

# Bionic Hand Movements Recognition: A Unified Framework with Attention-Guided ROI Identification and the Bionic Fusion Net Approach

Prakash. S<sup>1</sup>, Josephine H. H<sup>2</sup>, Priya. S<sup>3</sup>, M. Batumalay<sup>4</sup>

Department of Electrical and Electronics Engineering, Bharath Institute of Higher,  
Education and Research, Chennai, 600073. Tamil Nadu, India<sup>1,2</sup>

Department of Electrical and Electronics Engineering, AMET Deemed to be University,  
Kanathur, Chennai 603112, Tamil Nadu, India<sup>3</sup>

Faculty of Data Science and Information Technology, INTI International University, Malaysia, Nilai, Malaysia<sup>4</sup>

**Abstract**—In prosthetics, bionic hand movement recognition is crucial to developing sophisticated systems that can effectively understand and react to human motions. Recent advances in image processing, feature extraction, and deep learning have improved bionic hand movement detection systems' accuracy and flexibility. This study proposes a unified framework called using attention-guided ROI detection and a unique Bionic Fusion-Net architecture to overcome these difficulties which contributes towards Sustainable Development Goal (SDG) Good Health and Well Being. Initially pre-processing undergoes dataset augmentation and image enhancement. The ROI Identification approach uses an attention-guided U-Net with sophisticated convolutional components. Spatial Features, BionicNet-1, and BionicNet-2 learn spatial and temporal features together during feature extraction. Optimized Red Fox Falcon Algorithm (O-RFF) which is a hybrid of Red Fox and Falcon Optimization Algorithms improves the feature selection. The Bionic Fusion-Net Architecture combines Xception, Squeeze-Net, Shuffle-Net, optimized Bi-LSTM, and Huber Loss function application. The recommended technique improves bionic hand movement recognition flexibility that attained an accuracy of about 99% which outperformed other approaches in use for well-being and future health policy.

**Keywords**—Bionic Hand; Optimized Red Fox Falcon Algorithm; Xception; Squeeze-Net; Shuffle-Net; Bi-LSTM; Huber Loss; Sustainable Development Goals (SDG); good health; well-being; health policy

## I. INTRODUCTION

The human hand is composed of three essential bone groups—phalanges, metacarpals, and carpals—which together contribute to the hand's dexterity and usefulness. The engineering of a 3D hand model seeks to accurately reproduce the complex anatomy of the hand. This is achieved by specifically addressing the distal, middle, and proximal phalanges, and combining the metacarpal and carpal bones into a cohesive unit to enhance overall functioning [1]. This engineering is significant because to its potential application in prosthetics, namely in the evolution of bionic hands. People who have physical impairments, strokes, or damage to their nervous system frequently struggle with a decrease in their capacity to use their hands normally. This affects their ability to adapt to complicated situations and makes it difficult for them to do

everyday chores. Prosthetic devices are becoming a viable way to tackle these issues. However, existing choices are often limited in terms of flexibility and the range of gestures they can do. There is a need for increasingly sophisticated prosthetic technology in order to accurately replicate the subtle motions of the human hand. The identification of motions in a bionic hand plays a crucial role in connecting human intention with the reaction of an artificial limb [2, 11]. The control system and interface of the prosthetic hand should be simple and user-friendly [8]. The firms engaged in the development of bionic prostheses are primarily focused on two specific areas of advancement. These prostheses are more affordable. Enhancing the functionality of the prosthesis control system [14]. There are two main sensor control methods. One uses stretch sensor to imitate the motions of a human hand, while the other uses machine learning (AI) and a camera for more advanced control. Electromyography (EMG) pattern recognition algorithms are fundamental for categorizing and regulating hand motions [3].

Surface electromyogram (sEMG) is a method used to gather electrical impulses produced by muscles [12, 13]. It has potential in the field of bionics and is being used in wearable devices for medical and healthcare reasons. Myoelectric pattern-recognition algorithms [10] based on EMG data, aid in the identification of finger motion intents, which is essential for commanding prosthetic robots that imitate known intentions [4, 7, 9]. Hand gesture pattern recognition algorithms may be divided into two categories: traditional approaches and deep learning techniques. These algorithms aim to tackle issues associated with identifying hand poses and trajectories, as well as performing regression of continuous parameters [5]. Robotics, namely in the realm of bionic limbs, is crucial in both engineering and medical fields. It provides mechanically and electrically driven alternatives to replace missing limbs in individuals who have had amputation [6].

The creation of bionic hands is an important achievement, considering the nearly 1.7 million individuals in the United States who have experienced limb loss. The control system and interface of prosthetic hands should be designed to be simple and intuitive, in order to provide broad accessibility and usage. The extensive motion planning algorithms are necessary to provide stable and dexterous grasp control of five-finger bionic

hands, which are characterized by a high number of Degrees of Freedom (DoFs) [8]. Machine learning algorithms play a crucial role in hand gesture detection as corporations prioritize the development of cost-effective prostheses and advancements in control systems. Time and frequency characteristics, which are used as inputs for algorithms like Artificial Neural Networks (ANN), Support Vector Machines (SVM), and k-Nearest Neighbor (k-NN), play a role in classifying hand gestures [15]. The main contribution of this paper is as follows:

- Initially the dataset augmentation and Image enhancement is done.
- ROI are identified by employing a novel approach called Attention guided U-Net with atrous convolution.
- The spatial and temporal features are extracted and the features are selection by employing a novel hybrid approach called Optimized Red Fox Falcon Algorithm (O-RFF).
- Finally, Bionic Fusion-Net Architecture is developed which is a combination of Xception, Squeeze-Net, Shuffle-Net and optimized Bi-LSTM and Huber loss function for the classification.

The remaining section is provided below. Section II discusses the relevant papers, Section III describes the suggested methodology, Section IV shows the results and discussions and finally, Section V concludes the paper.

## II. RELATED WORKS

Hui Li et al. [16] discussed about the development of human-computer interaction and the use of different devices such as Leap Motion, Gloveone, and Lingzhi. These devices allow users to control computers and perform various operations using gestures. The text highlights the accuracy and speed of the bionic manipulator in accurately interpreting instructions and enabling human-computer interaction in different situations.

Sapto Budi Priyatno et al. [17] provided motor imagery brain wave categorization research for bionic hand control. It created a bionic hand control system using EEG sensors to assess brain bioelectric activity. Alpha and beta waves, associated to motor imagery, are extracted and classified in the research. Features are extracted using the Fast Fourier Transform (FFT) approach and categorized using the Multilayer Perceptron (MLP) method for five bionic hand movement classes. Tests show system accuracy of 77.20% and 84.40% for two situations. The research shows that motor imagery can operate a bionic hand.

Binish Fatimah et al. [18] proposed a Surface Electromyogram (sEMG)-based hand movement recognition method. Entropy and kurtosis are calculated for each Fourier intrinsic band function (FIBF) after the Fourier decomposition method (FDM) decomposes the signals. Train machine learning classifiers using statistically significant features. On two publicly accessible datasets, the suggested technique outperforms current algorithms in accuracy of 93.53% on NinaPro DB5 and 99.49% on the UCI dataset. Its Fourier theory foundation makes it appropriate for real-time implementation with minimal processing cost. The program might help develop efficient and user-friendly prosthetic hands.

Sehyeon Kim et al. [19] suggested a multimodal fusion system for transf forearm amputee hand movement detection. The method improves motion categorization by combining EEG and EMG inputs. Use a transfer learning paradigm and convolutional neural network technique to train a model from 2D EEG and EMG input pictures. Comparing the proposed approach to single-model EEG signal trained models demonstrates considerable classification accuracy increase. The research involves five transf forearm amputees and nine healthy controls. Both groups show that multimodal fusion works for motion categorization. This method may improve amputee prosthetic arm control.

Shudi Wang et al. [20] emphasized the necessity of building a dexterous hand control system to assist forearm disabled people regain limb functionality. The use of EMG signals for hand control gesture recognition is investigated. Deep learning may improve identification accuracy compared to traditional machine learning. An attention module is added to a multi-stream fusion network using CNNs and GRUs to improve feature extraction. Accuracy is also improved by adding acceleration signals. This strategy may help hand-disabled people in everyday life, according to the paper.

Yang et al. [21] introduced an approach called MResLSTM, which is a multi-stream network that merges the residual model with variant ConvLSTM model to accurately identify and analyse dynamic hand movements. The network attains cutting-edge outcomes and enhances the precision of behavioural action detection. sEMG signals and accelerometer (ACC) signals are gathered to create datasets including various dynamic motions. The research finishes by emphasizing the benefits of deep learning in surpassing the constraints of feature engineering and enhancing the precision of EMG signal detection.

Ricardo. V. Godoy et al. [22] explored the use of EMG based interfaces for the purpose of controlling robotic and bionic systems and suggested using a technique using Temporal Multi-Channel Vision Transformers, a deep learning methodology, to extract intricate characteristics from unprocessed EMG data. The efficacy of this approach is evaluated by comparing its performance with that of other approaches, using the Ninapro dataset. It emphasized the capacity of EMG-based interfaces to attain skilful manipulation of robots and bionic hands.

Dianchun bai et al. [23] investigated a human-computer interaction gesture recognition using sEMG. A multichannel sEMG amplification unit-based feature model creation and optimization approach are employed. A multistate muscle action recognition feature model using CNN and LSTM is produced. The testing findings reveal that this 1 MB model can recognize 91.40% of complicated movements. The CNN+LSTM hybrid framework predicts complicated hand motions better than standard machine learning approaches.

Contemporary studies on the control of bionic hands and the connection between humans and computers demonstrate progress, yet obstacles remain. Research investigates the precision of understanding gestures, the dependability and verification of hand movement detection based on sEMG across various situations. The concerns revolve on the capacity of the suggested systems to handle large-scale operations and be user-friendly, as well as the adaptability of deep learning methods to

different hand movements and human characteristics. To advance the practical usefulness of bionic hand control systems in real-world applications, it is crucial to tackle these obstacles.

### III. PROPOSED SYSTEM

Initially, the pre-processing step called dataset augmentation is done by means of rotation, flipping, and scaling and the quality of the augmented image are enhanced using median filtering and Histogram equalization. ROI Identification is done by Attention guided U-Net with atrous convolution followed by that the spatial and temporal features are extracted using Bionic Net-1 and Bionic Net-2. The features are selected by using the novel hybrid algorithm called Optimized Red Fox Falcon Algorithm (O-RFF). Finally, the Bionic Fusion-Net Architecture combines Xception, Squeeze-Net, Shuffle-Net, optimized Bi-LSTM and Huber Loss functions employed for classification. Fig. 1 depicts the recommended methodology.

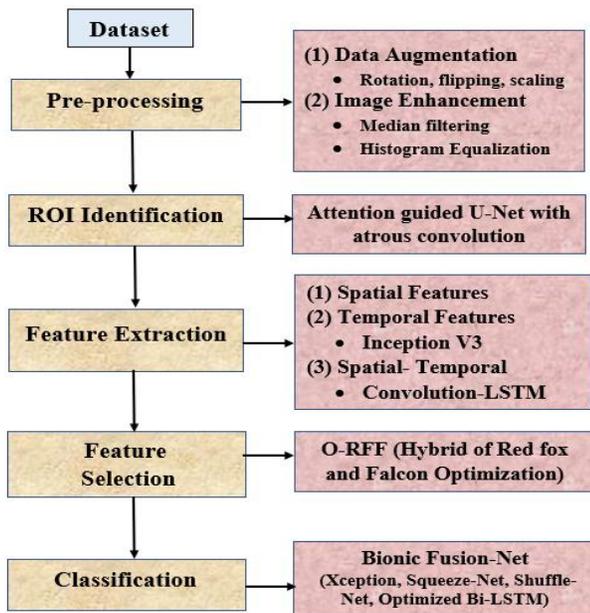


Fig. 1. Recommended methodology.

#### A. Pre-processing

It is the initial process involved in the model which is a crucial step in converting the unprocessed information into the best format. The data are pre-processed by Data augmentation and image Enhancement.

1) *Data augmentation*: Despite the quantity of data, finding correct information that matches for our research is difficult. To improve model performance, data must be diverse in measurements, positions, shades, and illumination. To address the limitation of a finite dataset, we employ data augmentation techniques. The data are augmented by the following parameters.

a) *Rotation*: Image rotation may be done at 90-degree angles or more angles, depending on needs. No background noise is added when an image is rotated 90 degrees. In contrast, rotation at minute angles may cause noise. For black or white backgrounds, the noise may mix, but for various colours, the

network treats the noise as a distinguishing feature during learning.

b) *Scaling*: For the majority of time series data, if we introduce a little alteration to the data in each individual sample, it is probable that the sample will still be assigned the same label as before. Scaling refers to the process of altering the size of the source data by multiplying the sample data by a random scalar. Contrary to the AddNoise approach, Scaling involves the addition of uniform noise to all samples simultaneously, whereas AddNoise introduces distinct noise to each sample. The source data is typically scaled using a normal distribution function with a mean ( $\mu$ ) of 1 and a variance ( $\sigma^2$ ) of 0.1.

c) *Flipping*: The images may be flipped either horizontally (HF) or vertically (VF). It generates pictures by rotating the image in increments of 90° degrees. Some frameworks do not have the capability to do VF. To achieve VF, one may rotate the picture by 180° and then apply a HF.

2) *Image Enhancement (IE)*: Enhancement refers to the manipulation of a data to improve its visual quality or adapt it to a certain objective. The following two approaches are used for the purpose of IE which are detailed below.

a) *Median filtering (Me-F)*: Noise cancellation is often achieved using a nonlinear technique called Me-F. The information from both neighbourhoods is selected for sorting based on the window length selected, and each data point to be processed is filtered before the value of the middle size is determined. Enhancing the outcomes of further analysis—such as edge identification on an image, for example—is a routine procedure. Compared to linear filters, it offers a significant advantage in that it completely eliminates the impact of extremely high magnitude input noise levels. They are used as smoothers to eliminate noise from salt and pepper. The Me-F's output  $y$  is

$$z(t) = \text{median} \left( y(t - T/2), \dots, y(t), \dots, y(t + T/2) \right) \quad (1)$$

where  $t$  refers Window size of Me-F

b) *Histogram equalization*: Image contrast is determined by dynamic range, the ratio of brightest to darkest pixel intensities. Enhancing low-contrast images with contrast enhancement techniques has several uses. Histogram equalization (HE) is a popular approach. The probability distribution of input grey levels is used to map grey levels. Considering image  $I$ , the probability density function  $P(I_x)$  is given as,

$$P(I_x) = \frac{n_x}{n} \quad (2)$$

Where  $x = 0, 1, \dots, L - 1$ ,  $n_x$  is the number of times the level  $I_x$  in input,  $n$  represents the overall samples.

The Cumulative Density Function (CDF) is defined as,

$$c(i) = \sum_{j=0}^x P(I_j) \quad (3)$$

Where  $I_x = i$  for  $x = 0, 1, \dots, L - 1$ ,  $c(I_{L-1}) = 1$  (constant). The transform function  $f(i)$  based on CDF is stated as,

$$f(i) = I_0 + (I_{L-1} - I_0)c(i) \quad (4)$$

Output of HE,  $S = \{S(q, r)\}$  is given as,

$$S = f(I) = \{f(I(k, j)) | \forall I(k, j) \in I\} \quad (5)$$

The high performance of the HE in enhancing the contrast of an image because of the dynamic range expansion.

### B. ROI Identification

The quality of the data has a significant impact on how well the model's function. A precise localization of the ROI is vital to avoid biased learning and subpar model performance caused by irrelevant features. Attention guided U-Net with atrous convolution is employed for the ROI Identification.

The U-Net design uses the encoder-decoder structure including compression and expansion path. It consists of three primary components: the left (down-sampling) segment, the center (copy and crop) segment, and the right (up-sampling) segment. During the first stage, the left portion performs four down-sampling procedures to decrease picture size, while extracting characteristics from superficial data. The primary component is performing four concatenation procedures, smoothly merging profound and superficial information to augment feature representation. During the up-sampling step, four processes are performed to retrieve detailed information from the expanded picture. Significantly, the up-sampling procedure entails reducing the number of channels in the picture by half, in contrast to the feature extraction in the left section. During the process of up-sampling, the shallow information on the left is combined and the features are joined together. By include skip connections within the same stage, it guarantees that the reconstructed feature map encompasses a wide range of low-level characteristics and features of different scales. This method enables the prediction of many scales and incorporates deep supervision to enhance the accuracy of segmentation maps by improving details such as edge recovery. The incorporation of a Hybrid Atrous Convolution block, a modified version of a residual block, and a Redesigned skip connection strengthens the U-Net architecture, leading to enhanced performance without any loss of information. Fig. 2 shows the attention guided U-Net with atrous convolution.

1) *Hybrid atrous convolution block*: The Hybrid Atrous Convolution block utilizes the well-recognized attribute of atrous convolution to greatly increase the receptive field. By using a 3×3 convolution kernel, the combination of 1-dilated and 2-dilated convolutions replicates the effects of a 7×7 convolution kernel. By combining 4-dilated convolution with 1-dilated and 2-dilated convolutions, the resulting receptive field is comparable to that of a 15×15 convolution kernel. This method results in a rapid and significant increase in the receptive field as compared to conventional convolution processes. However, using just stacked convolutions with the same dilation rate creates a discontinuity in the kernel, where not all pixels are included in the computations. The fragmented technique leads to a checkerboard pattern, which undermines continuity and becomes less efficient for little things. The Hybrid Atrous Convolution block is offered as a solution to this problem. It combines several dilation rates to improve the

continuity of information and boost performance, particularly when dealing with tiny objects.

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \quad (6)$$

Where  $M_i$  maximum dilation rate at layer  $i$ ,  $r_i$  is the dilation rate of the  $i^{th}$  layer.

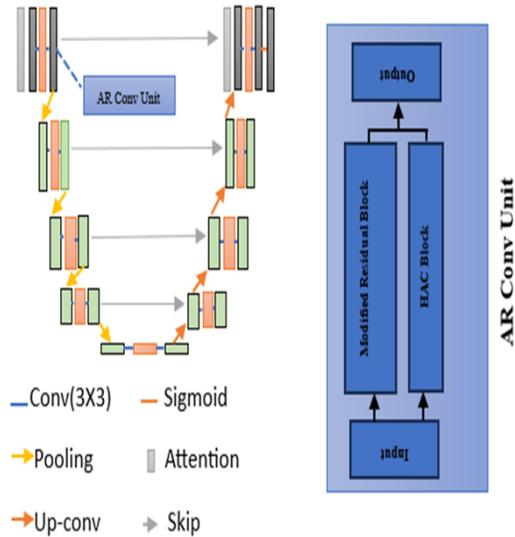


Fig. 2. Attention guided U-Net with atrous convolution.

2) *Modified residual block*: To improve the performance of the network, a squeeze-and-excitation block has been added to ResNet, which demonstrates variations from the original residual network as seen in Fig. 2. The Squeeze operation, shown by the red box in Fig. 2, is crucial in modifying the spatial dimensions of every input feature map from  $H \times W$  to a condensed  $1 \times 1$  format. This conversion is accomplished by means of global average pooling. During the Squeeze process, there is a compression of spatial dimensions, which results in the conversion of each two-dimensional feature channel into a single real number that encompasses the whole field of perception. The output dimension corresponds to the number of feature channels in the input. This technique captures the worldwide dispersion of responses on the characteristic channel, giving the layer near the input a thorough global receptive field. This adjustment is very helpful in a wide range of jobs, enhancing the capabilities of the network.

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (7)$$

$z_c$  represents the channel descriptor for channel  $c$ ,  $F_{sq}$  indicates the represents global average pooling,  $u_c$  represents channel  $c$  of the input,  $H, W$  represent the height and width of the input

In the excitation phase (highlighted in the green box in Fig. 2, the feature dimension undergoes reduction to  $1/r$  of the input. Subsequently, it is elevated back to the original dimension through a fully connected layer following ReLU activation. This approach, characterized by increased nonlinearity, proves

superior in capturing complex correlations between channels compared to the direct utilization of a fully connected layer. Notably, this method significantly diminishes the number of parameters and computational load, offering a more efficient and streamlined computational process.

3) *Attention module*: The U-Net includes an attention module to estimate the most likely area. This uses a circular bounding box to approximate area coordinates. The feature layer section of the network has an attention module underneath the fully connected layer. The circular bounding box coordinates are acquired by this attention module. Left vertex (x1, y1), right vertex (x2, y2) defines the circular area. Therefore, the attention module's regression component creates an array (x1, y1, x2, y2). The regression results are used to create the loss function, optimizing and training the network. This integration helps the U-Net concentrate and improve its prediction for the area of interest, enhancing task performance.

4) *Redesigned skip connection*: To use the unique attributes of each level inside the network, a modification to the skip connection has been executed. In contrast to the traditional dense connections that establish connections between every node in the encoder and decoder via intermediary connections, this novel strategy specifically retains skip connections alone between the decoder and each node. The original design's extensive connections led to a large number of model parameters and increased computing complexity. This skip connection, which has been simplified and revamped, successfully decreases the parameter count while retaining all relevant information. This ultimately leads in enhanced segmentation outcomes. This achieves a compromise between computational efficiency and the preservation of crucial information, resulting in a segmentation conclusion that is both more effective and efficient.

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), j = 0 \\ H([x^{i,j-1}, u(x^{i+1,j-1})]), j < 4 - i \\ H([\underbrace{x^{i,k}}_{k=0}^{j-1}, u(x^{i+1,j-1})]), j = 4 - i \end{cases} \quad (8)$$

### C. Feature Extraction

It transforms the preprocessed data into features which is used to categorize the data. The multi-features like spatial and temporal features are extracted using the various approached which are provided below.

1) *Spatial features*: They include the characteristics of data that relate to the spatial organization or configuration of objects or components within a certain area. These traits are crucial for activities that require comprehension of the geometric connections, locations, and structures of things. Dynamic Filters are used to extract the characteristics. Dynamic filters in CNNs are adjustable or modifiable dependent on the input data, and are used to identify patterns and characteristics. They modify their weights and configurations in order to effectively record changing spatial attributes in real-time. Joint positions

refer to the precise locations or coordinates of the joints within a given system. Angles are the quantitative representation of the inclination or relative orientation of lines or surfaces. This encompasses the comparative distances, dimensions, configurations, and characteristics as well.

2) *Temporal features*: The dependencies delineate the exact timing relations between several components of the motion, governing the seamless progression and precision of the activity. In this Inception V3 is employed for extracting the temporal features.

The selection of InceptionV3, one of the Google Inception Nets, was based on its high accuracy and adaptability. The module's design includes two convolutional layers with a patch size of 3x3 and two unique strides. In order to enhance processing efficiency after the convolutional layers, the Inception model underwent modification by dividing the 5 × 5 layers into several 3 × 3 layers and ReLu levels. The max-pooling methodology is used to extract critical features, namely edges. The average pooling method is used to progressively extract features. Classification is then performed using the SoftMax function in the fully connected layer. The activation function is defined as,

$$j(k) = \frac{1}{1+e^{-k}} \quad (9)$$

The ReLu activation function may be defined as follows:

$$j(k) = \begin{cases} 0; k \leq 0 \\ 1; k > 0 \end{cases} \quad (10)$$

Including one or many pooling layers in the feature maps generated by convolutional layers helps to mitigate computational challenges. To do this, the size of the CL's maps was reduced. The two dominant techniques are average pooling and maximal pooling.

3) *Integration of spatial and temporal features*: A model called Convolutional-LSTM is employed to integrate the spatial and temporal features. Convolutional Neural Networks (CNNs) include many filter stages and one classification step. In sequence learning, Recurrent Neural Networks (RNNs) are popular. Their efficiency may be limited by the vanishing gradient issue during training backpropagation. Long Short-Term Memory (LSTM) architecture improves data feature long-term dependency. This patch mitigates the vanishing gradient problem, improving the RNN's sequential data handling and long-term pattern learning and retention. Fig. 3 shows the Conv-LSTM.

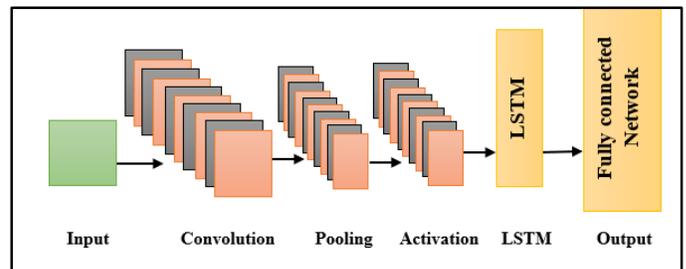


Fig. 3. Conv-LSTM.

a) *Input layer*: The first layer in CNN is the 'input layer'. Its purpose is to receive data and resize them before forwarding them to subsequent layers for the extraction of features.

b) *Convolution layer (conv)*: This function as image filters. These layers are responsible for identifying features inside images and are also used for determining the matching feature points during testing.

c) *Pooling Layer (PL)*: The feature sets obtained are then sent to the PL. This layer resizes huge photos by reducing their dimensions while retaining the crucial details. It retains the highest value from each window and maintains the most accurate matches of each characteristic inside the window.

d) *Activation layer*: The inclusion of an activation function layer is crucial in every convolutional block. The network enables a non-linear expression of the input signal, hence improving the representation and increasing the distinguishability of the learnt characteristics. Several activation functions may be used, including ReLU, Identity, Tanh, and Sigmoid, which are often employed.

e) *LSTM layer*: The purpose of the LSTM is to address the issue of long-term reliance by including the forget gate, which controls the use of information in the cell state.

f) *Fully connected Layer (FCL)*: The final is the FCL which receives the high-level filtered images and converts them into categories with corresponding labels.

#### D. Feature Selection

The process of choosing a subset of pertinent features from the initial set of features is known as feature selection. Simplifying the model, lowering the complexity of the feature space, and enhancing generalisation performance are the objectives. Utilising the Optimized Red Fox Falcon Algorithm (O-RFF) which is a hybrid of Red Fox and Falcon Optimization Algorithms the features are chosen.

The Red Fox population consists of individuals displaying various habits, with some occupying well defined territories and others adopting a wandering existence. Every territorial group is organized with an alpha pair at the top of the hierarchy. Once the young foxes reach adulthood, they have the option to leave the group and create their own territories, especially if there are favorable chances to win authority over new regions. Alternatively, kids have the option to remain within the family and ultimately inherit the hunting zone from their parents. Red foxes, skilled at capturing tiny animals, whether they are domestic or wild, take advantage of every opportunity to find food as they go across their region. Their hunting style is surreptitiously approaching their victim and slowly reducing the distance before executing a successful assault. The algorithm employs a global search approach by exploring territory in search of food. Next, the second stage entails moving through the environment to approach the prey as closely as feasible before commencing the assault, which is represented as a local search.

Falcons have unique and complex hunting behaviors, using both clear, clearly identifiable techniques and more elaborate, convoluted tactics while chasing and capturing their prey.

During the initial phase, the FOA factors and the governing limits are initialized. Following this, the motion and location of falcons are altered according to the supplied values which is represented as,

$$y = \begin{bmatrix} y_{1,1} & \cdots & y_{1,V} \\ \vdots & \vdots & \vdots \\ y_{A,1} & \cdots & y_{A,D} \end{bmatrix} \quad (11)$$

Where  $y$  is the falcon location, regarding the total applicants  $A$  every dimension  $V$ . The speed is generated randomly within the  $v_{Max}$  and  $v_{Min}$  limits.

$$v_{Max} = 0.1Up_{li} \quad (12)$$

$$v_{Min} = -v_{Max} \quad (13)$$

Where  $Up_{li}$  is the upper limit in every measurement.

Determine the fitness value at each iteration. During that period, the optimal individual is designated as  $I_b$  and the optimal position for each falcon is denoted as  $y_b$ .

Create two random elements ( $Q_B, Q_C$ ) with a normal distribution on every bird of attack to study the association between awareness and leap probability. The main probability considered where  $Q_B$  is smaller than the falcon is given as,

$$y_t = y_{t-1} + v_{t-1} + M_t r (y_{b,t-1} - y_{t-1}) + O_t r (I_{b,t-1} - y_{t-1}) \quad (14)$$

$y_{t-1}$ ,  $v_{t-1}$  are present location and Falcon's motion.  $M_t r$ ,  $O_t r$  are cognitive rate and social

If  $Q_B$  exceeds  $B$  (Adaptive prob) the jump is compared to  $Q_C$ . If  $Q_C$  exceeds  $C$  (Dive Prob), the falcon selects one prey ( $y_{ch}$ ) and performs its hunting evolution using logarithmic twisting expressed as,

$$y_n = y_{t-1} + |(y_{ch} - y_{t-1})| \exp(ze) \cos(2\pi e) \quad (15)$$

$y_n$  is the new position,  $z$  is accurate observation logarithmic twisting that equates to 1,  $e$  is an irregular value in range  $[-1, 1]$  indicates how close the falcon is to its actual target.

When  $Q_B$  is lower than  $Q_C$ , the fitness of the picked prey is compared to the falcon's fitness and this condition is expressed as,

$$y_n = y_{t-1} + v_{t-1} + f_c r (y_{ch} - y_{t-1}) \quad (16)$$

The constant  $f_c r$  describes the falcon's ability to measure its precise position when uncertain. The vision radius from RFO is defined as,

$$r = \begin{cases} a \frac{\sin(\theta_0)}{\theta_0} & ; \theta_0 \neq 0 \\ \theta & ; \theta_0 = 0 \end{cases} \quad (17)$$

$\theta$  is a random value in range  $[0,1]$ ,  $\theta_0 \in [0, 2\pi]$  is chosen for all individuals to simulate falcon observation angle,  $a$  is the scaling parameter set once each iteration for all individuals in the population to simulate variable distances from prey during falcon approaches.

Using the O-RFF, the best features are selected based on which the classification is performed.

### E. Classification

This is the final step in the process which classifies the final output based on the results from the previous stages. Bionic Fusion-Net Architecture is used for the classification process which is the combination of Xception, Squeeze-Net, Shuffle-Net and optimized Bi-LSTM which results in accurate classification. Fig. 4 depicts the Bionic Fusion-Net Architecture.

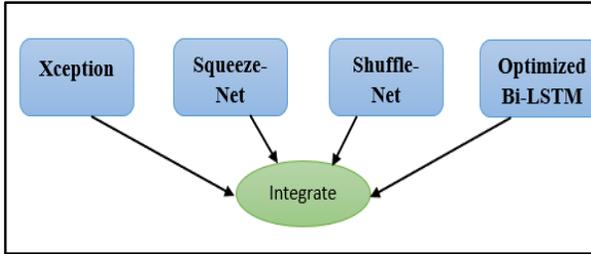


Fig. 4. Bionic fusion-net.

1) *Xception*: The Xception architecture consists of a total of 36 convolutional layers, which make up the fundamental feature extraction part of the network. Our experimental inquiry only focuses on picture categorization. Thus, the convolutional base is immediately succeeded by a logistic regression layer. Additionally, there is the possibility of including fully-connected layers before to the logistic regression layer. The 36 convolutional layers are arranged in 14 modules, with linear residual connections encompassing them. It is worth mentioning that the initial and final modules lack these linear residual connections. The Xception architecture may be described as a sequential arrangement of depth wise separable convolution layers with residual connections. This specific design decision makes the architecture easily definable and adaptable, providing convenience for modifying and experimenting.

2) *Squeeze-Net*: With two convolution layers (CL), eight fire layers, three max-pooling levels, one global average pooling layer, one SoftMax output layer, and so on, Squeeze Net is a convolutional network. Squeeze Net has 8 Fire modules (fire2-9) after the 1st CL (conv1). Finally, there is a final convolution layer. From the start of the network to its end, each Fire module has more filters than before. Max-pooling is carried out by Squeeze Net using a two-stride method following layers conv10, fire8, fire4, and conv1.

The RGB channels of the network's input (i/p) measure 227 × 227 pixels. Max pooling is used to further specialize the i/p pictures after convolution. Using 3 × 3 kernels, the convolution layer connects the limited areas and weights in the i/p volumes. Each component is independently activated as the real component by each convolution layer. Each convolution layer performs activation. It utilizes fire layers which squeeze phase (S-p) and Expanded phase (E-p) between the convolution layers. The i/p and output (o/p) tensor scales of the fire are equal. 1 × 1 filters are used in S-p, whereas along with the above filter, 3 × 3 are used in the E-p. The o/p is given as,

$$f(y) = \sum_{fm1=1}^{FM} \sum_{c=1}^c W_c^f x_c^{fm1} \quad (18)$$

Where FM, C are feature maps (fm's), channels. The weighted total of the feature maps of each individual tensor is the squeeze results. Max pool down samples along spatial dimensions, whereas the global average pool aggregates the class feature mappings into a single value. Multiclass probability distributions are returned by the SoftMax activation method at the output end.

The improved sliding windowing technique is employed for feature extraction, in contrast to considered whole signal at once, because of its stochastic nature. The procedure is used for segmentation, either adjacent or overlapped [24-25]. The outcomes demonstrate the accuracy of categorization is higher for the overlain windowing strategy compared to the adjacent or disjunct windowing scheme. Data stability during feature extraction is ensured by segmenting the data into brief windows. Each temporal series was divided into ideal segments or sub-frames using the overlapping windowing technique in this investigation.

3) *Shuffle-Net (Sh-Nt)*: A very effective DL architecture called Shuffle-Net was developed using mobile devices. It contains 50 learnable layers, which are comprised of an FC layer, 48 group convolution layers, and 1 convolution layer. To reduce the total computing complexity, pooling layers are used. The output of the convolution layer of the network is given as,

$$s(i, j) = \sum_n \sum_m I(m, n) K(i - m, j - n) \quad (19)$$

K, i are kernel and i/p image. The o/p of size e=((i-k)+2p)/(s+1) is produced after convolution, where p means padding, and s represent steps.

The o/p fm's of the 1st CL is sent to the Sh-Nt unit. Three convolutional procedures make up the Shuffle Net unit: three × three depth wise convolutions and two 1 × 1 pointwise group convolutions. Channel shuffle operation, ReLu activation function, and BN come after the initial pointwise group convolution. ReLu activation is employed due of its simplicity and effectiveness. It is stated as

$$f(y) = \begin{cases} 0; & y < 0 \\ 1; & y \leq 0 \end{cases} \quad (20)$$

ReLu sets neurons to 0 (deactivates neurons) and activates neurons with positive values. BN comes after the second and third convolution procedures, which are the 3 × 3 depth- wise convolutions and the 1 × 1 pointwise group convolution. A 3-by-3 average pooling on the shortcut paths is included in the model. The model is made up of sixteen sequential Shuffle-Net components. There are fifty layers in the model, and each one offers trainable feature maps. Additionally, these layers extract features. After submitting these feature maps to FC, the final classification layer uses Soft-max activation for classification probability determination.

$$a_i = \sum_{j=0}^{m \times n - 1} w_{ij} \times x_i + b_i \quad (21)$$

where i means index of the FC layer's output; n, m, d, and i denote the height, width, depth, and index of FC layers output.).

4) *Optimized Bi-LSTM*: The forget gate, i/p gate, and o/p gate are the three gated units that make up the majority of the LSTM neural network. Bi- LSTM can processes sequences in

bi-directions. Two LSTM networks are combined, one of which processes the sequence from left to right and the other from right to left. The output of a Bi-LSTM is typically a concatenation of the hidden states from both the left-to-right and right-to-left LSTM networks which provides higher efficiency for the model. This enables it to record each element in the sequence's context from both the past and the future. In order to maintain the interdependence of time series information over long distances and achieve accurate prediction, the sequence data is acquired and stored by specialized gated units. The input data is largely processed by the input gate. The current neuron's capacity to remember prior knowledge is contingent upon the forget gate. The o/p of the neuron is given in the output gate. Assuming that the i/p sequence is  $x_1, x_2, \dots, x_t$ , the calculation formula for each LSTM neuron parameter at time  $t$  is as follows:

$$i_t = S(W_i * [h_{t-1}, x_t]) \quad (22)$$

$$f_t = S(W_f * [h_{t-1}, x_t]) \quad (23)$$

$$o_t = S(W_o * [c_t, h_{t-1}, x_t]) \quad (24)$$

$$c_t = f_t * C_{t-1} + i_t * \tanh(W_c * [h_{t-1}, x_t]) \quad (25)$$

$$h_t = o_t * \tanh(c_t) \quad (26)$$

The LSTM neural network's weights between nodes are optimized using a predetermined procedure, which can make the weights between neurons more rational and enhance the model's capacity for generalization and prediction. The weights are optimized by using the novel O-RFF Optimization Algorithm.

5) *Integration*: All the outputs from the above models are integrated to form a Bionic Fusion-Net model which results in higher classification accuracy than the other models.

#### F. Huber Loss Function

The Huber loss function is used for regression, especially with outliers. It balances mean squared error and mean absolute error, making it resilient to extreme dataset values. This makes it beneficial when accuracy and robustness must be balanced. Its smooth gradient transition provides robust optimization during training, especially in neural network applications. The Huber loss is adaptable for regression problems with different data properties because the absolute difference between predicted and actual values is a more relevant error statistic than the squared difference.

It is defined as,.

$$L_\delta(x) = \begin{cases} \frac{1}{2}x^2 & ; |a| \leq \delta \\ \delta \left[ |a| - \frac{1}{2}\delta \right] & ; else \end{cases} \quad (27)$$

#### IV. RESULT AND DISCUSSION

Using the dataset taken, the results are evaluated using the performance metrics. The results are computed in comparisons of the suggested and the existing methods Xception [24], Squeeze-Net [25], Shuffle-Net and Bi-LSTM which is implemented in the python platform.

#### A. Dataset Description

The evaluation of the suggested and the current techniques are done using the performance metrics which are explained in detail below and Table I and II provides the values of the Performance metrics for the recommended and methods in use at the rate of 70/30 and 80/20 respectively.

TABLE I. VALUES OF PERFORMANCE METRICS IN THE 70/30 RATE

Model	Accuracy	Sen	Spe	Pre	F-measure	FPR	FNR	MCC
Bionic Fusion-Net	0.986	0.987	0.988	0.987	0.984	0.043	0.039	0.988
Bi-LSTM	0.952	0.963	0.966	0.958	0.967	0.073	0.089	0.965
Squeeze-Net	0.930	0.950	0.94	0.933	0.943	0.063	0.079	0.952
Shuffle-Net	0.930	0.951	0.94	0.934	0.945	0.073	0.089	0.953
Xception	0.909	0.949	0.947	0.921	0.930	0.072	0.062	0.953

TABLE II. VALUES OF PERFORMANCE METRICS IN THE 80/20 RATE

Model	Accuracy	Sen	Spe	Pre	F-measure	FPR	FNR	MCC
Bionic Fusion-Net	0.986	0.987	0.989	0.991	0.989	0.035	0.0362	0.988
Bi-LSTM	0.990	0.967	0.968	0.963	0.970	0.063	0.079	0.967
Squeeze-Net	0.958	0.952	0.943	0.937	0.949	0.053	0.069	0.953
Shuffle-Net	0.95	0.954	0.952	0.941	0.953	0.063	0.079	0.953
Xception	0.943	0.950	0.948	0.923	0.934	0.061	0.061	0.954

1) *Accuracy*: It is the proportion of true forecasts to all i/p Observations. It is determined using (28)

$$Accuracy = \frac{\text{Number of correct predictions}}{\text{Total number of samples}} \quad (28)$$

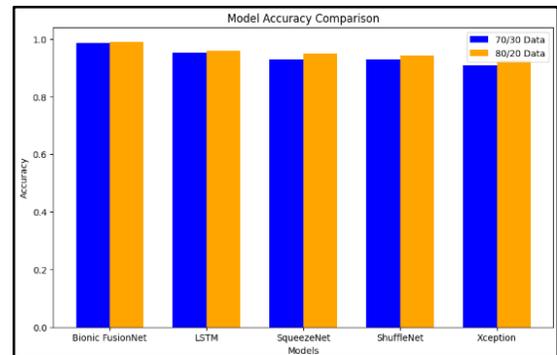


Fig. 5. Examination of suggested and existence approaches in terms of accuracy.

Fig. 5 shows the examination of the recommended and methods in use about Accuracy. From the graphical representation, it is seen that the proposed approach has a higher Accuracy.

2) *Sensitivity*: The fraction of real positives that are correctly identified is measured by sensitivity. It is given as,

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (29)$$

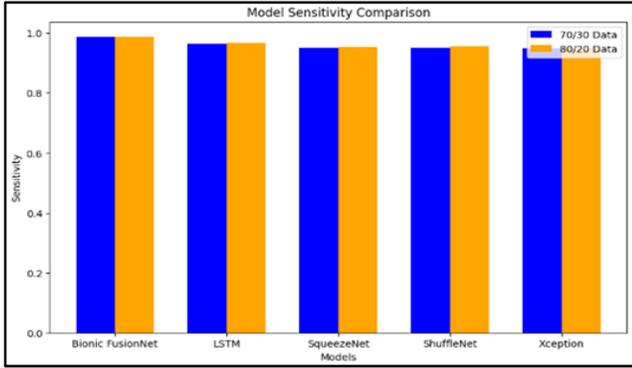


Fig. 6. Examination of suggested and existence approaches in terms of Sensitivity.

Fig. 6 shows the examination of the recommended and methods in use in regard to Sensitivity. From the graphical representation, it is seen that the proposed approach has a higher Sensitivity.

3) *Specificity*: The percentage of real negatives that are accurately identified is measured by specificity. It is calculated using

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (30)$$

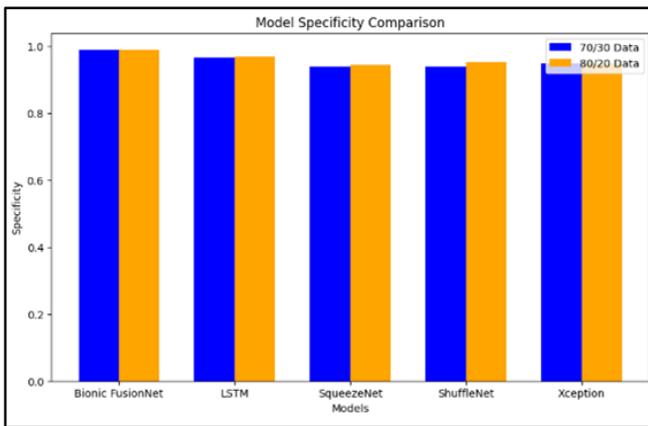


Fig. 7. Examination of suggested and existence approaches in terms of Specificity.

Fig. 7 shows the examination of the recommended and methods in use in regard to Specificity. From the graphical representation, it is seen that the proposed approach has a higher Specificity.

4) *Precision*: How much of a model's positive predictions are actually right is determined by its precision, which is a

performance indicator. In order to assess how well what you detect is actually present, precision is important. It is given as,

$$\text{Precision} = \frac{TP}{TP+FP} \quad (31)$$

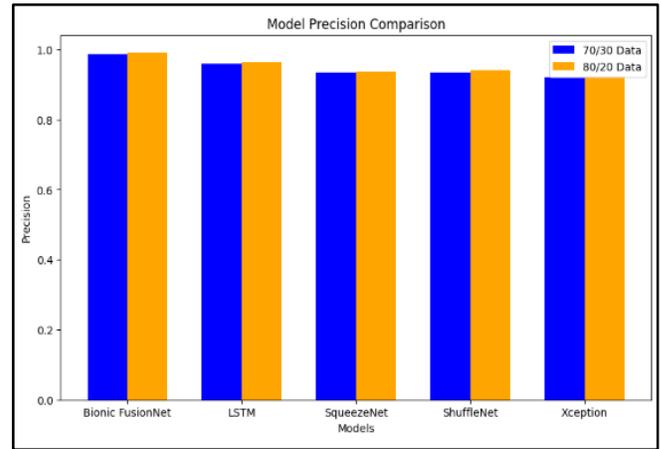


Fig. 8. Examination of suggested and existence approaches in terms of Precision.

Fig. 8 shows the examination of the recommended and methods in use in regard to Precision. From the graphical representation, it is seen that the proposed approach has a higher Precision

5) *Precision*: A general score for performance evaluation, it is a combination statistic that combines Precision and recall. It is given as,

$$\text{F measure} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (32)$$

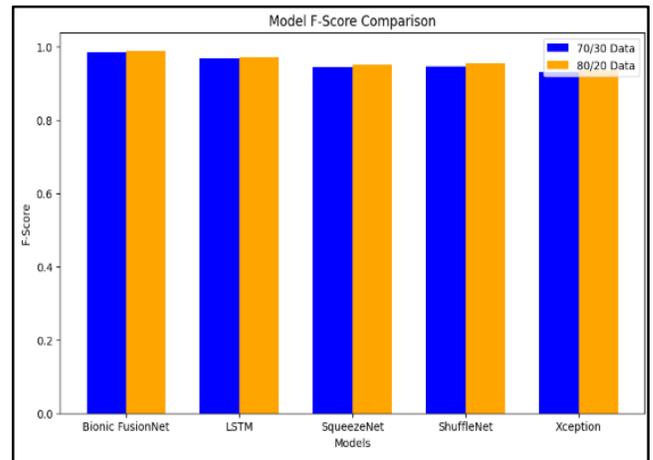


Fig. 9. Examination of suggested and existence approaches in terms of F-measure.

Fig. 9 shows the examination of the recommended and methods in use in regard to F-measure. From the graphical representation, it is seen that the proposed approach has a higher F-measure.

6) *FPR*: FPR refers to the values that are actually negative but predicted to be positive. It is calculated using the formula,

$$FPR = \frac{FP}{FP+TN} \quad (33)$$

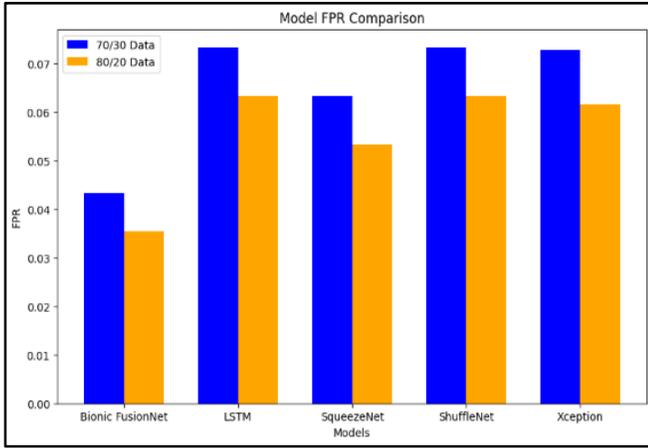


Fig. 10. Examination of suggested and existence approaches in terms of FPR.

Fig. 10 shows the examination of the recommended and methods in use in regard to FPR. From the graphical representation, it is seen that the proposed approach has a lower FPR.

7) *FNR*: FNR refers to the values that are actually positive but predicted to negative. It is calculated using the formula,

$$FNR = \frac{FN}{FN+TP} \quad (34)$$

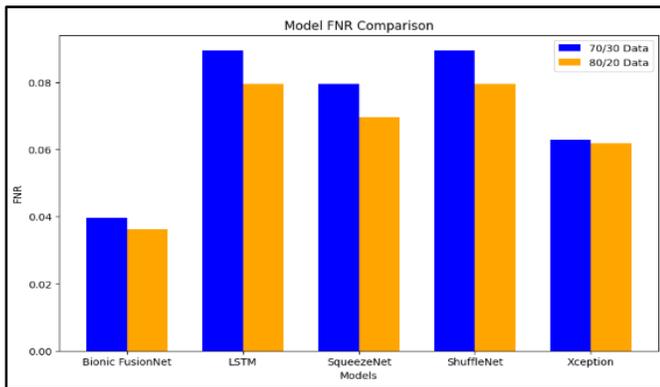


Fig. 11. Examination of suggested and existence approaches in terms of FNR.

Fig. 11 shows the examination of the recommended and methods in use in regard to FNR. From the graphical representation, it is seen that the proposed approach has a lower FNR.

8) *Mathew's Correlation Coefficient (MCC)*: MCC measures the degree of correlation between expected and actual values. It's stated as,

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (35)$$

Fig. 12 shows the examination of the recommended and methods in use in regard to MCC. From the graphical representation, it is seen that the proposed approach has a lower MCC.

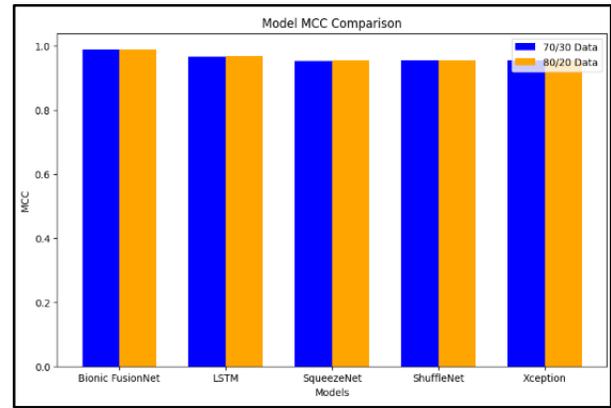


Fig. 12. Examination of suggested and existence approaches in terms of MCC.

## V. CONCLUSION

To create advanced prosthetic devices that recognize and respond to human gestures, bionic hand movement detection is essential. Recent developments in image processing, feature extraction, and deep learning have enhanced bionic hand movement detection accuracy and versatility. This paper presents a unified framework employing attention-guided ROI identification and a novel Bionic Fusion-Net architecture to solve these issues. Initial pre-processing includes dataset and picture improvement. ROI Identification employs an attention-guided U-Net with advanced convolutional components. During feature extraction, BionicNet-1 and BionicNet-2 learn spatial and temporal features jointly. O-RFF, a hybrid Red Fox-Falcon Optimization Algorithm, optimizes feature selection. Xception, Squeeze-Net, Shuffle-Net, optimized Bi-LSTM, and Huber Loss function application form the Bionic Fusion-Net Architecture. The proposed attained a performance of 99% accuracy, 98.7% Sensitivity, 98.9% Specificity, 99.1% Precision, 98.9% F-score, 3.5% FPR, 3.6% FNR and 98.8% MCC. Thus from the results, it is seen that the recommended technique results better in contrast to the methods in use.

## REFERENCES

- [1] Mohammadreza Aghaei, Hossein Ebadi, Aline Kirsten Vidal de Oliveira, Shima Vaezi, Aref Eskandari, Juan M. Castañón (2020), Chapter 11 - New concepts and applications of solar PV systems, Photovoltaic Solar Energy Conversion, Academic Press, 349-390.
- [2] Gurung, A. et al. (2017). Highly efficient perovskite solar cell photocharging of lithium ion battery using DC-DC booster. *Advanced Energy Materials*, 7(11), 1602105.
- [3] Rathore, N.; Panwar, N.L.; Yettou, F.; and Gama, A. (2021). A comprehensive review of different types of solar photovoltaic cells and their applications. *International Journal of Ambient Energy*, 42(10), 1200-1217.
- [4] Mohamed, N.; Aymen, F.; Altamimi, A.; Khan, Z.A.; and Lassaad, S. (2022). Power management and control of a hybrid electric vehicle based on photovoltaic, fuel cells, and battery energy sources. *Sustainability*, 14(5), 2551.
- [5] Safayatullah, M.; Elrais, M.T.; Ghosh, S.; Rezaii, R.; and Batarseh, I. (2022). A comprehensive review of power converter topologies and control methods for electric vehicle fast charging applications. *IEEE Access*, 10, 40753-40793.
- [6] Kumar, P.V.; Suresh, A.; and Rashmi, M.R. (2016). Optimal Design of Fused Chopper based Standalone Hybrid Wind Solar System. *Indian Journal of Science and Technology*, 9(21), 1-6.

- [7] Kumar, P.V.; Athithya, R.S.; Valli, R.I.; Abinaya, S.; and Hema, B. (2023). Battery Management for Electric Vehicle Using Low Voltage DC-DC Converter. Proceedings of the 4th International Conference on Signal Processing and Communication (ICSPC), 62-67
- [8] Kumar, P.V.; Kumar, A.R.; and Tiwari, R. (2021). Performance Analysis of Solar Connected Fly-Back Boost C-onverter for Electric Vehicle applications. Proceedings of the 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, 1638-1643.
- [9] Nandankar, P.V.; Bedekar, P.P.; and Dhawas, P.V. (2021). Efficient DC-DC converter with optimized switching control: A comprehensive review. Sustainable Energy Technologies and Assessments, 48,101670.
- [10] Aljafari, B.; Ramu, S.K.; Devarajan, G.; Vairavasundaram, (2022). Integration of Photovoltaic-Based Transformerless High Step-Up Dual-Output–Dual-Input Converter with Low Power Losses for Energy Storage Applications. Energies, 15(15), 5559.
- [11] Belqasem Aljafar.; Gunapriya Devarajan.; Selvi Arumagam.;Indhiragandhi Vairavasundram (2022). Design and Implementation of Hybrid PV/Battery – Based Improved Single Ended Primary \_Inductor Converter-Fed Hybrid Electric Vehicle. International Transactions on Electrical Energy System, 2022,1-11.
- [12] Mondzik, A.; Stala, R.; Piróg, S.; Penczek, A.; Gucwa, P.; and Szarek, M. (2021). High efficiency DC–DC boost converter with passive snubber and reduced switching losses. IEEE Transactions on Industrial Electronics, 69(3), 2500-2510.
- [13] Prabhu, N.; Thirumalaivasan, R.; and Ashok, B. (2023). Critical Review on Torque Ripple Sources and Mitigation Control Strategies of BLDC Motors in Electric Vehicle Applications. IEEE Access.
- [14] Lee, J.; Lim, G.C.; and Ha, J.I. (2022). Pulse width modulation methods for minimizing commutation torque ripples in low inductance brushless DC motor drives. IEEE Transactions on Industrial Electronics, 70(5), 4537-4547.
- [15] K. Xia.; Y. Ye, J. Ni.;Y. Wang and P. Xu (2020) .Model Predictive Control Method of Torque Ripple Reduction for BLDC Motor. IEEE Transactions on Magnetics, 56(1), 1-6.
- [16] Hussain, M.T.; Sulaiman, N.B.; Hussain, M.S.; and Jabir, M. (2021). Optimal Management strategies to solve issues of grid having Electric Vehicles (EV): A review. Journal of Energy Storage, 33,102114.
- [17] Thangavel, S. et al. (2023). A Comprehensive Review on Electric Vehicle: Battery Management System, Charging Station, Traction Motors. IEEE Access, 11, 20994-21019.
- [18] Aishwarya, M.; and Brisilla, R.M. (2023). Design and Fault Diagnosis of Induction Motor Using ML-Based Algorithms for EV Application. IEEE Access, 11, 34186-34197.
- [19] Vinoth Kumar, P. et al. (2023). Battery Management for Electric Vehicle Using Low Voltage DC-DC Converter. In Proceedings of the 2023 4th International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 2023, pp. 62-67.
- [20] Chatterjee, D.; Biswas, P.K.; Sain, C.; Roy, A.; and Ahmad F. (2023). Efficient Energy Management Strategy for Fuel Cell Hybrid Electric Vehicles Using Classifier Fusion Technique. IEEE Access, 11, 97135-97146.
- [21] Nelson, L.M.; Nihat, I.; and Murat, L. (2023). Control and performance analyses of a DC motor using optimized PIDs and fuzzy logic controller. Results in Control and Optimization,13, 10036.
- [22] Alsharekh, M.F. et al. (2022). Improving the efficiency of multistep short-term electricity load forecasting via R-CNN with ML-LSTM. Sensors, 22(18), 6913.
- [23] Lipu, M.S.H. et al. (2024). Artificial Intelligence Approaches for Advanced Battery Management System in Electric Vehicle Applications: A Statistical Analysis towards Future Research Opportunities. Vehicles, 6, 22-70.
- [24] Akram, M.N.; and Abdul-Kader, W. (2023). Sustainable Development Goals and End-of-Life Electric Vehicle Battery: Literature Review. Batteries, 9, 353.
- [25] Chan, C. K., Chung, C. H., & Raman, J. (2023). Optimizing Thermal Management System in Electric Vehicle Battery Packs for Sustainable Transportation. Sustainability, 15(15), 11822.