CHAPTER 3

KERNEL BASED DIMENSIONALITY REDUCTION AND FUZZY CONVOLUTIONAL LONG SHORT-TERM MEMORY BASED CONVOLUTIONAL NEURAL NETWORK

3.1 INTRODUCTION

Currently, Parkinson's Disease (PD) has emerged to be the second prominent neurodegenerative disease all over the world. Even if there is no evidence of Parkinson's disease, its symptoms can be mitigated easily if detected in premature stages. The current forecast specifies that nearly 90% of patients having Parkinson's Disease show symptoms of dysphonic. Therefore, vocal measurements could be utilized in the form of an efficient diagnostic tool for PD. Lately, several studies associated with the application of dysphonic indicators for the diagnosis and monitoring Parkinson's Disease have been carried out. But, not one of them provide satisfactory accuracy in the results (Zham et al. 2017). This study has developed an improved model for PD detections to address this issue.

Initial part of the work involves designing an algorithm to operate in the real feature space, computing the reduced feature subset, and applying the classifier afterward. The following stage involves the proposal of a novel method that, through dimensionality reduction and transformation, directly utilizes the new feature space properties (Pereira et al. 2016, Wahid et al. 2015). Fuzzy Convolutional Long Short-Term Memory **(**FCLSTM-CNN) classifier is employed in the conclusion with a variety of feature ensembles, such as Vocal fold features, TQWT, WT, MFCC, and Time frequency features. Ensemble feature sets will be different as fuzzy membership functions compute biases and weights in the proposed classifier. Data from the UCI Machine Learning repository is utilized to train the suggested FCLSTM-CNN classifier.

3.2 Convolutional Neural Network

In this work, PD was identified using FCLSTM-CNN. ConvNets also referred to as CNN in Deep Learning Technique contexts come within the ANN category and are widely used for visual image analysis. Also known as shift invariant or space invariant artificial neural networks (SIANN), these networks adopt a shared-weight architecture in their convolution kernels or filters. This architecture is inclined towards capturing input characteristics, producing translation-equivariant outputs commonly referred to as feature maps. (Sim et al., 2016). CNN offers societalized, multi-layered viewpoints. The term "multilayer perceptron" frequently connotes fully interconnected networks, in which each neuron in the layer is connected to neurons in the layers above it. Since this network has "complete connection," data over fitting can affect them. Common techniques for normalization, or preventing overfitting, include penalising training parameters (such weight decay), or shortening connectivity (ignored connections, dropout, etc.).

CNN employ a different method for normalization. It takes advantage of the data hierarchical structure and combines patterns with increasing complexity by using simpler, smaller patterns built into their filters. CNN therefore rate quite poorly in terms of connectedness and complexity. The structure of interconnections between neurons in biological reactions, which resembles the structure of the animal visual cortex, was the paradigm for convolution networks. Each cortical neuron can respond to stimuli in the receptive region, which is a defined area of the visual field. Because there is some overlap between the receptive fields of different neurons, the whole visual field is covered. CNN need more pre-processing time when compared to other techniques in image classifications. This suggests that, in contrast to conventional algorithms where these filters are created manually, the network learns by automatically learning how to optimize the filters (or kernels). This lack of reliance on preexisting data and human engagement in feature extraction has significant advantages.

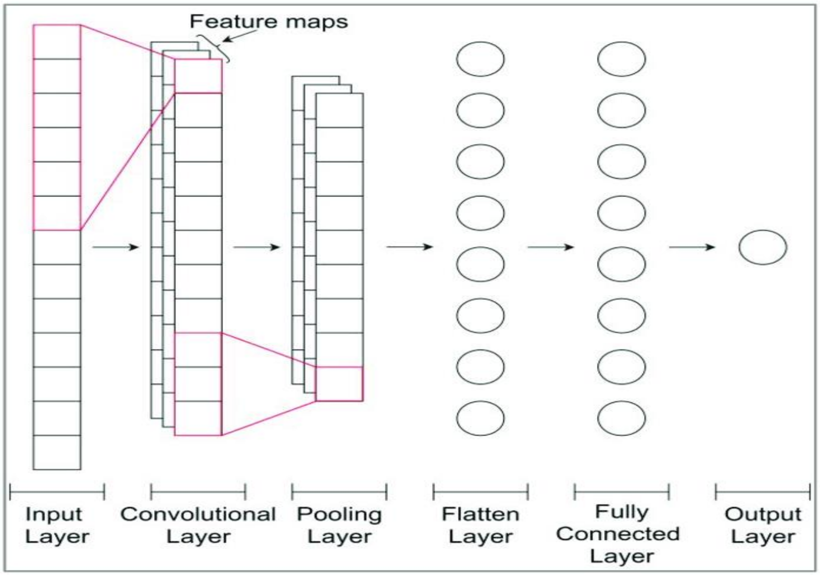


Figure 3.1 Convolutional Neural Network Architecture

Input, hidden, and output layers make up CNN. Intermediary layers of Feed-Forward Neural Network (FFNN) or hidden layers hide their inputs and outputs using activations and final convolutions in these layers. They typically consist of a layer that employs the convolution kernel dot product on the layer input matrix. The convolution process creates a feature map, which once more serves as the input to the subsequent layer, because convolution kernels advance along layer input matrices. The subsequent layers are completely linked layers, pooled layers, and normalization layers.

Long Short-Term Memory (LSTM)

In essence, LSTM are a framework for artificial recurrent neural networks (RNNs) used in DLT as shown in figure 3.2. As opposed to conventional FFNN, LSTM display feedback links. It has the ability to handle an entire sequence of data, not just a single data point (Greff et al. 2016). Normal units in LSTM consist of input, output, forget gates, and cells. The information flowing into and out of the cells which maintain values in its memory over arbitrary time periods are regulated by the three gates. LSTM were created to address the vanishing gradient issue that traditional RNN encountered during training. In comparison to RNN, hidden Markov models, and other sequence learning algorithms, LSTM are advantageous due to their relative insensitivity to gap length in a number of applications.

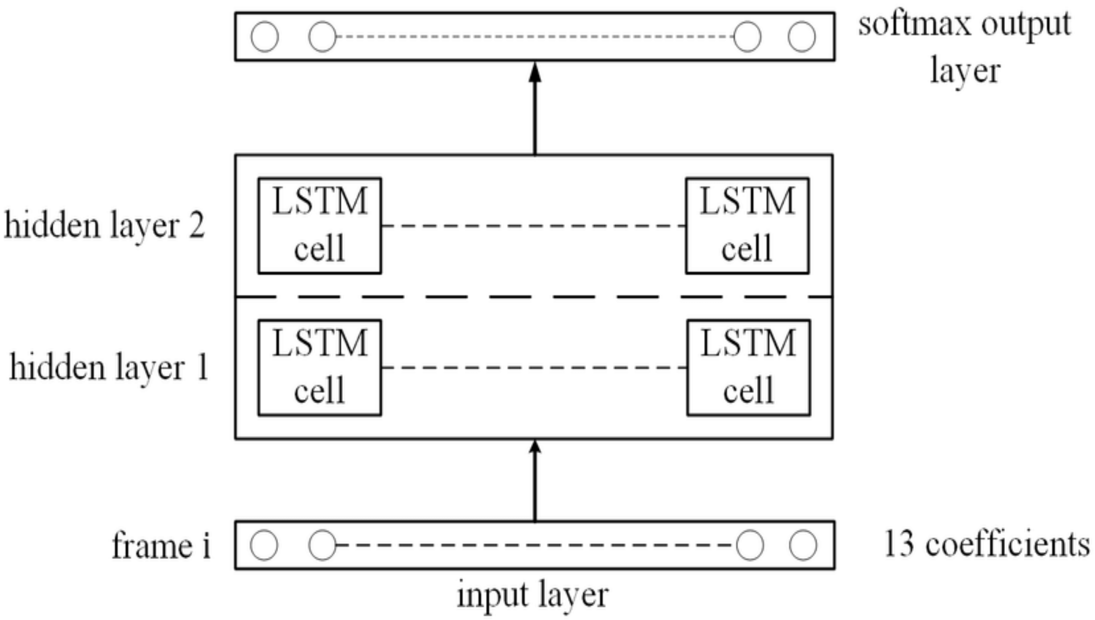


Figure 3.2. LSTM Architecture

3.3 PROPOSED METHODOLOGY

KPCA dimensionality technique is used in this work to reduce dimensionality of the specific dataset before supplying it into classifications for PD. Recurrent features are eliminated, and mRMR selects the features with the highest relevance score based on class label. That feature collection has a total of 60 features that were produced with the assistance of MFCC. Two feature sets are given to the FCLSTM-CNN classifier to integrate the various voice properties at the feature and model levels in order to discern between individuals with Parkinson's disease and those in good health, as well as performances compared to SVM in assessments of the suggested work. In Figure 3.3, the suggested system flowchart is displayed.

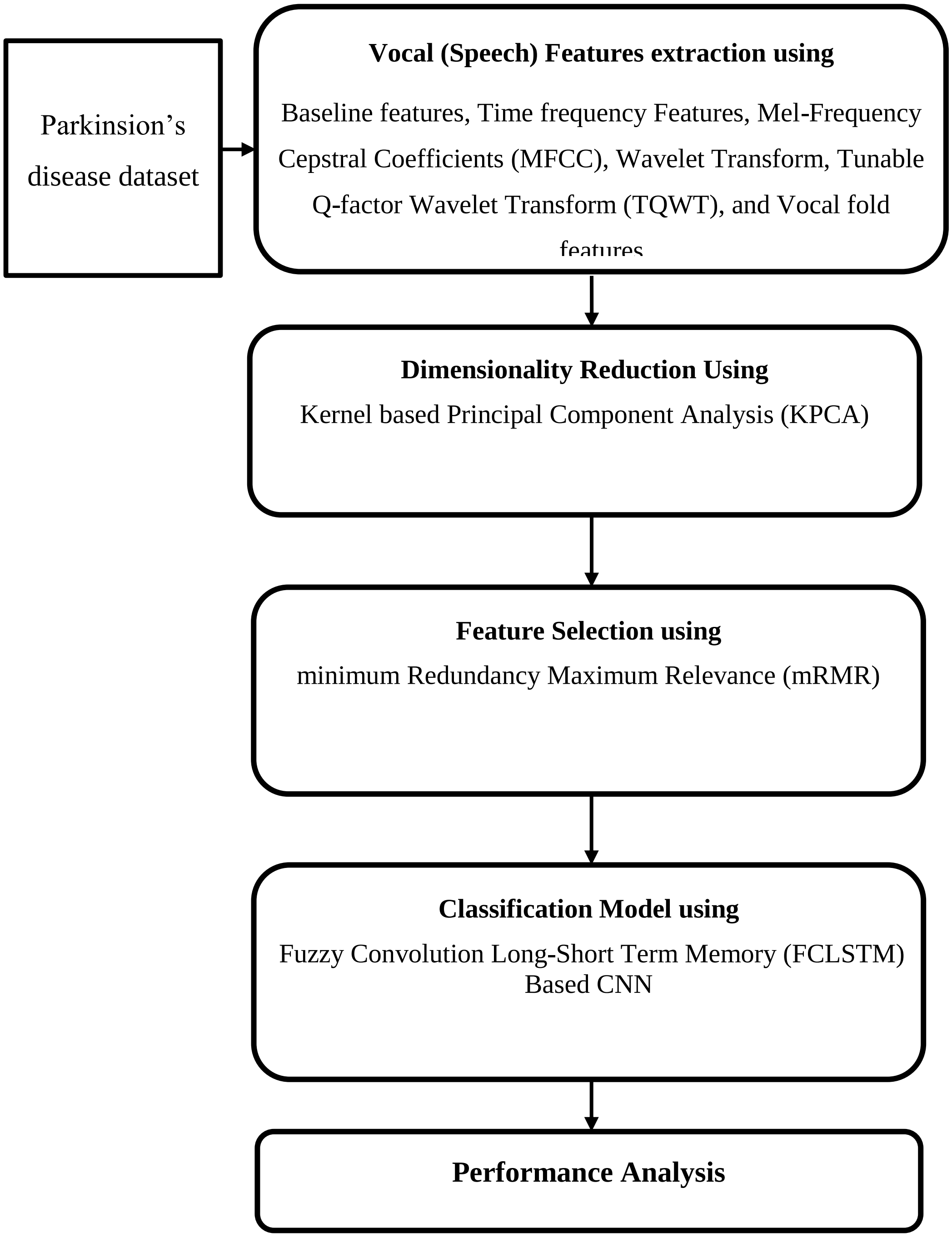
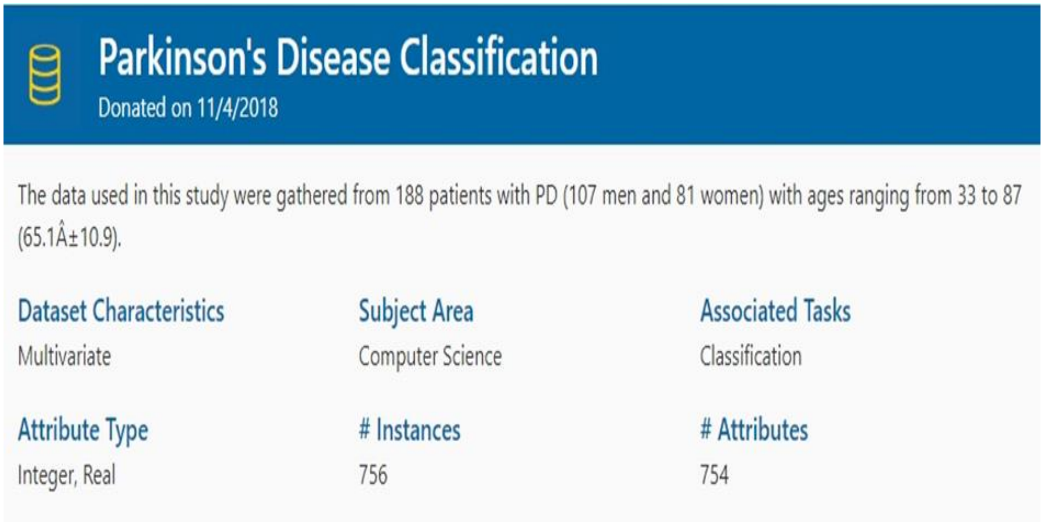


Figure 3.3 Overall Flowchart of the Proposed System

3.3.1 Parkinson’s Disease Dataset

The UCI machine learning repository is the source of the Parkinsion’s disease dataset, which has been utilized in this research. In addition, data is obtained from the Neurology the Department at Istanbul University's Cerrahpasa Faculty of Medicine, which has 188 PD patients 107 men and 81 women, along with 64 normal persons, made up of 41 females and 23 males. Healthy patients were between 41 to 82 years of age while Parkinsion’s disease affected had age ranges of 33 to 87. Microphone frequency response was fixed at 44.1 KHz while collecting the data and after the physician review, vowel /a/ letter repetition again and again in every person is gathered with three duplicated.



Parkinson's Disease Classification

Donated on 11/4/2018

The data used in this study were gathered from 188 patients with PD (107 men and 81 women) with ages ranging from 33 to 87 (65.1Â±10.9).

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Computer Science	Classification
Attribute Type	**# Instances**	**# Attributes**
Integer, Real	756	754

The above dataset is downloaded from the below
https://archive.ics.uci.edu/dataset/470/parkinson+s+disease+classification

Jitter and shimmer variances, fundamental frequency parameters, harmonicity parameters, Recurrence Time Density Entropy (RPDE), Detrended Fluctuation Analysis (DFA), and Pitch Period Entropy (PPE) are commonly employed as foundational features in various research studies, often denoted as baseline features. Additionally, other features, including formant frequencies, intensity parameters, and bandwidth, are extracted using specialized software designed for acoustic analysis.

Mel-Frequency Cepstral Coefficients (MFCCs) have proven to be a robust technique for feature extraction in Automatic Speech Recognition, Speaker Recognition, Biomedical

Voice Recognition, and Parkinson's Disease (PD) diagnosis. MFCCs utilize triangular overlapping filter banks to combine cepstral analysis with spectral-domain partitioning. In Parkinsion’s disease research, this feature set is applied to promptly identify changes in articulatory movements. In the current research data, the 84 features associated with MFCCs are derived from the mean and standard deviation, calculated from the original 13 MFCCs, along with their first and second derivatives, and the log-energy of the signal.

The wavelet transform (WT) is widely employed for analyzing signals, especially in cases involving minor fluctuations at a regional scale. Numerous studies have utilized wavelet transform to extract features from the fundamental frequency (F0) for predicting pathological conditions. Researchers have applied Wavelet Transform -based features to assess the degree of deviation in voice samples, particularly focusing on rapid alterations in the complete periodicity of long-term vowels in unhealthy speech samples. During data collection, voice signals undergo a 10-level discrete Wavelet Transformation to extract wavelet transform -based features from F0 and log transformation.

This procedure generates a total of 182 Wavelet Transform based attributes, encompassing Shannon's energy, log energy entropy, and Teager-Kaiser energy, derived from both detailed coefficients and approximation.

The dataset utilized in this study considers the time-domain properties of speech signals when configuring the parameters of the Tunable Q-factor Wavelet Transform. This involves determining the value of the Q-factor parameter, which governs the oscillatory behavior of wavelets. To avoid undesirable ringing in wavelets, the parameter value needs to be set equal to or greater than three. The accuracy values for various Q-r pairs are assessed by exploring several levels (J) within specified intervals

3.3.1.1 Data Preprocessing:

The dataset consists of 755 features (columns) and one binary label (output feature) that indicates whether the patient's health condition is classified as having Parkinson's disease or not. There are no missing values identified taken from the vocal feature dataset.

Standardization of data involves converting it into a uniform format, facilitating evaluation and processing by users. This process is essential for several reasons. Given that various data columns may exhibit diverse value ranges, encompassing both negative and positive values, as well as a mix of integers and decimals, the absence of a standardized format makes comparison and analysis challenging. Therefore, standardization serves as a rescaling technique, ensuring that the mean of variable values is centered around 0, and the standard deviation is approximately 1.

3.3.2 Feature extraction

The features including Vocal fold, TQWT, WT, MFCC, Time frequency and baseline features were extracted from dataset.

Baseline features: Even during young ages, PD can affect speech and therefore the evaluation of PD can be done with the help of speech characteristics with much ease. Once the medical treatment is done, it can be utilized to monitor its development. Features of speech are pitch period entropy (PPE), detrended fluctuation analysis (DFA), recurrence time density entropy (RPDE), harmonicity parameters, fundamental frequency parameters and Jitter and glow-based features. These features are known as baseline features in the extracted data. Description (#5) indicates that 5 features are extracted from jitter and this is prevalent among all features.

Jitter: In the fundamental frequency, it identifies alterations per cycle (#5).

Shimmer variants: In fundamental amplitude, it identifies alterations per cycle (#6).

Fundamental frequency parameters: Vocal fold vibration maximum and minimum frequency values, standard deviation, median and mean (#5).

Harmonicity parameters: It measures the signal information ratio over noise (in PD speech samples, increased noise elements) (#2).

RPDE: Vocal folds capability to offer stable vocal fold vibrations (#1)

DFA: It measures the turbulent noise random self-similarity (#1).

PPE: To measure the fundamental frequency affected controls through logarithmic scale (#1).

Time frequency Features: In this**,** Bandwidth, Formant Frequencies and Intensity Parameters features are present.

Intensity Parameters: Decibels (dB) representing the maximum, minimum, and mean strength of the voice signal (#3).

Formant Frequencies: Frequencies, whose amplification is done using the vocal tract (first four formants) (#4).

Bandwidth: Ranging from formant frequencies to other frequencies (first four bandwidths) **(**#4).

MFCC: Divisions of Spectral domains and cepstral analyses are integrated in extractions approaches of MFCC which employ triangular overlapping filter banks. In data, there are nearly 84 MFCC relevant features and these features are obtained from the actual 13 MFCC mean and standard deviation along with the signal long energy and with their first and second derivatives. In vocal tract, for catching PD affects separately from vocal folds (#84).

WT: In signal shavings light variations in regional scale, decisions are achieved with the popular prominent tool known as WT. Different studies use Particular features computed through the WT from the unprocessed basic speech signal frequency (F_0), to diagnosis PD. The data gathered for the extraction of WT-based features generated from crude (F0) contour and log modification of (F0) contour uses speech sound with 10-level discrete wavelet transformation. With the help of this method, about 182 characteristics that are dependent on WT are produced. These characteristics include Shannon, log energy entropy energy, and Teager-Kaiser energy of precise and approximation coefficients.

Tunable Q-factor WT: The use of another method known as TQWT can be used to extract the characteristics. TQWT have the advantages of three tunable parameters, such as Q-Q factor, r-redundancy, and J-levels count, to transform the signals in a higher quality based on the nature of the signal. Signal oscillation count in the time domain is related to the Q-factor parameter. During the decomposition phase, J-Represents the number of levels. Once decomposition is over, there are J+1 sub-bands that one final low-pass filter generates along with J high pass filter outputs. In time, the localization of the wavelet can be achieved with parameter r, which regulates extreme ringing without the shape being affected. Nearly 432 TQWT-related features are generated in this dataset with different experiments.

Vocal fold features: On vocal fold, to investigate the effects of noise, vocal fold vibration based features are used. In the same scenario, features such as Empirical Mode Decomposition (EMD), vocal fold excitation ratio (VFER), glottal to noise excitation (GNE) and glottis quotients (GQ) are brought into use.

GQ: It indicates the glottis opening and closing times (#3).

GNE: It measures the threshold of turbulent noise. Incomplete vocal fold closure generates this (#6).

VFER: It measures the noise quantity generated due to pathological vocal fold fluctuation (#7).

EMD: Adaptive basis functions are utilized for the generation of fundamental signal components through the decomposition of speech signal and using these components, entropy and energy values are derived.

Concat features: The ensemble of time-frequency, vocal folds and baseline characteristics develops the Concat features.

3.3.3 Dimensionality reduction using KPCA

Dimensionality reduction can be achieved with the help of a well-known technique known as Kernel Principle Component Analysis (KPCA). In comparison to the actual sound-recording feature spaces, wherein the new sound-recording characteristics of Parkinsion's disease exhibit significant variation, a lower dimensions linear subspace is generated with the help of KPCA. Suppose $\{a_i\}, i = 1, \dots N$ to be Parkinson Disease (PD) dataset and D-dimensional sound recorded features vector is denoted as a_i. An M-dimensional sound reordered feature subspace is mapped with Parkinson Disease data, where $< D$. Equation (3.1) has been given,

$$b = Xa, \qquad X = [u_1^T, \dots u_M^T], u_k^T u_k = 1, k = 1, \dots M \tag{3.1}$$

It needs the maximized $\{b_i\}$ variance. Therefore, X has to be computed applying the Equation (3.2) as,

$$X^* = \arg\max_X tr(S_b) \tag{3.2}$$

$$S_b = \frac{1}{N}\sum_{i=1}^{N}(b_i - \bar{b})\left(b_i - \bar{b}\right)^T \tag{3.3}$$

$$\bar{b} = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{3.4}$$

Suppose dataset $\{a_i\}$ covariance matrix to be S_a. Hence,

$$tr(S_b) = tr(XS_aX^T) \tag{3.5}$$

Lagrangian multipliers are utilized for deriving the Equation (3.5) and the last Equation (3.6) is got by computing its derivative as,

$$S_a u_k = \lambda_k u_k \tag{3.6}$$

It is found that, u_k indicates S_a eigenvector. Equation (3.7) is considered for defining a_i as,

$$a_i = \sum_{k=1}^{D}(a_i^T u_k)u_k \tag{3.7}$$

Equation (3.8) is utilized for the approximation of a_i as,

$$\tilde{a}_i = \sum_{k=1}^{M}(a_i^T u_k)u_k \tag{3.8}$$

here, S_a eigenvector is denoted as u_k , which is associated with the kth highest Eigenvalue. Using the actual D-dimensional sound recorded features vector space, nonlinear transformation $\phi(a)$ is used to establish a Kernel matrix of M dimensional feature space using the sound recorded feature vector, with $M < D$. So, each PD $(a_i)$is mapped onto a point $\phi(a_i)$. Take zero mean of projected new sound recorded features applying Equation (3.9).

$$\frac{1}{N}\sum_{i=1}^{N}\phi(a_i) = 0 \tag{3.9}$$

The mapped sound recorded features covariance matrix is given by M × M by Equation (3.10),

$$Co = \frac{1}{N}\sum_{i=1}^{N}\phi(a_i)\,\phi(a_i)^T \tag{3.10}$$

Equation (3.11) provides its eigenvectors with eigenvalues.

$$Cov_k = \lambda_k v_k \tag{3.11}$$

Where, k = 1, 2,….., M. Using Equation (3.10-3.11),

$$\frac{1}{N}\sum_{i=1}^{N}\phi(a_i)\,\{\phi(a_i)^T v_k\} = \lambda_k v_k \tag{3.12}$$

Equation (3.13) has been described as,

$$v_k = \sum_{i=1}^{N} y_{ki}\phi(a_i) \tag{3.13}$$

Value of $v_k$gets substituted in Equations (3.12-3.13) and it yields,

$$\frac{1}{N}\sum_{i=1}^{N}\phi(a_i)\phi(a_i)^T\sum_{j=1}^{N} y_{kj}\phi(a_i) = \lambda_k\sum_{i=1}^{N} y_{ki}\phi(a_i) \tag{3.14}$$

The feature vector of the Sound recording with reduced dimension is computed from above Equations. Later, mRMR is considered for the Feature Selection.

3.3.4 mRMR based feature selection

The goal of the mRMR based feature selection, which have found widespread use in many research disciplines. Itis to provide the best classification performance by reducing feature redundancy and boosting feature relevance to the target class (Ramezani et al. 2017). Among the filter techniques, the inspiration behind using mRMR approach is that it is efficient in reducing the repetitive features when maintaining the features relevant for the model. In the case of the stated PD classification task, mRMR grades the significance of PD features set. As per the relevance to the target, PD features are graded in this approach and the redundancy in PD features is also penalized. Between class c and PD features set 'fe', computing the maximum dependency forms an important goal and it makes use of Mutual Information (MI), which is denoted by I. Equation (3.15) specifies, MI between a PD features pair and it can be attained with efficiency using known joint probability pr(x, y), marginal probabilities pr(x) and pr(y) between PD features pair.

$$I(X,Y) = \sum_{y \in Y} \sum_{x \in X} pr(x,y) \log\left(\frac{pr(x,y)}{pr(x)pr(y)}\right) \tag{3.15}$$

Maximum relevance searches those PD features, satisfying the expression below,

$$\max D(A,c)\,; D = \frac{1}{|A|} \sum_{A_i \in A} I(A_i, c) \tag{3.16}$$

A high level of redundancy may be developed, when PD features are chosen on the basis of maximum relevance condition. Therefore, Equation (3.17) is included with minimum redundancy condition as,

$$\min R(A)\,; R = \frac{1}{|A|} \sum_{A_i, A_j \in A} I\left(A_i, A_j\right) \tag{3.17}$$

The mRMR condition is generated through the combination and optimization of the parameter R and D. Greedy algorithm is used for the same practically, where, the chosen PD features set is denoted as S.

$$\max_{A_i \notin S} \left[I(A_i; c) - \frac{1}{|S|} \sum_{A_j \in S} I\left(A_j; A_j\right) \right] \tag{3.18}$$

Algorithm 3.1 depicts the process involved in the mRMR variant. In this algorithm, between two components, which may be an input of PD features pair (line 7) or the stated PD feature with class (line 4). Between each PD features pair, this function is computed, in spite the fact that, the ultimate classification is not affected by PD features pair due to their irrelevance.

Algorithm 3.1. FS algorithm based on MRMR

Input: dataset D, Chosen features count $S_{fe}$from dataset 'D'

Output: lastly selected features$St_{Fe}$from dataset 'D'

```
Start
For each feature $Fe_i$in Dataset (D)
do
    Relevance =mutual Info ($Fe_i, class$); Redundancy =0;
For each feature $Fe_j$in Dataset (D)
  Do
          Redundancy += mutual Info ($Fe_i, Fe_j$);
End for
  mRMR values [$Fe_i$] =relevance-redudancy;
End for
$St_{Fe} = sort$(mrmrvalues)$.found(S_{fe})$
End
```

Average distribution of the top-50 features selected by the mRMR filter after KPCA dimensionality reduction

Features	**All Feature Subsets**	**All Feature subsets except TQWT**	**All Feature subsets except MFCC**
Baseline (n = 26)	4	5	5
Intensity (n = 3)	0	1	1
Bandwidth + Formant (n = 8)	2	5	2
MFCC (n = 84)	10	27	-
WT applied to F0 (n = 182)	1	4	1
Vocal Fold (n = 22)	3	8	4
TQWT (n = 432)	30	-	37

3.3.5 PD classification through FCLSTM-CNN classifier

In this research, FCLSTM-CNN is utilized in PD classification. For PD classification, proposed FCLSTM-CNN classifier is used in this work. A FCLSTM-CNN model is introduced and figure 3.4 illustrates the architecture of this classifier. Fuzzy weights are computed in FCLSTM-CNN classifier employing the membership function and to extract the highly relevant estimated Parkinson features, Parkinson features response is modified. To achieve step-by-step Parkinson classification, hidden states and Parkinson features are utilized in FCLSTM-CNN and in the inherent memory state; context data are stored so that correlation between PD features can be mined.

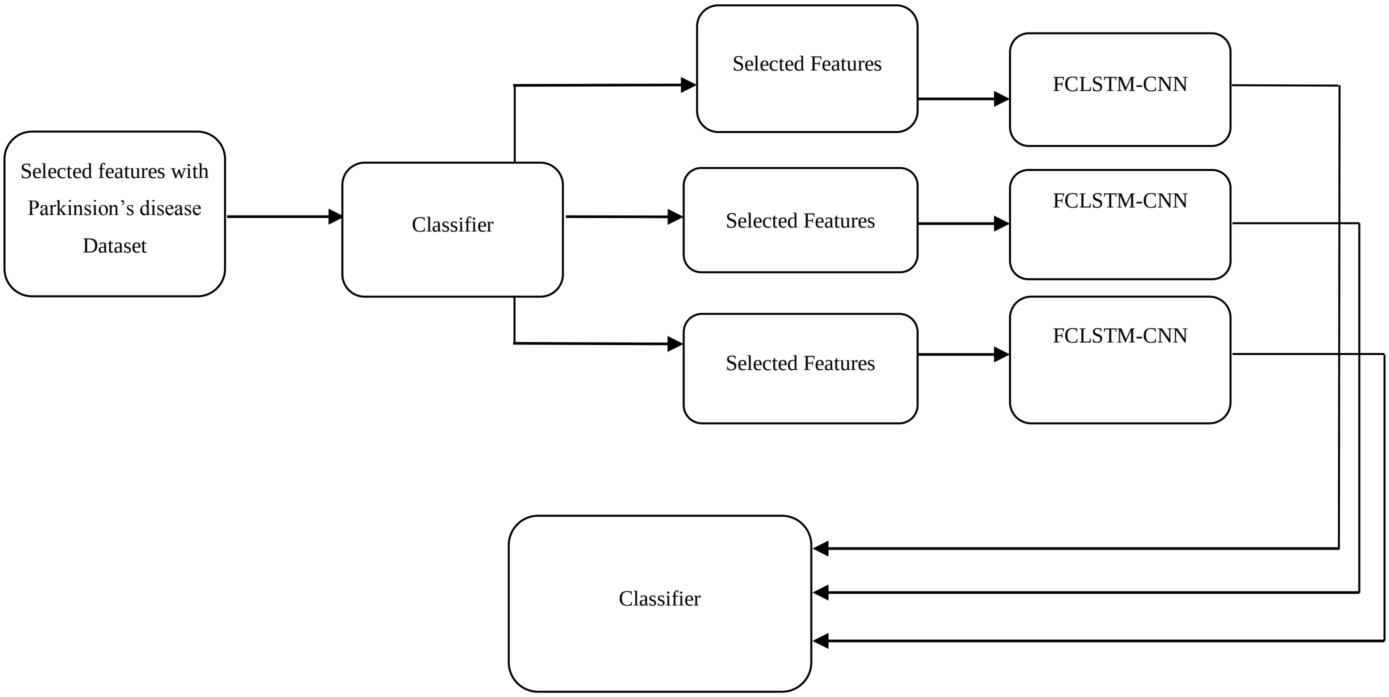


Figure 3.4 Architecture of Proposed FCLSTM-CNN Classifier

To examine the correlation between features, FCLSTM is utilized in this work and the prediction of the features is done sequentially. During the next prediction of the genes, hidden state is associated with CLSTM with past data. In sequence modeling, LSTM shows improved performance. This is owing to the fact that, typically LSTM classifier, data sequence is sent through a full connection layer and input data is projected onto a one-dimensional vector, resulting in the data being lost. To achieve PD performance enhancement, it is necessary to have the relevant data. In this classifier, CLSTM (Xingjian et al. 2015) is utilized to maintain the structure of PD features. In FCLSTM, the input is transformed to state and the conversion between states is achieved with the help of convolution operation. Features data can be acquired superior in comparison with standard LSTM and between PD features and the correlation can be achieve defectively. The Equations (3.19-3.23) of FCLSTM has been described as,

$$i_t = sigmoid(FW_{ix} * x_t + FW_{ih} * h_{t-1} + FW_{ic} \odot c_{t-1} + b_i) \qquad \textbf{(3.19)}$$

$$f_t = sigmoid\left(FW_{fx} * x_t + FW_{fh} * h_{t-1} + FW_{fc} \odot c_{t-1} + b_f\right) \qquad \textbf{(3.20)}$$

$$o_t = sigmoid(FW_{ox} * x_t + FW_{oh} * h_{t-1} + FW_{oc} \odot c_t + b_o) \qquad \textbf{(3.21)}$$

$$g_t = tanh\left(FW_{gx} * x_t + FW_{gh} * h_{t-1} + b_g\right) \qquad \textbf{(3.22)}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t, h_t = o_t \odot tanh(c_t) \qquad \textbf{(3.23)}$$

where, * stands for the logistic sigmoid functions, *tanh*(·) denotes the hyperbolic tangent functions, the subscripts t specifies the tth step of CLSTM, i_t signifies the input gate, f_t stands for the forget gate, o_t signifies the output gate, g_t represents the input modulation gate, x_t stands for the input data, c_t indicates the cell states,h_t represents the hidden states. Three dimensional tensors are expressed using by $x_t, c_t, h_t, i_t, f_t$and o_t. The PD dataset is represented by the first dimension, and its rows and columns are shown by the second and third dimensions. Each gate, fuzzy weight values are denoted using $W_{ix}, FW_{ih}, FW_{ic}, FW_{fx}, FW_{fh}, FW_{fc}, FW_{ox}, FW_{oh}, FW_{oc}, FW_{gx}, FW_{gx}$. Fuzzy membership function with initial weights $W = (w_1, \ldots w_i)$, i=1,...n is utilized for the generation of these weights, where,n represents the samples count of gene chosen dataset. The PD dataset features are stored here using the convolution procedure. The LSTM and CLSTM cores are nearly identical. The inputs of next layers are linked to the outputs of previous layers. Moreover, the data will affect the Convolution operations due to potential activating response of regions that belong to the targeted pedestrian features. CLSTM highlights on the key regions of pedestrian attribute prediction, its yields excellent performance, in comparison with the common LSTM as the experiments reveal. Figure 3.5 illustrates the inherent structure with Fuzzy weight values of CLSTM.

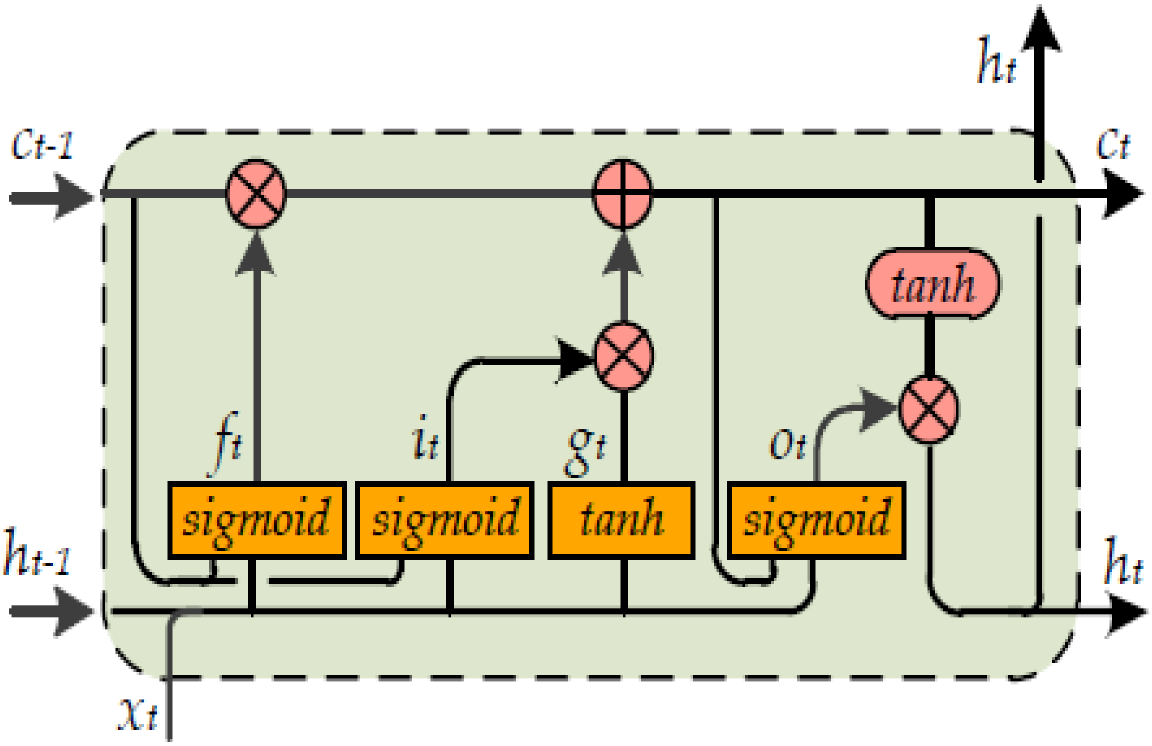


Figure 3.5. Internal Structure of FCLSTM Classifier

Triangular membership function is utilized for the computation of fuzzy weights. An upper limit b, lower a and a value m, whose range is between a and b are utilized for specifying triangular function. Equation (**3.**24) formulates the triangular membership function.

$$\mu_A(W) = \begin{cases} 0, W \leq a \\ \dfrac{W-a}{m-b}, a < W \leq m \\ \dfrac{b-W}{b-m}, m < W < b \\ 0, W \geq b \end{cases} \tag{3.24}$$

The fuzzy weights are generated applying this and the above stated Equations (**3.**19-**3.**23) are provided using these values.

3.4 EXPERIMENTAL RESULTS

The experimental results of the suggested FCLSTM-CNN classifier are explained and contrasted with those of other methods, including Subtractive Clustering Features Weighting Approach (SCFW) - Kernel-Based Extreme Learning Machine (KELM), Support Vector Machine, and Convolution Neural Network. Every class label is determined by taking majorities of class labels allocated to recordings since there are 3 recordings for every individual.

The parameters used for triangular membership function are upper limit which is set to 4, lower limit set to 0 and peak is set to 2. The number of hidden layers used in LSTM is 100. Here the constrains used is upper limit, lower limit and peak and the optimal value which is obtained is 4,0,2 for the given constrains. In classification the constrains and its optimal values are given as follows,

LSTM hidden layer=100

Fuzzy membership function used variable FW→ measured by trimf (triangular membership function.

$$\mu_A(W) = \begin{cases} 0, W \leq a \\ \frac{W-a}{m-b}, a < W \leq m \\ \frac{b-W}{b-m}, m < W < b \\ 0, W \geq b \end{cases}$$

Evaluation Metrics: The predictability performances of classifiers can be evaluated in terms of performance metrics. Accuracy is a popular metric and in data with imbalanced class distribution, its results are not satisfactory. The capability of the classifier in differentiating between different classes, even during class unbalance is measured in terms of a metrics such as Matthews Correlation Coefficient (MCC) and F-Measure. It has been described by Equations (3.25-3.27),

$$Precision = \frac{tp}{tp + fp} \tag{3.25}$$

$$recall = \frac{tp}{tp + fn} \tag{3.26}$$

$$F - measure = \frac{2 * precision * recall}{precision + recall} \tag{3.27}$$

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} \tag{3.28}$$

$$MCC = \frac{(tp * tn) - (fp * fn)}{\sqrt{}(tp + fp)(tp + fn)(tn + fp)(tn + fn)} \tag{3.29}$$

The number of tn, fn, fp, tp are taken into consideration in MCC and it is presumed to be a balanced metric due to its usage in imbalanced class distribution. Between predicted and the real instances, the MCC indicates the correlation coefficient and its value ranges between -1 and +1. The accurate prediction is specified with +1 and discord between the real and predicted labels is denoted with -1 value. The confusion matrix for two-class classification is illustrated in table 3.1. Table 3.1 shows the confusion matrix and in the case of a binary classification, the count of wrongly and correctly classified instances per class is given in this table. tn, fn, fp and tp specifies the counts of true negatives (tn), false negatives (fn), false positives (fp) and true positives (tp) in that confusion matrix. F-Measure is computed in terms of these counts.

TABLE 3.1. Confusion Matrix for Two-Class Classification

ACTUAL/ PREDICTED AS	**POSITIVE**	**NEGATIVE**
POSITIVE	**TP**	**FN**
NEGATIVE	**FP**	**TN**

True Positive (TP): denotes the number of instances accurately identified as positive

False Negative (FN): represents instances wrongly identified as negative when they are actually positive.

False Positive (FP): signifies instances incorrectly identified as positive when they are actually negative.

True Negative (TN): indicates instances correctly identified as negative.

In simpler terms, these metrics help assess how effectively a classification model distinguishes between positive and negative cases.

Results Comparison: Three kinds of features and classifiers such as the available Support Vector Machine, Convolutional Neural Network and proposed FCLSTM-CNN classifiers are utilized in the experiments. Classifier performance is evaluated in terms of metrics like accuracy, f-measure and Mathews Correlation Coefficient. In the case of novel experiments, triple feature sets are used in feature-level ensembles.

There are 4 feature types in the dataset such as TQWT, Wavelet, MFCC and Concat (Concatenation of baseline, vocal fold and time frequency features). Due to the number of individuals in the dataset, prediction performances of the frameworks are evaluated with LOPO CV using F-Measure and MCC metrics.

Using leave-one-person-out cross-validation is a good way to classify Parkinson's disease using vocal data. This method means we systematically exclude one person's data at a time to check how well our model works. It makes our classification results more trustworthy.

TABLE 3.2. Results of FCLSTM-CNN Classifier with Triple Feature (KPCA+ mRMR)

FEATURE EXTRACTION COMBINATION	ACCURACY (%)	F-MEASURE (%)	MCC (%)
TQWT+MFCC+Wavelet	92.1457	93.0257	66.3669
TQWT+MFCC+Concat	92.1854	90.3252	66.4060
TQWT+Wavelet + Concat	92.2252	92.2252	64.1457
MFCC +Wavelet + Concat	94.2557	90.4729	65.9960

Table 3.2 shows the results of classification. Nearly 86.52% accuracy rate is obtained with TQWT, MFCC and Wavelet features along with 90.14% F-Measure rate, while, 83.6% accuracy rate is achieved with TQWT+MFCC+Concat ensemble and 84.25% is attained with TQWT+Wavelet+Concat ensemble. As far as MCC is concerned, F-Measure and accuracy scores, very low performance is shown through feature ensembles with no TQWT features (MFCC+Wavelet+Concat).

Table 3.3. Accuracy Comparison Results of Classifiers with Triple Feature with KPCA+ mRMR

Feature Extraction Combination	FCLSTM-CNN	CNN Classifier	SVM Classifier
TQWT+MFCC+Wavelet	92.1457	86.6697	84.2025
TQWT+MFCC+Concat	92.1854	91.0315	82.7640
TQWT+ Wavelet + Concat	92.2252	87.1695	86.4662
MFCC + Wavelet + Concat	94.2557	93.2752	87.2293

Three feature sets are used to train the FCLSTM-CNN. The triple feature findings reveal that the MFCC + Wavelet + Concat ensemble performs better than any other classifiers when looking at all metrics. According to table 3.3, the accuracy rate of the FCLSTM-CNN is 94.2557%, CNN classifier is 93.2752%, whereas the accuracy rate of this ensemble is 87.2293%. CNN classifier and features boost performance in MFCC+Wavelet+Concat, TQWT+MFCC+Concat, TQWT+Wavelet+Concat, and TQWT+MFCC+Wavelet. The FCLSTM-CNN classifier is trained using a triple feature set. In regard to each metric, among every classifier, MFCC + Wavelet + Concat ensemble shows remarkable performance in triple feature results as depicted in table 3. 3.

Table 3.4. F-measure Comparison Results of Classifiers with Triple Feature with KPCA+ mRMR

Feature Extraction Combination	FCLSTM-CNN	CNN Classifier	SVM Classifier
TQWT+MFCC+Wavelet	93.0257	84.6697	81.7152
TQWT+MFCC+Concat	90.3252	89.0315	79.5960
TQWT+ Wavelet + Concat	92.2252	85.1695	85.1589
MFCC + Wavelet + Concat	90.4729	91.2752	86.3510

FCLSTM-CNN classifier in MFCC + Wavelet + Concat feature ensemble yields 90.4729% of F-Measure rate. The F-measure values of CNN classifier is achieved by MFCC+Wavelet+Concat feature ensemble yields 91.2752% and SVM classifier is achieved by MFCC+Wavelet+Concat feature ensemble yields 86.3510%.

Table 3.5. Mathews Correlation Coefficient Comparison Results of Classifiers with Triple Feature with KPCA+ mRMR

Feature Extraction Combination	FCLSTM-CNN	CNN Classifier	SVM Classifier
TQWT+MFCC+Wavelet	66.3669	55.9007	54.9000
TQWT+MFCC+Concat	66.4060	60.1656	53.3000
TQWT+ Wavelet + Concat	64.1457	62.0993	57.5000
MFCC + Wavelet + Concat	65.9960	63.2384	58.4000

FCLSTM-CNN exhibits superior Mathews Correlation Coefficient (MCC) values across various feature extraction combinations, implying its efficacy in capturing intricate patterns within the data. Notably, the combination TQWT+MFCC+Concat achieves comparatively higher MCC values, reaching 66.3669%, across all classifiers than other combinations. This suggests its potential for enhanced classification performance in the specified context. The Convolutional neural network gives 60.1656% in the TQWT+MFCC+Concat and Support Vector Machine achieves 53.3000% of MCC value in the combination of TQWT+MFCC+Concat.

3.5 SUMMARY

This research uses a collection of speech or vocal cues to classify Parkinson's disease, and a deep FCLSTM-CNN classifier is proposed to achieve this goal. With this goal in mind, a Convolution Neural Network-based model is created that aids in distinguishing between Parkinson's disease patients and healthy individuals. Only a few characteristics from the input dataset are retrieved for the categorization of Parkinson's disease detection. KPCA is used to reduce the dimensionality of the features were extracted. MRMR is utilized for choosing the meaningful features from the whole features set. As per the class label, mRMR chooses the features with better relevance score and repetitive features are removed. Based on the relevance between target and Parkinson's disease features, Parkinson's disease features are graded using the minimum Redundancy Maximum Relevance approach. It also helps penalizing the redundancy in Parkinson's disease features. Finally, correlation between features is investigated applying the FCLSTM classifier and the prediction is done. Dataset obtained from UCI Machine Learning repository was utilized to train this classifier. Compared with other classifier, proposed FCLSTM-CNN classifier attains better results.