostly affecting the elderly, though it can also affect patients considerably younger. It is ranked number two among neurodegenerative diseases. The exact cause of Parkinson's disease is unknown; however, it is believed to be a result of a combination of hereditary and environmental factors. Estimates of incidence and prevalence vary significantly, positively in part because various investigations employ different methodologies, yet they tend to be higher in males than in women. Around, 1-2% of individuals aged 65 and older experience Parkinson's disease (H Ullah et al. 2018, H Khan et al. 2018). This chapter covers the main deep learning techniques for Parkinson's disease classification using vocal features. To enhance classification performance, dimensionality reduction, feature selection, and classification techniques are also discussed. Finally, this chapter addresses the aims of the investigation, the primary results of the research, and the overall structure of the thesis.

### 1.2 DEEP LEARNING

It is a type of Machine Learning inspired by the structure of a human brain. In terms of Deep Learning this structure is called Artificial Neural Network. Deep Learning maps the provided input to certain labels using a vast amount of data instead of requiring human-designed rules to function. Artificial neural networks, or ANNs, which are constructed in layers and provide several interpretations of the data they are fed, are the algorithms used in the construction of deep learning (Shrestha A et.al 2019).

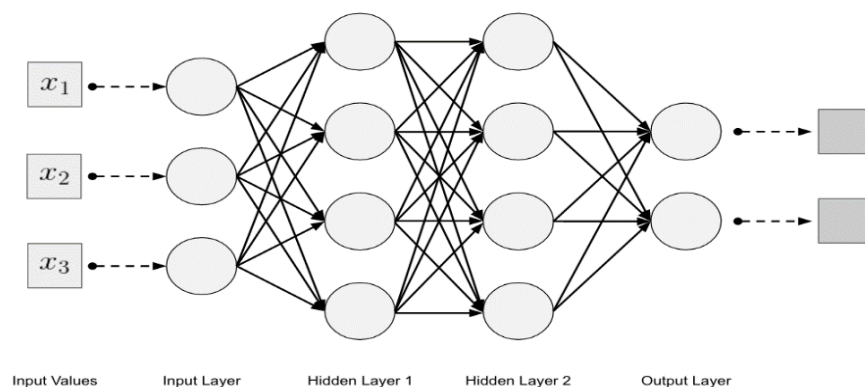
Deep learning is becoming more and more popular in many pattern recognition applications due to its low generalization error and strong learning capacity. The two primary deep learning methods for identifying Parkinson's disease from speech are end-to-end and feature engineering-based methods. Many deep learning models, such as multi-layer perceptron's (MLPs), convolutional neural networks (CNNs), and long short-term

memory (LSTM), have been investigated for the diagnosis of Parkinson's disease using hand-engineered speech features in recent publications. Using raw speech signals or voice time-frequency spectrograms as the source of feature representations, deep learning models are trained end-to-end, bypassing the need for hand-crafted features.

### 1.2.1 Deep Learning Architectures

Deep Learning architectures like CNN (convolutional neural networks), RNN (recurrent neural networks), and FNN (feed forward neural networks) are widely used in various fields.

Artificial neural networks, specifically CNNs, are most commonly utilized in deep-learning for the analysis of visual data. In at least one of its layers, it substitutes a mathematical operation known as convolution for general matrix multiplication. The multilayer stages in the perceptron's feature selection processes essentially consist of fully interconnected networks as shown in figure 1.1. In these networks, each neuron in the single layer is connected to all neuron in layers located below it. Because these networks are so intertwined, overfitting of the data might occur. CNN approaches regularization differently. Using smaller, simpler patterns incorporated within their filters, they build patterns of increasing complexity by utilizing the hierarchical pattern in the data.



**Figure 1.1. Deep Learning Architecture**

The relevant information is extracted by applying these filters to various sections of the input. The network is able to learn increasingly abstract representations of the input as it

moves through the layers because these features are built and blended into more complicated patterns. With a reduced chance of overfitting, this hierarchical method enables CNN to discover intricate patterns in data quickly. Convolutional networks resemble the architecture of an animal's visual cortex because of its neural connection structure, which is based on biological processes. The receptive field, a small area of the visual field, is the sole area in which individual cortical neurons react to inputs. The whole visual area is covered by partial overlaps between the receptive zones of various neurons.

A Recurrent Neural Network, or RNN, is capable of processing sequential data, including time series and plain language. A directed or undirected graph is created along a temporal sequence by the connections between nodes in an RNN type of Artificial Neural Network (ANN). It can therefore display temporally dynamic behavior. RNNs process the input sequences of different lengths by utilizing their internal state, or memory. They are based on feed forward neural networks. Recurrent Neural Network refers to an infinite class of networks with infinite impulse response, while Convolutional Neural Network refers to a finite class of networks with finite impulse response.

A time-varying behaviour is identified in both network groups. Directed acyclic graphs come in two varieties: finite impulses recurrent systems, which can be unrolled and replaced with completely feed-forward neural networks, and infinite impulsive recurrent networks, which create directed cyclic graphs that are unrollable. Finite and infinite impulses recurrent networks have the capacity to store more states, and the neural network has the direct ability to control the storage. If the storage system has feedback loops or time delays, it can be swapped out with another network or graph.

An ANN that operates on a linear information flow is called a Feed Forward Neural network (FNN). Essentially, FNNs represent the initial and simplest form of artificial neural networks. In these networks, data follows a unidirectional path, moving from input nodes through any potential hidden nodes to the output nodes. The basic structure of a feed-forward neural network is linear, consisting of a single layer of output nodes. Within this linear network, inputs are directly sent to the outputs through a set of weights. Each node calculates the sum of the products between inputs and weights. The network minimizes the

mean squared errors of the calculated outputs by adjusting the weights in relation to a specified target value.

### **1.2.1.1 Long Short-Term Memory (LSTM)**

RNN with LSTM (Bhati et al. 2019) have developed into a useful and flexible model for various learning problems associated with sequential data. The important objective of the classification technique involves exactly predicting the target class for every case in the dataset. The process comprises of training algorithm that makes use of the pre-determined examples for finding a solution to the parameter set required for suitable differentiation. To achieve the classification task, training examples are utilized for specifying a model, which is capable of classifying the data samples into named classes. On the other hand, LSTM are popular and useful for acquiring long-term temporal dependencies. They are unaffected by the optimization setback that the Simple Recurrent Network (SRN) are subjected to, and they have helped to raise the benchmark for a number of difficult problems (Rizvi et al. 2020, Hssayeni et al. 2019).

### **1.2.1.2 Bi-Directional Long Short-Term Memory (Bi-LSTM)**

Bidirectional long-short-term memory, or Bi-LSTM, is the training process for any neural network to store and process information in two directions: forward, from previous times to the future times, and backward, from the potential future to past times. Unlike a conventional LSTM, a Bi-LSTM receives its input in two directions. The Bi-LSTM is identical to its unidirectional equivalent. The network is linked to both the past and the future, which makes a difference. Additionally, on the backward pass, the next letter in the sequence would be supplied to Bi-LSTM, providing it with access to future data.

### **1.2.1.3 Fuzzy Convolution LSTM Classifier**

CNN is applied to the transformed vocal features to extract relevant patterns and characteristics that may indicate the presence of Parkinson's disease. This step is crucial for learning spatial hierarchies of features. The output from the CNN, which represents high-level features of the vocal signals, is then fed into an LSTM network. The LSTM can model the temporal dynamics of these features, capturing the sequence and evolution of vocal characteristics over time. Fuzzy logic can be integrated at various stages, such as in the

decision layers of the network, to handle uncertainty in the data and predictions. For example, fuzzy rules can be applied to the outputs of the LSTM to classify the severity of Parkinson's disease based on the vocal features with a degree of uncertainty. The final output of the system is a classification decision, indicating whether the vocal features suggest the presence of Parkinson's disease and potentially its severity.

#### **1.2.1.4 Fuzzy Convolution Bi- Directional LSTM Classifier**

The Fuzzy Convolution Bi-LSTM strategy is designed to leverage the synergies among fuzzy logic, convolutional operations, and bidirectional LSTM to capture intricate patterns within vocal datasets associated with Parkinson's disease. This approach provides a holistic framework that takes into account both spatial and temporal characteristics of the data. Ensuring the model's efficacy in real-world applications necessitates meticulous fine-tuning and validation of its performance on independent datasets.

#### **1.2.1.5 Ensemble Deep Learning Classifiers**

The classification approach considers the training dataset in the form of input, and information about the class identifiers is already available. Classification helps in the study of the training dataset and establishes a model associated with the class identifier, and designates the class identifiers to the future unknown records.

#### **Sparse Auto-encoder:**

Utilizing a Sparse Autoencoder in a deep learning framework facilitates efficient dimensionality reduction and extraction of crucial features for Parkinson's disease classification in vocal datasets. The incorporation of sparsity constraints not only improves the interpretability of the learned features but also contributes to the overall effectiveness of the classifier.

#### **Contractive Auto-encoder:**

The design of a Contractive Autoencoder involves acquiring a concise and resilient representation of the input through the inclusion of a penalty term in the loss function, ensuring stability in the learned features. Employing a Contractive Autoencoder in this fashion not only promotes efficient dimensionality reduction but also fosters feature

stability. This approach has the potential to improve interpretability and enhance the overall performance of the model when applied to Parkinson's disease classification using vocal datasets.

**Stack Generalization:**

Stacked Generalization, often referred to as stacking, is an ensemble learning technique where multiple models are trained and their predictions are combined to improve overall performance. Stacking allows for the combination of diverse models, potentially capturing a broader range of patterns in vocal data associated with Parkinson's disease. It helps mitigate overfitting and enhances the robustness of the overall classification system.

### 1.3 PARKINSON'S DISEASE DIAGNOSIS

Parkinson's disease is categorized as neurological diseases where phonation of patients was impacted by a variety of neurological conditions, and their voice can be a significant neurological parameter in diagnosis. In PD, nearly 45% of patients are affected by voice ailments. PD is the prominent reason behind parkinsonism, a set of identical symptoms (Tysnes et. al. 2017). PD is also known as "primary parkinsonism" or "idiopathic PD" ("idiopathic" implies no known cause). There are very few instances where the symptoms are brought on by toxicity, medicines, genetic mutation, head trauma, or other health issues when numerous aspects of Parkinsonism show idiopathic behavior.

#### 1.3.1 Symptoms and Diagnosis of Parkinson's Disease

PD impacts mobility (motor symptoms). The other common symptoms are mood disturbances, behavior, thought, and sensation (non-motor symptoms). The symptoms of each patient may be nearly different and the advancement also differs person to person.

#### 1. Motor Symptoms

A. Cardinal/primary symptoms include:

**Tremors:** Despite the fact that 30% of patients are expected to have very little shivering that can be felt; these people are classified as kinetic-rigid, this is the most evident and well-known symptom. It is often referred to as "rolling pills." a typical 4–7Hz tremor that is lessened with intentional movement and is greatest when the limb is at rest. Usually, it begins unilaterally.

**Stiffness:** Many people with Parkinson's disease exhibit rigidity, or a reluctance to move. If there is a passive movement of the limb, rigidity, also known as "cogwheel rigidity," is produced by stiffness, an increase in muscular tone, and a resting tremor.

**Bradykinesia/akinesia:** Bradykinesia, or the decelerating and absence of instant and automated movement, sluggishness or loss of mobility. Quick, redundant movements lead to dysrhythms and reduced amplitudes. In addition, "dysdiadokinesia", involves losses in capabilities of carrying out quick switching movements.

**Postural instability:** impaired postural reflexes that cause balance problems and slips.

Dopamine levels in SN (substantianigra) neurons are depleted, which causes an increase in above mentioned symptoms including tremors, stiffness, and slow movements. A signal in humans is transmitted down neurons from SN areas of the brain and received in the striatum, where these signals travel via intercellular spaces as chemical messenger dopamine (Srinivasan et al. 2017). In Patients of PD, dopamine generated by the cells in the SN is quite less, therefore this pathway is disrupted, resulting in the manifestation of symptoms stated above (Fig. 1. 2).

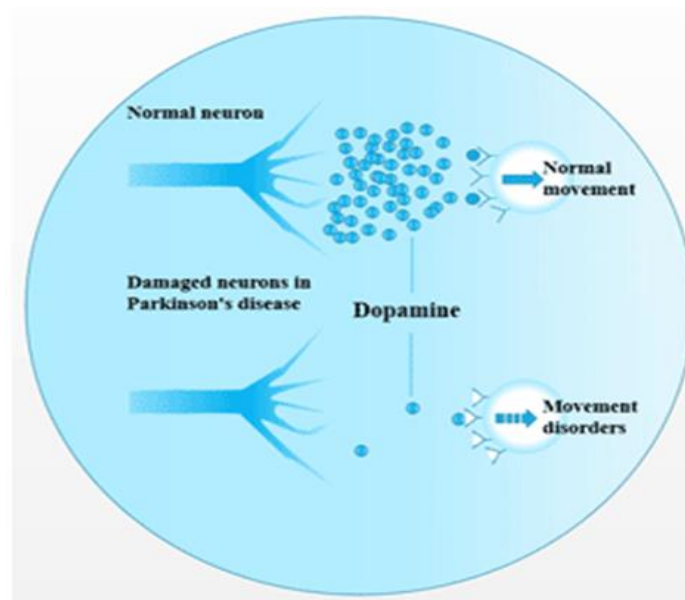


Figure 1.2. Normal and Parkinson's Disease Neurons

In healthy individuals, the striatum gets a signal that is delivered by neurons from the SN area of the brain, where the signals go across the gap between the neurons in the form of the chemical messenger dopamine. Reduced productions of dopamine impact these pathways in PD patients.

**Secondary symptoms:**

(i) Disturbances of Gait and postures

**Shuffling:** A shuffling gait is characterized by short strides that are taken with little time spent off the ground. The patients stumble with each light obstacle scan.

**Reduced swing of the arms:** It is a kind of bradykinesia; Turning "en bloc": PD patients have stiff necks and require multiple slow steps in turning of their necks and trunks. Their movements are bent, with flexed forward positions around the toes. In severe PD instances (camptocormia), patients show slumped heads or upper parts.

**Festinations:** They are combinations of curved shoulders with quick unbalanced movements. They cause strides that become increasingly hurried and frequently result in patient's fall.

**Gait freezing:** Akinesia or inability of movements is "frozen." Freezing in Walks are characterized by difficulties in moving the feet, especially in confined spaces or at the start of gait.

**Dystonia:** Nearly 20% of people have dystonia, which is characterized by irregular, difficult-to-control twisting muscle contractions that typically affect the foot and ankle and cause gait disruption. However, dystonia, which affects a large number of skeletal muscles, may be reasonably explained; these occurrences are really bad and completely paralyzing.

(ii) Speech and swallowing disturbances

**Hypophonia:** It signifies quiet voice where speech is delicate, grating, and repetitious. Few PD patients describe their tongues as "heavy." Festinating speech is very brief, hushed, and difficult to understand.

**Drooling:** Drooling is caused in bent postures, fragile movements and uncommon swallows. Non-motoring result in cognitive disturbances related to emotional perspectives of speech and distorted languages in both expressions and receptions.

**Dysphagia:** Afflicted swallowing incapability can result in aspiration, pneumonia, and finally patient's death.

Analyzing and categorizing speech signals from patients are considered a potential method for early Parkinson's disease (PD) diagnosis, aiming to identify unique voice characteristics before conventional symptoms manifest (Al-Fatlaw et al. 2016). The purpose of this study was to determine whether speech problems might be recognized in the early stages of Parkinson's disease (PD) before typical symptoms appeared, and whether acoustic and classification analysis could be used to identify individuals who were at risk among the general population. The study found limitations in speech articulator motions (lips, tongue, and jaw), characterized by reduced range (hypokinetic), and centralization of vowels. This manifested as high-frequency formants likely having lower frequencies, and lower-frequency consonants having higher frequencies. Speech monitoring and repair assessments effectively differentiated between groups and correlated with linguistic performance tests. Hypokinetic dysarthria, a characteristic Parkinson's disease speech disorder, adversely affects comprehension, self-image, and productivity. Notably, the study addressed previous shortcomings by controlling task-related and specific cognitive factors, avoiding flaws from prior research. Common deficiencies observed included monolith, monologs, and inappropriate pauses in the prosody domain, attributed to disruptions in vocal-fold vibration frequency and language latencies. Restricted characteristics, like specific vowel pronunciations against consonants, were frequently employed to manifest speech regions (García, A.M et al 2021).

The use of speech signals holds promise for diagnosing diseases, particularly in individuals experiencing various speaking difficulties. Common speech impairments observed in affected individuals encompass reduced speech intensity, frequency component variations, hoarseness in the voice, and inconsistencies in speech articulation, such as hypokinetic dysarthria. Extracting and classifying speech features pose challenges due to the

non-stationary and discontinuous nature of speech signals. A significant hurdle lies in appropriately interpreting voice and speech data to identify Parkinson's disease (PD). The depletion of neurotransmitters, notably dopamine, gives rise to a range of symptoms affecting speech, vision, mobility, urinary functions, weight, mood, and sleep patterns. Affected individuals commonly experience vocal and speech issues like dysphonia, hypophonia, monotonic speech, and dysarthria. Dysphonia, prevalent in over 90% of patients, and gait unpredictability stand out as distinguishing criteria for the condition's progression. Consequently, there has been a growing interest in advancing speech signal processing technology to better detect and understand Parkinson's disease symptoms.

The dataset employed in this study is accessible on the UCI Machine Learning Repository. The study directed by Sakar et al. (2019) emphasized the potential introduction of bias in performance evaluations arises from the inclusion of multiple voice recordings per individual in existing datasets and the utilization of the same individual's recordings for both training and testing in Cross-Validation (CV). In our dataset, encompassing multiple voice recordings for both healthy individuals and those with Parkinson's disease (PD), we have opted for the Leave-One-Person-Out Cross-Validation (LOPO CV) approach to assess the performance of the proposed frameworks. In each iteration of LOPO CV, we exclude instances linked to one individual as a test set, utilizing the remaining instances for training purposes.

#### **1.4 FEATURE EXTRACTION**

Feature extraction through vocal features is crucial in deep learning, especially for applications involving speech processing, voice recognition, and audio analysis. The extraction of pertinent features from raw audio signals allows models to encapsulate fundamental characteristics, catering to diverse applications. The features including Vocal fold, TQWT, WT, MFCC, Time frequency and baseline features were extracted from dataset.

**Wavelet transforms (WT)**, including Tunable Q Factor Wavelet Transforms (TQWT), serve as feature extraction methods in deep learning applications for vocal

datasets. These methods entail the examination of audio signals across diverse scales and resolutions, capturing information in both the time and frequency domains. The application of wavelet transforms enables the extraction of features that represent various aspects of vocal characteristics.

**Tunable Q Factor in wavelet transforms (TQWT)** provides the ability to adjust the width of frequency bands, offering flexibility to accommodate different patterns within vocal data. These techniques support the creation of deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), which can be trained to identify and interpret intricate patterns within vocal datasets for tasks such as speech analysis or voice recognition Top of Form

**Mel-Frequency Cepstral Coefficients (MFCCs)** find extensive application in speech and audio processing, portraying the spectral attributes of audio signals through the capture of power spectrum across varying frequencies. RNNs or CNNs, especially in the context of speech recognition, can be trained using Mel-frequency cepstral coefficients (MFCCs)

In a recent investigation by Sakar et al. (2019), the tunable Q-factor wavelet transform (TQWT) was employed on vocal signals from individuals to diagnose Parkinson's disease (PD). The effectiveness of the extracted TQWT features was compared with commonly utilized vocal features in PD research. The study involved experiments with multiple voice instances from 252 individuals, extracting various feature sets from these instances. These feature subsets were provided to various classifiers, and the outputs of these classifiers were combined using a majority voting scheme. The study's findings indicated that TQWT features demonstrated comparable or superior performance compared to the widely used voice features in PD classification. Furthermore, it was observed that combining MFCC and TQWT features enhanced classification performance, particularly when mRmR selection was applied.

## **1.5 DIMENSIONALITY REDUCTION**

Medical knowledge is crucial for patient diagnosis and treatment in clinical research (Mostafa et al. 2018, Hassan et al. 2019). In the case of medical research, it also produces

valuable data to support medical researchers and enable therapeutic augmentation. Medical knowledge is crucial for patient diagnosis and treatment in clinical research. It is possible to see a pattern in the way that clinical research, policy, performance measurements, quality indicators, and research are distributed in the field of medical information. The dataset includes of higher dimensionality, and irrelevant number of features.

Dimensionality reduction (Ali et al. 2019) indicates the procedure of decreasing the number of random variables or features considered. Reduction of high-dimensionality data, in the form of a sub-phase in the data pre-processing-step, holds huge significance in several practical applications. High-dimensionality reduction has been found to be one among the important tasks in applications. For instance, a dataset may exist with many hundreds of features. Reducing the dimensionality involve removing those features of data by combining them in a such a way there will be huge loss in the important features of the actual dataset (Segovia et al. 2018, Peng et al. 2019). The representation of the features that can be achieved in the medical field including patient test report, patient history, side effects, are made up of several aspects, ranking the status of a patient health (Reddy et al. 2020). Few advantages of dimensionality reduction approaches used on a dataset are listed as below.

1. With the reduction in the number of dimensions, there is also a consequent reduction in the memory required too.
2. Its time taken for computation is low.
3. Recurrent, unnecessary, and noisy data can be eliminated.
4. Data quality can be enhanced.
5. Few algorithms do not exhibit better performance when number of dimensions used is more. Therefore, dimensionality reduction aids an algorithm in efficient functioning and the accuracy is improved.
6. Visualizing data in higher dimensions is difficult. Therefore, reduction in the dimension may help designing and exploring the patterns with better clarity.
7. It makes the process of classification simple and also increases the effectiveness.

Curse of Dimensionality is a term used to describe a significant problem with high dimensional data. Few dimensionality reduction techniques like Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Kernel Based Principal Component Analysis (KPCA).

**Linear Discriminant Analysis (LDA):** LDA makes use of the information from different features to generate a new axis and then the data is projected on to the new axis in such a manner that the variance is minimized and the variations between the classes means is maximized (Lahmiri et al. 2017). LDA is essentially a supervised method that only works with labeled data. It comprises data statistics that have been calculated for each class. This represents the mean and variation of the variable for each class for one input variable ( $x$ ). In the case of multiple variables, they imply same variables like means and covariance matrices which are computed using multivariate Gaussians. LDA is a supervised approach, therefore labeled data is required.

**Principal Component Analysis (PCA):** One technique for lowering dimensionality is PCA, that identifies the important links in the data, adjusts the available data in light of these correlations, and then measures the importance of these associations in order to keep the highly significant correlations (Cigdem et al. 2018, Rasheed et al. 2020). It can be split into four stages:

1. To find the correlation among the features using a Covariance Matrix.
2. To derive eigenvectors and eigenvalues using the Covariance Matrix's linear transformation or eigen decomposition.
3. Next, the data is transformed applying eigenvectors into principal components.
4. Finally, the significance of these correlations is measured with Eigen values and the significant principal components are found.

**Kernel Based Principal Component Analysis (KPCA):** Kernel PCA introduces a non-linear mapping function, often referred to as a kernel function, to project the data into a higher-dimensional space where linear techniques can be applied. The basic idea is to implicitly compute the dot products in this higher-dimensional space without explicitly

calculating the coordinates of the data points in that space. This is particularly useful when dealing with data that is not linearly separable in the original feature space (Wang et al. 2016).

## 1.6 FEATURE SELECTION

Feature Selection refers to the process that finds and eliminates the maximum amount of unnecessary and repetitive information (Chia et al. 2020). Feature selection stands for the process that chooses a subset of the actual variables such that the superior performance is achieved by a model constructed on data having just these features (Shrivastava et al. 2017, Adeli et al. 2017). The goal is to extract a subset of characteristic from the initial problem that still serves as a good representation of it. Methods of FS can be viewed as combinations of search strategies that add new feature subsets and evaluation scores of feature subsets. Common approaches include testing potential subsets of characteristics and selecting ones with lowest error rates. Algorithms are heavily impacted in evaluations that distinguish FS algorithmic types including wrapper, filter, and embedding techniques (Haq et al. 2019). Figure 1.3 illustrates the different phases of FS process. Its performance relies on the decision made at each stage.

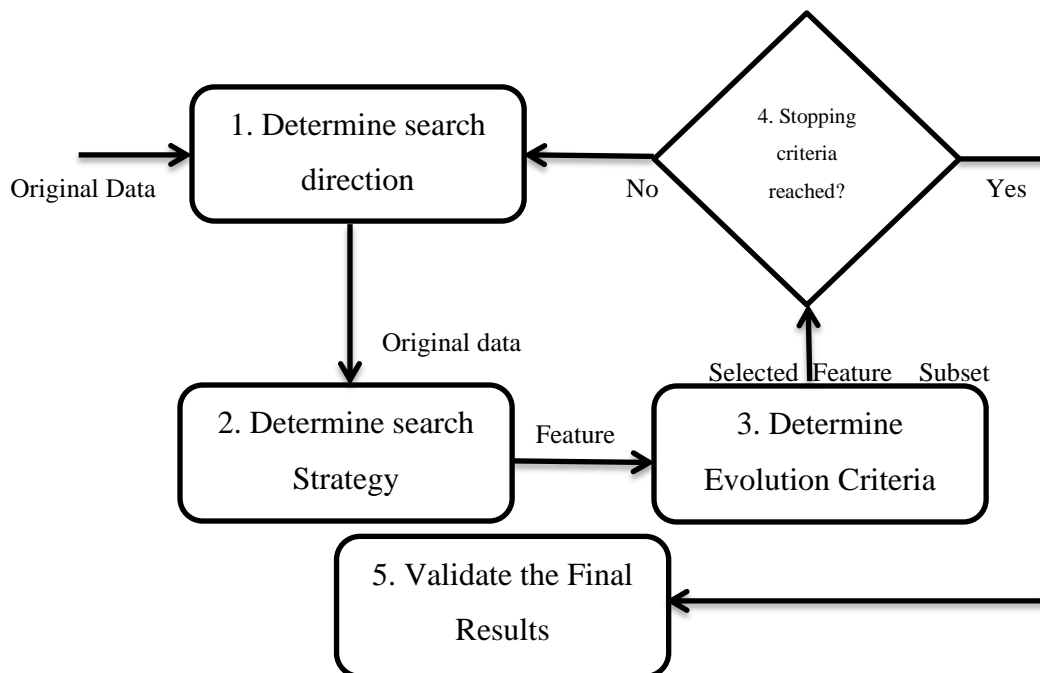


Figure 1.3. Stages in Feature Selection Processes

1. **Search Direction:** Choosing the search direction and the starting location are both part of the first step of FS. Three types of search directions—forward search, backward search, and random search—are fully classified. Forward searching is a method of conducting a search that can start with a null set and add additional features using recursion by each iteration (Ali et al. 2019). In contrast, the backward eliminations begin with complete feature collections and eliminate one by one for obtaining required subsets of characteristics. Another process involves a random search approach, which constructs the feature subset through the addition and elimination of the features through iteration. Once the search direction is determined, the search mechanism can be used.
2. **Decide Search Mechanism:** From the literature, search mechanism can be randomized, exponential and sequential search. Among these, the most commonly used one is exponential search and it faces  $2^N$  problems for N characteristics and different combinations of FS are required. It is a Non-Deterministic Polynomial (NP) hard issue that uses a strict search strategy. These difficulties have been overcome by the researchers' use of randomized search techniques. Sequential search entails deleting features from the entire set or adding them one at a time to a null set. Sequential Forward Selections (SFS) and Sequential Backward Selections (SBS) are the terms utilized to describe these processes, respectively. One drawback of these methods is that the traits that are eliminated won't be taken into account for subsequent iterations. This research has highlighted the use of Swarm Intelligence methodologies to overcome the shortcomings in all of these techniques.
3. **Evaluation Condition:** Based on the evaluation circumstance, the best characteristics are selected. FS can be filter, wrapper, and embedded according to evaluation techniques.
4. **Termination Condition:** Feature selection procedures must stop when the stopping condition is reached. An optimal feature subset may be obtained with less computing effort when the termination condition is good, and the over-fitting issue is also resolved. Choices made in previous stages impact termination condition decisions. Several common termination circumstances include:

- Pre-determined counts of features
- Pre-set counts of iterations
- Percentages (%) of progress over consecutive iterations
- Depending on evaluation functions.

5. **Validation of the Results:** Results are validated using feature set validation techniques, such as Rand Indices, cross-validations, confusion matrices, and Jaccard similarity-based metrics. Cross-Validations are highly popular evaluation approaches as they yield non-skewed error predictions. Confusion Matrices are produced for evaluating classifiers.

On the whole, Feature selection forms the cue for being capable of predicting values with any level of accuracy. FS is carried out by either considering the significant features or neglecting the unwanted features in the dataset with no alteration done to them (Chen et al. 2020). FS achieves several goals.

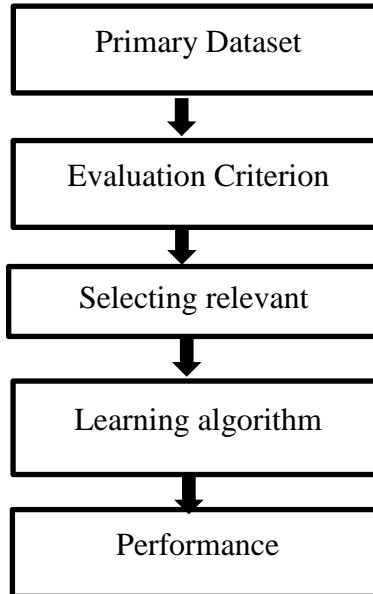
- It minimizes the computational time and complexity involved during the training and testing of classifiers, resulting cost optimality of models.
- It removes the unwanted and noisy features by maintaining the ones having the least redundancies and highest relevance's to target variables.
- It improves the performance of learning algorithms, guards against over fitting, and assists in generating superior generalized models.

### **Types of Feature Selection**

The three types of feature selection techniques are filter, wrapper, and embedding methods, which are categorized according to how they work with the classifier. (Chen et al. 2020).

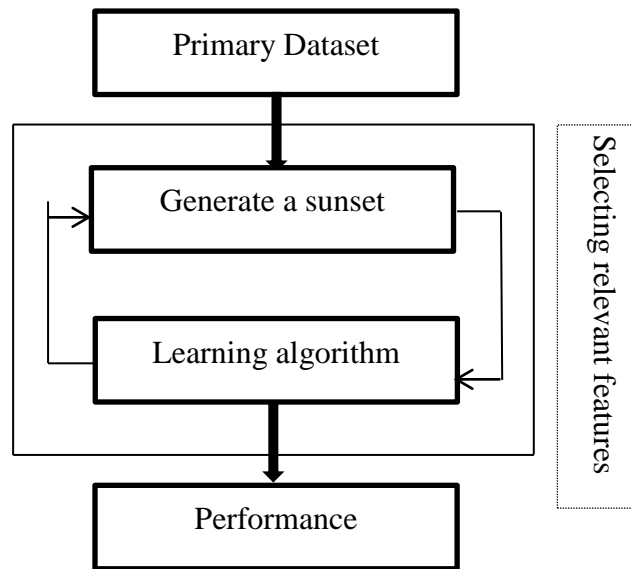
**Filter methods** possess scalability (reaching very high-dimensional data) and be able to do quick FS before classifying so that the bias of a learning algorithm does not conflict with the bias of the FS method. They primarily function as rankers, arranging the traits in descending order of importance. The inherent characteristics of the data, such as variance,

consistency, distance, information, and correlation, etc., are used to grade the features. Each filter technique uses its own criteria for determining the relevancy of the data, and new ones are continually being developed. A popular definition of relevance is that a feature is irrelevant if it has no conditional dependence on the class labels or if it has no influence on the class labels; in these cases, it can be removed. The filter approach applies several metrics via ranking to eliminate the undesired feature and recurrent columns from the model. Filter techniques have the advantage of having a shorter calculation time and avoiding overfitting the data. The selection process and the classification process are therefore made distinct when using the filter approach, as shown in Figure 1.4.



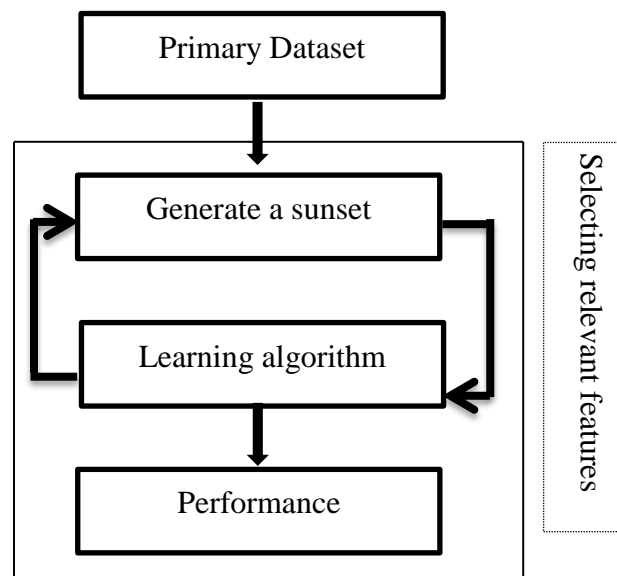
**Figure 1.4. Filter Based Method Process Flow**

**Wrapper methods** they rely on the classifier since they use MLT as a black box evaluator to obtain the best feature subsets. In actuality, any combination of the modelling method and search mechanism may be used as a wrapper. When wrappers are applied on datasets with more features, they are computationally expensive. The implementation of these techniques is easy and feature dependencies can be modeled. Therefore, classification is “wrapped” in the selection procedure, as shown in Figure 1.5.



**Figure 1.5. Wrapper Based Method Process Flow**

**Embedded methods** fill the area between filters and wrappers. They first combine measurable and statistical conditions, such as a filter, to choose a small number of characteristics, and then, with the aid of MLT, they choose the subset that performs the best in terms of classification. The computational complexity of wrappers is reduced with no re-classification of the subsets during every iteration and the feature dependencies can be modeled. Iterations are not required here. Figure 1.6 shows FS are primarily by-products of classification steps.



**Figure 1.6. Embedded Based FS Method Process Flow**

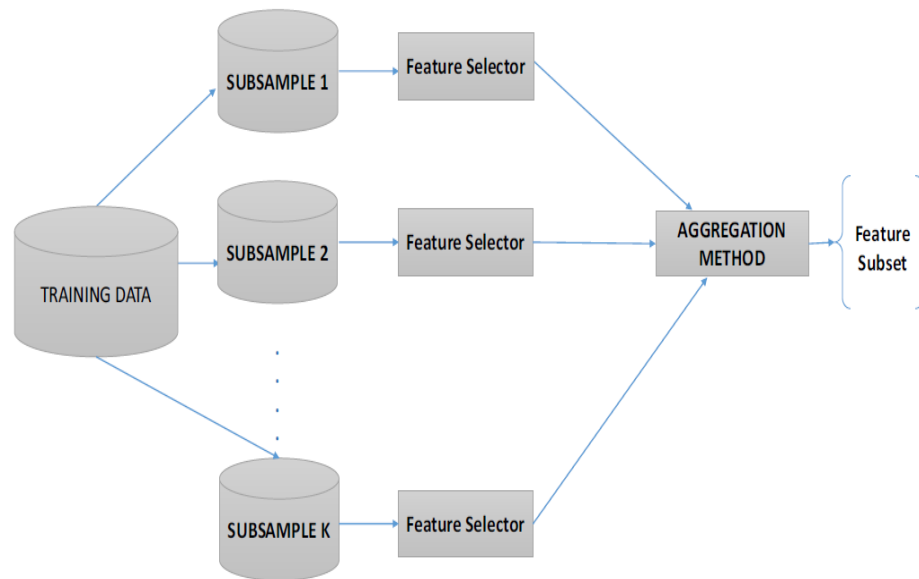
These constitute rapid processing approaches that are identical to the filter technique, however their accuracy is high compared to the filter method.

## **1.7 ENSEMBLE FEATURE SELECTION AND CATEGORIES**

Finding a collection of traits that will increase the classification models' predictive power and ease comprehension is the aim of feature selection strategies. Ensemble feature selection involves the concept that combines the outputs from different single feature selection framework that yields improved results when compared to individual mechanisms of feature selection. An ensemble feature selection is advantageous because it helps to mitigate and balance for these biases, contrary to past research that have demonstrated that single feature selection approaches might display specific biases. Particularly in high-dimensional and small sample size configurations, where it is quite naturally difficult to extract the consistent feature subsets, the ensemble concept has been examined in the form of a potential model to enhance the reliability of the selection procedure. However, this is not possible just with the presence of many models, like it is the scene with classification mixtures, however it is also due to the versatility of the feature subsets achieved (Kadam. et.al. 2019).

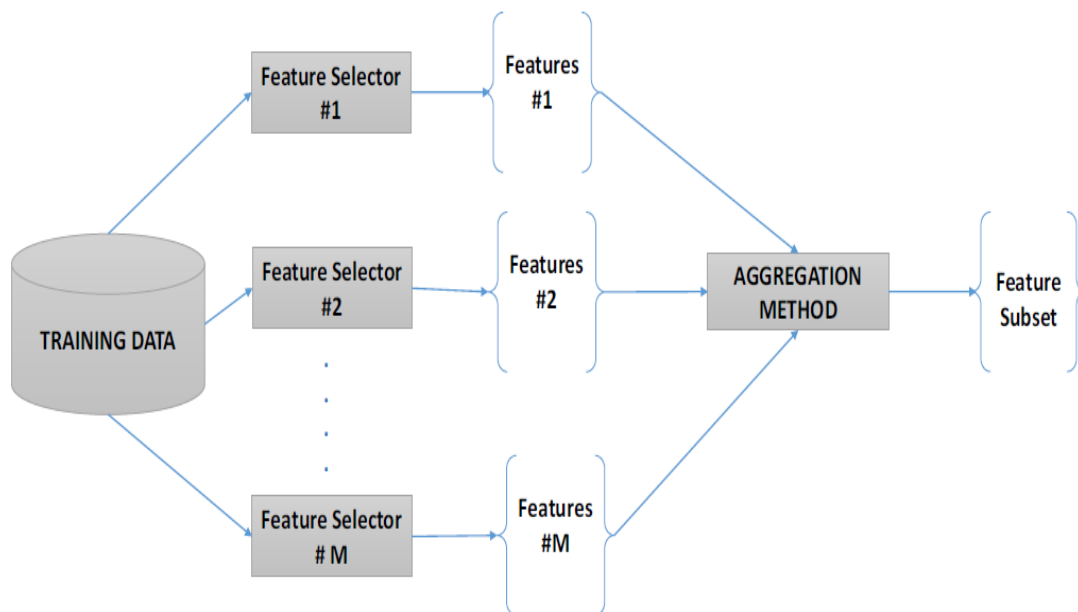
There are two types of ensembles created for feature selection namely homogenous which have the same underlying feature selection and heterogeneous which have multiple feature selectors (Guan et al. 2014, Seijo-Pardo et al. 2017). Both methods have been shown to provide useful templates. Attention must be paid to combination phases, which entail integrations of many outputs that feature selectors produce. They are crucial to have combination mechanisms that can process or sort subsets of features. Other factors are also studied in this work which includes performance evaluations of these combinations in terms of adaptability and constancy or manuals with software tools that include outcomes of feature selection.

The ensemble is regarded as homogeneous if the base selectors are members of the same type; otherwise, it is referred to as heterogeneous. The homogenous strategy uses a different selection of training data along with the same feature selection techniques as illustrated in Figure 1.7a.



**Figure 1.7a. Homogeneous Scheme**

This method makes the size of the divides into a design consideration. These methods are additionally known as data variation ensembles (Anisha, et.al. 2020). Few instances of homogeneous schemes, primarily with the objective of having the capability of managing with massive scale conditions, can be observed in the current work, and in the latest work, this latter with the additional objective of being capable of handling datasets.



**Figure 1.7b. Heterogeneous Scheme**

In the case of the heterogeneous mechanisms, several diverse FS techniques are used on the same training data like it can be observed in Figure 1.6b. In this method, as previously noted, the quantity of distinct feature selectors that will be used also counts as a design element. These methods are additionally referred to as function variation approaches. In contrast to homogeneous FS ensembles, heterogeneous FS ensembles are more common and can produce a large number of samples. Both homogeneous and heterogeneous mechanisms are used in only a handful of these analyses. Depending on the type of used feature selectors, feature rankings can be generated as outputs where in the latter case, extra threshold steps are needed. The final outcomes of base selectors need to be merged and hence aggregations are used.

## **1.8 SWARM INTELLIGENCE (SI)**

A swarm intelligence system generally includes a population consisting of considerably simple agents that have local interaction with themselves and their environment, with no global knowledge regarding their individual state and about the world position. Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) are presently the famous algorithms in the field of SI which demonstrate multifarious characteristic as systems with above stated features can be noticed in several fields. Research carried out in swarm intelligence can be categorized on the basis of various conditions (Sehgal et al. 2020).

**Natural vs. Artificial:** It is natural to classify the research of swarm intelligence into two categories based on the characteristics of the systems being analyzed. Here, natural swarm intelligence research is being considered, in which biological systems are examined; and in the case of artificial swarm intelligence, human artifacts are analyzed.

**Scientific vs. Engineering:** Swarm intelligence research can be divided into two streams: scientific and engineering. This is a less thorough classification based on the goals rather than just one alternative. Scientists research swarm intelligence systems and try to uncover and clarify the principles of how local relationships between individuals and their environment result in a cooperative system as a whole. Maximizing the application of knowledge gleaned from the scientific stream is the aim of the engineering branch, which

aims to enable the development of systems capable of addressing major practical difficulties.

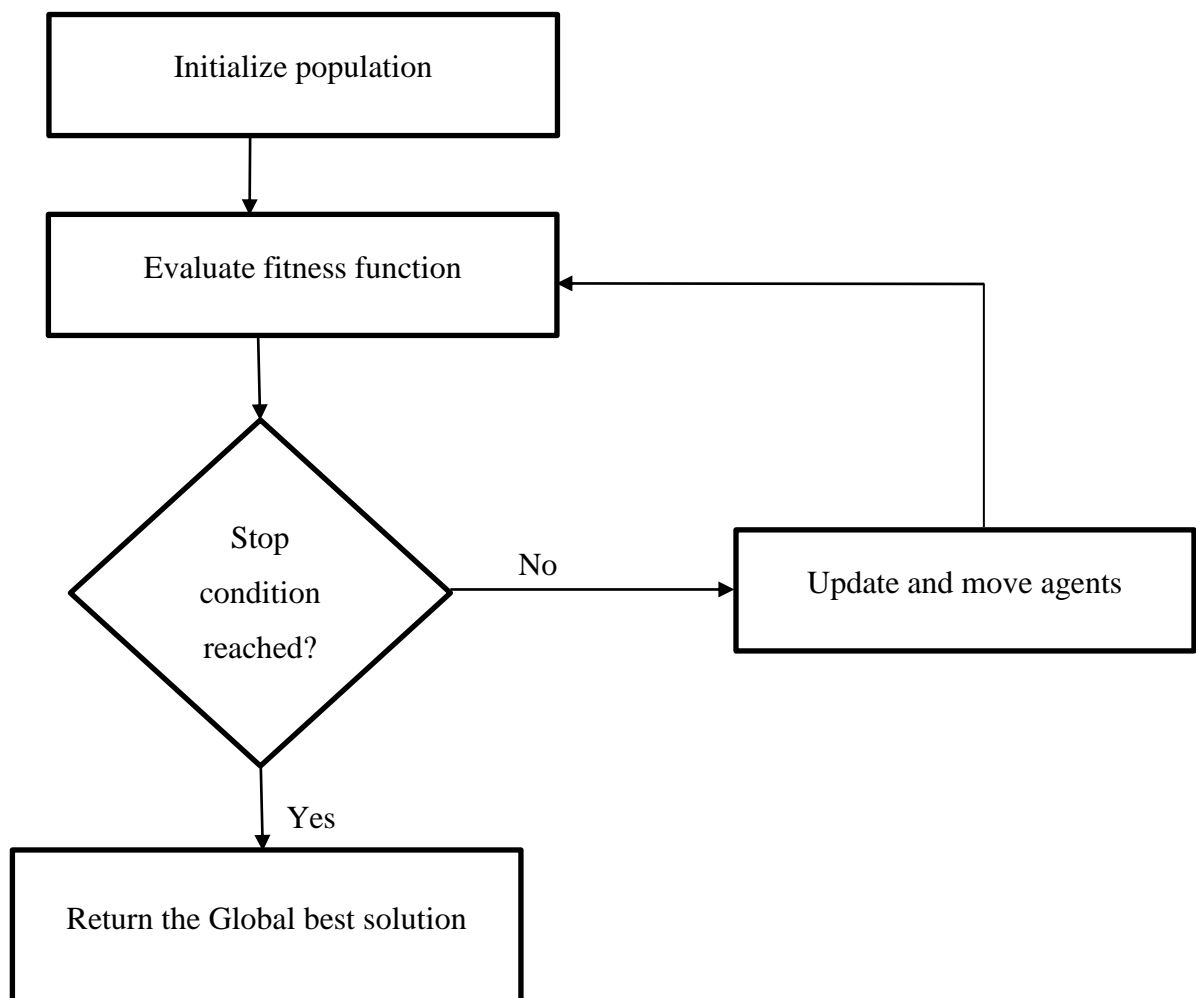
The unified behaviour present in localized, self-organized systems is the foundation of artificial intelligence types referred to as Swarm intelligence. These systems usually have a population of simple actors that interact with their environment and with one other on a local level. EA employ strategies that are based on biological evolutions, like reproductions, mutations, recombination, and selections. Possible results to the optimization issues should behave like people in their own populations, and fitness functions determine quality levels that these candidates must possess. Repetition of the evolutionary algorithm causes the original population to evolve and wanders toward global optimization (Gong et al. 2020). The plain populations of artificial agents that make up Swarm intelligence, on the other hand, are often those that interact locally with the environment. The role of each agent in this paradigm, which often draws inspiration from nature, is straightforward. However, through local interactions and largely random interactions, these agents come to exhibit "intelligent" global behaviour that is unknown to the individual individuals (Yong et al. 2016).

One of the best advantages of Swarm intelligence is autonomy. There is no external management for the Swarm, but every agent in the swarm regulates their behavior independently. Here, the agent indicates a probable solution for a particular problem. That makes it possible to foresee the second advantage, which is self-organization. Although it does not specifically target any one agent, the intelligence develops inside the swarm as a whole. As a result, the problem solutions (or agents) are unknown beforehand but change as the programme runs. An important aspect of flexibility is self-organization. The latter is apparent in variable environments where the response of the agents is quite better as per the modifications stated; they change their behavior, and are adaptive to them independently. Furthermore, because there is no centralized coordinating, the swarm is dependable, with no single point of failure. Furthermore, the swarm permits redundancy, which conceals two further benefits. The first is about scalability, which means that the swarm can have anywhere from a few hundred to thousands of agents, and the control architecture is the

same in both cases. In addition, due to the swarm holding importance to no individual agent, the Swarm intelligence benefit of flexibility is completely reasonable.

### 1.8.1 Fundamental phases of SI framework

Algorithmic Swarm intelligence hold to a few basic steps and their frameworks (Figure 1.8) can be summarized as below: 1. Initializations of populations, 2. Specifying termination criteria 3. Fitness functions for assessments, 4. Updating and shifting agents, and 5. Retrieving global best solutions.



**Figure 1.8. Flow of Swarm Intelligence Algorithm**

**Initialize Population:** Initializing the population forms the first step in any Swarm intelligence approach.

**Define stop condition:** The important function of the termination criterion is to halt the algorithm execution. A parameter supplied for ending the algorithm execution is called a stopping condition or termination criteria. The termination criterion in the vast majority of Swarm intelligence algorithms is either one or two separate criteria. The number of iterations is the most common termination criteria.

**Evaluate fitness function:** The search agents are evaluated in the third stage of the Swarm intelligence model, which evaluates the fitness function. The fitness function might be solitary, such as any fundamental parameter like classification accuracy, or an ensemble, much like the termination criterion. Accurate categorization is a common way to define fitness function.

**Move and update the agents:** According to the algorithm mathematical derivation, agents in a SI algorithm are updated and moved.

**Return the global best solution:** The ideal search agent is formed by the outcome of a SI algorithm.

### **1.8.2 Wrapper models based on Swarm intelligence**

In general, Swarm intelligence are preferred optimizers for feature subset selection procedures within wrapper model approaches as they evaluate the quality of the features using specified Machine Learning Technique (MLT) (i.e., classifiers), and their Feature Selection stop algorithm biases. However, because Machine Learning Technique have a high computational cost, they must be performed numerous times to check the quality of the selected feature subsets. As a consequence, Swarm intelligence methods may be utilized to optimize the Feature Selection process within a wrapper model. Their objective is to choose a subset of all characteristics. That, when combined with preset Machine Learning Technique, produces satisfactory predictive performance. A popular wrapper approach, given a specified classification algorithm, iterates through the two stages below until the required objective is reached (i.e., a specific accuracy is achieved):

- Pick a portion of the characteristics.
- Apply the selected classification algorithm on the selected subset to evaluate it.

### **1.8.2.1 Butterfly Optimization Algorithm (BOA)**

A population-based meta-heuristics programme called Butterfly Optimization Algorithm mimics butterflies' foraging patterns. Butterfly Optimization Algorithm is the most recent meta-heuristics algorithm that mimics the foraging and mating behaviour of butterflies (Dubey 2021).

The Butterfly Optimization Algorithm (BOA) is a heuristic optimization method inspired by the collective behavior of butterflies. It is employed to seek optimal solutions for various optimization challenges. In the context of classifying Parkinson's disease using vocal datasets, the algorithm is utilized to fine-tune the parameters of a classification model. The primary goal is to enhance the model's performance in effectively differentiating between individuals afflicted with Parkinson's disease and those who are not.

### **1.8.2.2 Minimum Redundancy Maximum Relevance (mRMR)**

The goal of the Minimum Redundancy Maximum Relevance based feature selection, which has found widespread use in many research disciplines. It is to provide the best classification performance by reducing feature redundancy and boosting feature relevance to the target class (Ramezani et al. 2017, Sonu et al. 2017). Among the filter techniques, the inspiration behind using Minimum Redundancy Maximum Relevance approach is that it is efficient in reducing the repetitive features when maintaining the features relevant for the model. It is obvious that the  $m$  optimal features are not necessarily the best  $m$  features, the reason being that several significant features are related and repetitive. Minimum Redundancy Maximum Relevance technique finds a solution to this issue by choosing the features with due consideration to both the relevance for estimating the outcome variable and the redundancy within the chosen features.

### **1.8.2.3 Cuckoo Search Algorithm (CSO)**

Nest eggs denote solutions where cuckoo eggs signify fresh approaches. The main aim is to utilize cuckoos, which are new and maybe superior alternatives, in place of a less than

ideal solution in the nests. Simply said, there is only one egg per nest (Faraji et al. 2019). The method may be altered to handle extremely complicated situations in which each nest includes a number of eggs representing a variety of potential solutions. The effectiveness of utilizing CSA for Parkinson's disease classification with vocal datasets relies on various elements, including the selection of features, the quality of the dataset, and the appropriateness of the chosen classification model. Additionally, to attain optimal outcomes with optimization algorithms like CSA, it may be essential to conduct comprehensive fine-tuning and experimentation. These methods have been introduced for feature selection.

Meta heuristics exhibit a general characteristic: they get the optimal solutions through the exploitations and explorations of the search space. So far, a plenty of metaheuristic algorithms have been examined to find solution to practical combinatorial or global optimization problems like artificial bee colony (ABC), cuckoo search (CS), differential evolution (DE), firefly algorithm (FA), GA, MBO (Wang et al. 2019, Alweshah et al. 2020), and PSO. These algorithms show better performances in comparison with traditional optimization techniques, particularly when used for solving non-convex optimization problems. Exploitation is dominant and will make the exploration become weak. However, when exploration is improved, exploitation is made weak. In search algorithms, it is vital to define a right tradeoff between exploration and exploitation in swarm optimization techniques. BOA is a swarm-based metaheuristic algorithm (Arora & Singh 2019). BOA develops an exploitation and exploration process on the basis of the foraging behavior and information-sharing mechanisms of butterflies.

## **1.9 PROBLEM STATEMENT**

Parkinson's disease is a serious neurodegenerative disorder. It is reported that most of Parkinson's disease patients have voice impairments and are not perceptible to common listeners. Therefore, different machine learning methods have been developed for automated Parkinson's disease detection. However, these methods either lack generalization and clinically significant classification performance or face the problem of subject overlap. Finding the optimal subset is a decisive issue and leads to an extremely high computational cost that also leads to model overfitting, which in turn leads to poor results. When dealing

with high-dimensional data with multiple optimal feature subsets the single feature selection methods can have specific biases leading to performance degradation.

### **1.10 OBJECTIVES OF THE RESEARCH**

- To extract, transform the features using dimensionality reduction technique , rank the features using mrMR and use Fuzzy Convolution Long Short-Term Memory Based CNNs for Parkinson's disease classification.
- To optimize the features using Fuzzy Monarch Optimization algorithm and then Classify the Parkinson's disease using Bi-Directional LSTM.
- To apply Ensemble Feature Selection method for the Parkinson's disease classification.
- To apply Ensemble Deep Learning Classifiers for the optimized features.

### **1.11 CONTRIBUTIONS OF THE RESEARCH**

Principal Component Analysis based on Kernels is used to reduce dimensionalities (KPCA). Additionally, an innovative approach that makes use of condensed feature spaces is given. Through reductions and transformations, the following features are reduced: WT (Wavelet Transform), MFCC (Mel-Frequency Cepstral Coefficients), TF (Time-Frequency), TQWT (Tunable Q-factor Wavelet Transforms), and Vocal Fold Features. For instance, classification, CNN uses Fuzzy Convolution Long Short-Term Memory (FCLSTM-CNN).

Introducing a multidimensional approach to Principal Component Analysis (KPCA) based on Kernels and Feature Selection based on the Fuzzy Monarch Butterfly Optimization Algorithm (FMBOA). In this case, the Gaussian fuzzy membership function computes the weight value. Two classifiers are implemented: FMBOA for Feature Selection and Fuzzy Convolution Bi-Directional Long Short-Term Memory (FCBi-LSTM).

Optimization Based Ensemble Feature Selection (OBEFS) has been introduced for features selection. The findings of several FS including FMBOA, Levy Flight Cuckoo Search algorithms (LFCSA), and Adaptive Firefly Algorithms (AFA) are combined using

ensembles which result in three feature subsets and have their correlations assessed by the use of ensembles. For classification purposes, the OBEFS best features are trained on FCBi-LSTM.

## **1.12 ORGANIZATION OF THE THESIS**

Chapter 1 presents an overview on Deep Learning techniques in the diagnosis of Parkinson's disease. The kinds of dimensionality reduction and its types also studied in this chapter. The significance of Feature Selection techniques, Ensemble modeling and deep learning methods are utilized for finding a solution to the problems are explained in detail in this chapter. At the end, the issues observed in the research work, aims and contribution made by the research are also given.

Chapter 2 it outlines the relevant research concerning various approaches to tasks related to identifying Parkinson's disease. Furthermore, it provides a summary of the constraints found in existing systems for Parkinson's disease identification.

Chapter 3 describes a comprehensive explanation of how Kernel-based Principal Components Analysis (KPCA) is applied to achieve dimensionality reduction. Next, the discussion on Feature Selection by mRMR is also provided. FCLSTM-CNN classification techniques also studied in detail. The advantages and drawbacks of the proposed techniques are also elaborately discussed in this chapter.

Chapter 4 provides a detailed explanation of Feature Selection by Fuzzy Monarch Butterfly Optimization algorithm. Then, a thorough examination of the FCBi-LSTM classification method is shown.

Chapter 5 includes explanations of the Optimization Based Ensemble Feature Selection through Fuzzy Monarch Butterfly Optimization Algorithm, Levy Flight Cuckoo Search Algorithm, and Adaptive Firefly Algorithm. Parkinson's disease classifications by FCBi-LSTM are also discovered in this chapter.

Chapter 6 explains how the Optimization Based Ensemble Feature Selection algorithm has been presented with the intent of choosing the features on the basis of Parkinson's

disease dataset. The optimal chosen features are helpful in training an Ensemble Deep Learning classifier. The classifiers such as FCBi-LSTM, CAE (Contractive Autoencoders), and SAE (Sparse Autoencoder) are aggregated through stacked generalization for ensemble classification.

Chapter 7 this section presents the performance comparisons of classifiers with triple feature sets.

Chapter 8 concludes with the results and discusses for the future improvements that could be done.

### **1.13 SUMMARY**

This chapter comprehensively studies about Parkinson's diseases detections in deep learning, their diagnosis system, dimensionality reduction, feature selection, and classification model. The important advantages and drawbacks in deep learning are also studied. The operational process of feature selection and the means of selecting the optimal features are also discussed.