© September 2026 | IJIRT | Volume 13 Issue 4 | ISSN: 2349-6002

Comprehensive Overview of Deep Neural Network Architectures for Gesture Recognition

Nowreen Afsana[1], Sumaiya Tabassum Dipty[1], Ishmam Fatima[1], Ferdous Karim Lucy[1], Jafrin Sultana[5], Md Ushama Shafoyat[6], Mahamudul Hasan Ashik[7]

[1,2,3,4] *Department of Biomedical Engineering, Bangladesh University of Engineering & Technology (BUET)*

[5,6] *Department of Biomedical Engineering, Military Institute of Science and Technology (MIST)*

[7] *Department of Architecture, Rajshahi University of Engineering and Technology (RUET)*

doi.org/10.64643/IJIRTV13I4-208315-459

***Abstract*—Hand gesture recognition (HGR) is a human–computer interaction (HCI) task that predicts the class and timing of a given hand movement. This comprehensive literature review examines state-of-the-art real-time hand gesture detection models employing deep learning. Deep learning algorithms have gained prominence in recent years due to their ability to automatically learn discriminative features from vast volumes of data. This paper will cover data acquisition methods, feature extraction, hand gesture classification, recently proposed applications, prominent environmental factors that affect accuracy, and a discussion of the performance of hand gesture recognition systems developed by various researchers over the past decade. Surface electromyography (sEMG) sensors integrated with wearable hand-gesture devices were the predominant acquisition modality in the reviewed studies. The review indicates that convolutional neural networks (CNNs) were the most frequently used classifiers and among the most effective approaches for gesture identification.**

***Index Terms*—Deep Learning, Hand Gesture Recognition, Human–Computer Interaction, Surface Electromyography (sEMG)**

I. INTRODUCTION

Electromyography (EMG) is a technique for measuring electrical impulses in muscle tissue. The sites where nerves stimulate muscle fibers are called motor units (MUs), and the voltage generated by the MUs is called the motor unit nerve impulse (MUAP). EMG is now used in medical tests worldwide to assess the health of muscle fibers and the neurons that control them. EMG signals are also commonly used as control signals for prosthetics (arms, hands, and lower limbs) [1]. Because the surface EMG signal reflects nerve activity, it can be used to retrieve neural information

and offers the benefits of nonsurgical acquisition, bionics, and so on. It is well-suited for use in prosthetics regulation, clinical assessment, gesture recognition, and neurological restoration [2], [3]. Gesture recognition using surface EMG signals is a critical research area for practical surface EMG applications, as accurate and convenient gesture identification can aid the development of an effective human-machine interface. The EMG signal and limb motion are related. Muscle tension or relaxation results in distinct limb motions, eliciting a significant variation in bioelectrical signals [4]. EMG-based recognition systems are critical across a variety of domains, including neuromuscular disease diagnostics, human-computer interfaces, console gaming, sign language recognition, virtual reality (VR) applications, and amputee device control. There has been a surge of interest in automated motion detection for human-computer interaction (HCI) during the last several years. Among the several sensor modalities used to capture hand signal data, electromyography (EMG) is believed to be the most suited since it records the muscle's electrical activity. In contrast to invasive techniques that penetrate the skin to reach the muscle, surface electromyography (sEMG) maps a muscle's electrical activity from the skin surface.

For dealing with EMG-based inputs for gesture identification, tactics such as Deep Learning models have been well- employed. Deep learning is a subfield of machine learning that relies entirely on artificial neural networks (ANNs). Because neural networks are intended to resemble the human brain, deep learning is also a form of brain mimicry. Deep learning's capacity to handle a wide variety of features makes it extremely effective for managing large datasets. Deep Learning (DL) is a subset of Machine Learning (ML) algorithms that have revolutionized several disciplines of data analysis. Deep learning models represent a paradigm shift in the fields of artificial intelligence (AI) and machine learning. Recent breakthroughs in gesture recognition with this approach have sparked tremendous interest in this sector. On the other hand, the computational and statistical techniques that underpin deep learning models are quite intriguing.

Several studies on hand gesture recognition have been published in the past decade. The study [5] discussed methods for acquiring gestures, the feature extraction process, classification of hand gestures, recently proposed applications, the challenges researchers face in hand gesture recognition using ANNs, and the future of hand gesture recognition. Another study by [6] presented a survey on current deep learning methodologies for action and gesture recognition in image sequences, describing and discussing the major studies proposed so far, with a special emphasis on how they deal with the temporal component of data, highlighting their key aspects, and proposing possibilities and difficulties for future study. Other authors [7] evaluated the numerous strategies used to create the system, including multiple segmentation and tracking algorithms, feature extraction, and several hand motion detection techniques. In addition, authors [8] investigated hand gesture recognition systems, including the key stages of recognition and the approaches and techniques used at each step.

The papers above provide insight into several methods for gesture recognition systems and cover aspects such as feature extraction accuracy, database type, classification algorithm employed, application, research obstacles, and future scope. However, no review article evaluates the accuracy of gesture recognition using Deep Neural Networks (DNNs). As such, the purpose of this work is to conduct a comparative assessment of recent research on several Deep Learning models for gesture recognition and categorization using various technologies. This paper also discusses these approaches in detail. It summarizes recent research across various considerations (type of device, segmentation/Data acquisition technique, Classification algorithms, recognition rate/Accuracy, number of gestures, Dataset used, hardware used). Additionally, the study highlights the most frequently used applications in this field and the research gaps and obstacles that could be addressed in future studies. This paper is divided into V sections, with Section II describing hand gesture recognition methods. In section III, Application area of hand gesture recognition systems. In section IV, the research gaps and challenges are presented. Finally, Section V presents the conclusions.

II. HAND GESTURE RECOGNITION METHODS

Common Gesture recognition methods based on Deep Learning consists of architectures such as Convolution Neural Network (CNN), Region-based convolutional neural network (R-CNN), Recurrent Neural Network (RNN), Deep Neural Network (DNN), Transfer learning (TL), or TCN (Temporal Convolution Networks), Long short-term memory (LSTM) network, Feed-forward neural network (FFNN), general regression neural network (GRNN), Hybrid Recognition Model (HRM), Temporal Convolutional Network Model (TCNM) or TCN (Temporal Convolution Networks), ADCNN (Adaptive Distributed Convolution Network), DBN (Deep Belief Network), DCN (Deep and Cross Network), BiLSTM (Bidirectional Long Short-Term Memory), MPCNN (Max-Pooling Convolutional Neural Network). The examination with these advanced techniques and the investigation of the outcomes reveal how accurately any DL architecture performs and the improvements required to improve performance.

The primary goal in studying gesture recognition is to introduce a system that can detect specific human gestures and use them to convey information or for command-and-control purposes. Artificial intelligence is a reliable technique used in a wide range of modern applications because it relies on the principle of learning. Deep learning uses multiple layers to learn from data and produces accurate predictions. The most challenging aspect of this technique is the dataset required to train algorithms, which may affect processing time. Table I presents a set of research papers that use different deep learning-based techniques to detect ROI. A recent work based on deep learning shows that using a convolutional neural network, a transfer learning (TL) strategy for instantaneous gesture recognition can be proposed to improve the generalization performance of the target network. The dataset Capgmyo (128 channel) with a sampling rate of 1000 Hz and NinaPro DB (10 channel) with a sampling rate of 100 Hz was used to evaluate the validity of the strategy, where the average instantaneous recognition accuracy for TL strategy DB-C is 95.97 ± 2.95% and for non- TL strategy DB-C was 81.6 ± 18.9% [9]. Other authors take a different approach, combining handcrafted features from a time-spectral observation, such as wavelet transform (WT) coefficients, peak frequencies, mean absolute value (MAV), slope sign changes (SSC), and so on, with deep features derived from a convolutional neural network (CNN), which recognizes signals captured from a single-channel sensor to generate the feature vector. CNN and a multi-layer perceptron classifier (MLPC) are then used to classify the 54 errors' feature vector. The experimental findings indicated that for 8, 6, and 5 hand motions, the average classification performance was 81.54%, 88.54%, and 94.19%, respectively. Furthermore, the CNN model achieved accuracies of 66.64%, 74.44%, and 74.80% for 8, 6, and 5 hand motions, respectively [10].

TABLE IComparative Summary of Deep Learning-Based Hand Gesture Recognition Studies

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
1	2021	Source spatial feature extractor	-	Capgmyo (128 channel) NinaPro DB (10 channel)	CNN, TL	TL strategy 95.97±2.95% non-TL strategy 81.6±18.9% (DB-c)	3	CapgMyo Ninapro-DB1	-One Xeon 5122 CPU -One Titan Xp GPU
2	2021	Handcrafted features derived from time-spectral analysis and convolutional neural networks, as well as a hybrid feature technique utilizing both handcrafted and deep features.	-	Recognize EMG signal gestures using a single-channel device.	MLPC, CNN	MLPC: 98.00% CNN: 76.14%	4	10 gesture classes were selected by author; the subject repeated all the gestures 100 times	Intel i7, 7700hq processor with 16 Gb of RAM and a GTX 1060 GPU
3	2021	ACC, LOG, WL, AR4, DASDV, MNP, spatial features (CNN Feature, CNNFeat) extraction and temporal features (LSTM Feature,	-	-	SVM, LDA, KNN, CNN, LSTM	Hybrid CNN-LSTM 77.87%	Basic gestures - 5 Finger gestures - 5	Two sets of 10 gestures were used for this experiment from Elonxi DB, comprising	-

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
		LSTMFeat)						eight subjects	
4	2020	-	BIOPAC MP36 device	-	ResNet based CNN architecture	CNN-99.59%	7	NinaPro database	-
5	2020	Sensor wise method	Non-invasive wearable device consisting of 8 acquisition modules	-	ConvNet algorithm	100% (99.99%)	27	CapgMyo DB-a database	-
6	2020	Temporal features, spatial features	-	-	CNN, LSTM	98.7%	Basic movements of the wrist -9 isometric and isotomic hand movements-8	Second dataset from NinaPro called the DB2	-
7	2020	-	Myo Armband	Sliding window of n = 66 points. Window divided subsequently into 6 subwindows of length m = 11 each	Shallow feed-forward ANN	92:45±11: 00%	5	Dataset of EMGs of 120 users	4-core processor of 4GHz and16GB of RAM.
8	2020	PCA to reduce feature dimension RMS, WL, MAS and (SampEn)	Electromyography instrument	N-Sliding window	GRNN	95.1%	9 static gestures	-	-
9	2020	-	Six bipolar Ag/AgCl electrode	-	CNN	82.6% ±13.9%	10 active upper extremity movements	-	Quad-core CPU, 16 GB RAM and 8 GB NVIDI

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
									A GTX 1070M GPU.
10	2020	-	Myo armband	Sliding window strategy with 150 sampling points	CNN, LSTM	95.3%	27	Two benchmark datasets (i.e., CapgMyo DB-a and CSL-HDEMG)	-
11	2020	The WPT decomposes a signal into its time and frequency components (low frequencies).	-	Using the My Signals HW V2 card	R-CNN	96.48%	4	Sub dataset from Jester Dataset	-
12	2020	Traditional manual feature extraction method	Myo armband	Sliding window consisting of 250ms	CNN, RNN, TCN	99.31%	13	Author created dataset from previous literatures	GPU or cloud devices
13	2020	Spatial Features	Delsys Trigno Wireless EMG System with 12 wireless electrodes (channels)	Sliding windows with steps of 10 ms	CNN, LSTM	98.01%	17	Ninapro (HGR) Dataset	GEFORCE GTX 1080 Ti Graphic Card
14	2020	Hudgins' Time Domain feature, autocorrelation and cross-correlation coefficients and spectral power magnitudes of EMG signals	HD EMG arrays, each comprising 64 sensors	Heatmap Differencce Technique	CNN, VGG	98.96%	8	Dataset was recorded using two HD EMG arrays, each comprising 64 sEMG sensors	Built-in hardware filters
15	2019	TEMPONet has its own feature extraction	Pre-installed environmental cameras	15 ms of the sliding window approach	TCN, TEMPO Net	93.7%	8	NinaPro DB6 and 20-session dataset	SPI processing platform and the GAP8 process

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
									or
16	2019	-	DualMyo	SD along time of a window of 100-time steps (0.5s)	FFNN, GRN, RNN	95%	8	UC2018 DualMyo, NinaPro DB5 data set	Nvidia GTX970M GPU (6GB VRAM), Intel i7-6700HQ CPU and 32GB of RAM.
17	2019	-	Myo Armband	Sliding window approach	Feed-forward ANN	85.08 ± 15.21%	5	Dataset created by authors	Intel® Core™ i7-3770S processor and 8GB of RAM.
18	2019	MAV, ZC, SSC, WL,Skewness, RMS, Hjorth Activity, Integrated EMG	Myo Armband	a rectangle sliding window of 250 ms (50 samples) length	RNN	99.78%	7	Myo Armband Dataset	NVIDIA Quadro M4000 GPU.
19	2019	Signal obtained from each of the eight sensors and its associated feature maps.	Myo Armband	50 ms wide analysis window and a step of the same size for each analysis (Myo sampling time is 200 Hz)	CNN	99%	6	Author created own database	i7-7700HQ, 7th Generation Intel® Core™ NVIDIA® GeForce® GTX 1070, 8 GB GDDR5
20	2019	Five features in the time domain: MAV, SSC, WL, RMS, and HP	Myo Armband	sliding window of 400 ms	FFNN	98.7%	5	Author created own database	-
21	2019	CNN Feature	Webcam	Skin color	CNN	95.61%	6	4800	

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
		map		detection and morphology & background subtraction				images for trains and 300 for test	-
22	2019	Spatiotemporal feature extraction	Different mobile cameras	Features extraction by CNN	ADCNN	99%	7	Created by video frame recorded	Core™ i7-6700 CPU 3.40 GHz
23	2018	Automatic feature extraction	Myo armband	DTW distance for multi-channel time series	FFNN	90.1%	5	Author created own database	Intel® Core™ i7-3770S processor and 4GB of RAM memory.
24	2018	Time-domain features MAV, ZC, SSC	-	16 channel EMG acquisition system (Elonxi Ltd, UK)	CNN-SVM	68.2%	Basic gestures - 5 Finger gestures - 5	10 gestures were used for this experiment by eight healthy subjects	-
25	2018	GengNet is employed for feature extraction	Electrode	Sliding window with 200ms length	CNN-RNN	99.7%	8	CapgMyo-DBa	NVIDIA TITAN Xp GPU, one Intel 6850k CPU
26	2018	Manually designed feature extractors	Electrode	Sliding window with a length of Lms (2L samples)	CNN	78.86%	50	NinaPro database.	-
27	2018	sEMG features are extracted adaptively by trained convolutional networks	Trigno Wireless Electrode	Sliding window with a length of 100 ms	CNN	83.23%	17	Second database (DB2) from the Ninapro project	-
28	2018	Automatically extract characteristics from the input of raw EMG	Electrode	-	CNN-SVM	68.2%	10	Author created own database	-

Aut hor	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classific ation Algorith m	Recognitio n Rate	No. of Gesture s	Dataset	Hard- ware Run
		image using CNN structure							
29	2018	LCNN approach does not need feature design for sEMG.	Myo Armband	Sliding window of 260 ms (52 sample points)	LSTM, CNN, LCNN	98.14	5	MyoDatas et /DB5 dataset	-
30	2018	Time- and frequency domain features are extracted from 8 channel myoelectric signals i.e. Zero Crossings, Slope Sign Changes & Cepstral coefficients	10 High Density Electrode Array	Segmented into overlapping windows of 150ms length and 90ms overlap	DBN, kNN, SVM, LDA, RF	85%	5	Ninapro database	-
31	2017	-	Non-invasive wearable device	A band-stop filter (45–55 Hz, second-order Butterworth)	DCN	85.0%;	52	(NinaPro, CSL HDEMG and our CapgMyo dataset	NVidia Titan X GPU.
32	2017	-	Myo Armband	Hann windows of 28 Points per FFT	CNN	97.81%	7	Author created own database	6D of Robotic Arm
33	2017	-	Armband with one IMU and three MMGs	-	CNN	94%	8	Data was taken from for the offline trainning of the neural network from 27 randomly choosen students	Two crazy FLIE 2.0 Quadco pter and three webcam s
34	2017	-	Myo Armband	Conversion into time-frequency domain by a wavelet transform	CNN	83%	2	-	Core i5 CPU, 8 GB of RAM, CentOS 7, two GPUs (GTX 670)

Author	Year	Feature Extract Type	Type of measuring device	Techniques/ Methods for Segmentation	Classification Algorithm	Recognition Rate	No. of Gestures	Dataset	Hard-ware Run
35	2016	Spatial and semantic features of hand gesture	-	Preprocessed image by canny edge detection	CNN	98.2%	5	Cambridge Hand Gesture Data (CHGD) and Author constructed digit hand gesture data	-
36	2014	Binarized image features (position and orientation) using a threshold and GMM	XBOX Kinect camera	Conversion to the YCbCr color space (skin color segmentation) and the background removal by Gaussian Mixture model (GMM)	CNN	95.96%	7	Author constructed hand gesture images from 7 subjects	-
37	2013	Sampled linear acceleration and angular velocity	Android Nexus S Samsung device	Temporal segmented gesture images sampled at 40 ms	BLSTM -RNN	95.57%	14	Author constructed hand gesture images from 22 subjects	Intel Core i5 CPU at 2.67 GHz with 3.42 Go of RAM.
38	2011	Gestures are presented to the foot-bots using colored gloves, images are captured by the front (CMOS) robot camera using a Bayer color filter array	Foot-bot mobile robot	Using a parametric Single Gaussian Model (SGM) method	MPCNN	96%	6	Author constructed digit hand gesture data	ARM 11 533MHz Processor with 128MB RAM

Another study employed a hybrid CNN-LSTM network with two concurrent feature extraction stages: spatial features extracted with CNNs and temporal features extracted with LSTMs. Using Hybrid features, the CNN-LSTM network integrates spatial and temporal data. It feeds them to standard classifiers such as linear discriminant analysis (LDA), K-nearest neighbors (KNN), and support vector machines (SVMs). The hybrid CNN-LSTM network achieves an average accuracy of 77.87% [11]. One author used sEMG signals and segmented them into sections corresponding to each movement. The spectrogram images of the segmented sEMG signals were created using the Short-Time Fourier Transform (STFT). Then, the created colored spectrogram images were trained with a 50-layer Convolutional Neural Network (CNN) based on the Residual Networks (ResNet) architecture. Test accuracy of 99.59% and an F1 score of 99.57% were achieved for 7 different hand gesture classifications [12]. In another study, the Sensor-Wise method was introduced, which has a higher capability to extract features than the sEMG image method due to its convolutional network architecture and its high compatibility with the nature of sEMG signals. This method has shown no signs of overfitting and achieved an accuracy of almost 100% (99.99%) due to its optimal two-hidden-layer structure and strong generalization on the 96-electrode CSL HDEMG database [13].

One author proposes a model with two parallel paths: one using Long Short-Term Memory (LSTM) and the other using Convolutional Neural Networks (CNN), with a fully connected multilayer fusion network serving as the fusion center that combines the outputs of the two paths to perform classification. Temporal characteristics are extracted using the LSTM path, while spatial features are extracted using the CNN path. The suggested hybrid architecture was tested using the NinaPro DB2 dataset. The extensive collection of trials and analyses using state-of-the-art algorithms on the same dataset demonstrates that the proposed architecture significantly outperforms competitors and achieves an outstanding accuracy of 98.01% in gesture classification [14]. In [15], based on electromyographic (EMG) signals measured with the Myo Armband, a bag of functions for feature extraction, and a shallow feedforward neural network for classification, the hidden layer has 500 neurons, with the ReLU activation function. The recognition accuracy of this model was 92.45%, with a standard deviation of 11.00% and an average processing time of 40.58 ms, which was less than the permitted delay of 300ms for real-time gesture recognition models.

The principal component analysis (PCA) method and GRNN are used to construct the gesture recognition system in the paper [16]. The authors use four features to be integrated into one feature set. The feature extraction set includes root mean square (RMS), wavelength (WL), sample entropy (SampEn), and median amplitude spectrum (MAS) to reduce feature dimension, and the PCA algorithm eliminates redundant information. After dimension reduction, the first three principal components were used for pattern recognition in the GRNN classifier; the recognition rate was 95.1%, and the average operation time was 0.19 s. Another author used a convolutional neural network (CNN) to decode hand gestures from sEMG data recorded from 18 subjects, observing the effects of hyperparameters on each hand gesture. Results showed that the learning rate set to either 0.0001 or 0.001, with 80-100 epochs, significantly outperformed other settings (p ¡ 0.05). Moreover, it was observed that regardless of network configuration, some motions (close hand, flex hand, extend the hand, and fine grip) performed better (83.7% ± 13.5%, 71.2% ±20.2%, 82.6% ± 13.9%, and 74.6% ± 15%, respectively) throughout the research. For reducing the problem of vanishing gradient and improving training speed, activation function (ReLU), rectified linear units were used [17]. High-density surface electromyography (HD-sEMG) and deep learning have become increasingly employed in gesture recognition. In paper [18], depth experiments were conducted on 2 benchmark datasets (i.e., CapgMyo DB-a and CSL-HDEMG) using the planned 3D CNN, which was designed to support the platform. The accuracies of the lengths of 10 ms (20 frames) and one hundred fifty ms (307 frames) were 51.4% and 70.4% with 2nd CNN + MV, and were 67.0% and 75.9% with 3D CNN.

Another author proposes using R-CNN to identify hand movements from electromyography signals. The signal collection was performed using surface electrodes on the forearm, and biomedical signals were generated for signal cleaning and feature extraction using the wavelet packet transform. R-CNN was utilized to map the characteristics extracted from the wavelet power spectrum to verify and train the architecture’s framing. Additionally, the real-time test was performed to achieve a precision of 96.48% when compared to conventional procedures [19]. The approach of fine-tuning strategy, along with DL Methods, including CNN, RNN, and TCN, is a supervised learning approaches that end up in the recognition of an associated exaggerated number of gestures, ranging from a system trained for discriminating between a low number of categories of recognized gestures, with available online datasets. The authors of the paper [20] removed the 0.25s processing time at the start and at the top of the 5s, resulting in 4.5s of useful signal per gesture. The results show that the projected methodology can be constructed with 7 gestures and, with a comparative touch of additional coaching, will faithfully classify a total of 13 gestures, achieving an overall accuracy of 99.31% compared to 99.78% for 7 gestures. In [21], the authors implemented a Hybrid Recognition Model for which the input multichannel sEMG signals were used. The planned HRM design consists of 3 main blocks, i.e., the CNN block, the LSTM block, and the Fusion block. During this case, the length of trials for active signals was precisely up to five or 10,000 samples. It's ascertained that the planned design will turn out promising results with a high accuracy of 98.01% with 17 gestures. It's jointly observed that architecture can produce promising results with high accuracy, which was 8.71% larger than the progressive results. In [1], the authors use Heatmaps to indicate the electrical activity (energy level) of various sEMG sensors to obtain the input data, specifically because of different gestures that end in different levels of target muscle activity, with the intersection Algorithm with a VGG-based CNN as the deep learning architecture. The heatmap distinction approach considerably outperforms alternative methodologies in terms of accuracy, as several of the methodologies perform higher, reaching as high as 98.96% underneath and 98.96% under thirty-five sensors conditions, for eight hand gesture recognition supported by the combined two bands from twenty-three participants, together with fourteen amputees.

Moreover, another author developed a unique TCN topology (template) and tested the solution on a benchmark dataset (Ninapro) with 20 sessions, achieving 49.6% average accuracy, which is higher than the current progressive (SoA). Moreover, the authors wanted to associate energy-efficient embedded platforms with GAP8, a unique 8-core IoT processor. By victimizing their embedded platform, they collected a second 20-session dataset to validate the system in a setup representative of the ultimate deployment. This model reaches 93.7% average accuracy with the TCN, comparable to a SoA SVM approach (91.1%) with 8 gestures [22]. The author of [23] discusses and compares the performance of their method, which consists of a Feed- Forward Neural Network (FFNN), a recurrent Neural Network (RNN), a long short-term memory network (LSTM), and a Gated Recurrent Unit (GRU). The model has been evaluated on the UC2018 Dual Mayo and NinaPro DB5 dataset. It's incontestable that the static model (FFNN) and the dynamic models (LSTM, RNN, and GRU) attain similar accuracy for each data set, i.e 95% for the DualMyo and about 91% for the NinaPro DB5, specifically for the 8 gestures of diagnostic technique (EMG) signal non-inheritable from the forearm muscles. In paper [24], the Myo Armband was used to capture EMG data on a user's forearm for the hand gesture detection model suggested in this study. To identify the EMG data, a feed-forward multilayer neural network was implemented. The network architecture consists of a feed-forward ANN with 4 layers: input, autoencoder, ReLU hidden layer, and output. Their model can recognize 5 different hand gestures: fist, wave left, wave right, fingers spread, and double tap, with an average recognition accuracy of 85.08 ± 15.21%. The time of processing of the model was 3 ± 1 ms. Based on sEMG signals, the research in [25] presents a real-time automatic gesture detection system for seven basic gestures. This paper presents a fully connected architecture with a forward pass time of fewer than 4.5ms, including the feature extraction stage. Backpropagation is used to find the model parameters, rather than the traditional Stochastic Gradient Descent (SGD). The suggested network was made up of six hidden layers, each with 128, 128,128, 64, 32, and 16 neurons. To avoid overfitting, a batch normalizing step was conducted after each layer. Except for the output layer, which utilizes Softmax activation, other layers use ReLU activation. The overall accuracy was achieved by their recognition system (99.78%).

Another model in [26] used an adaptable convolutional neural network architecture (CNN) to recognize six distinct hand gestures using EMG inputs. This was accomplished by creating a database of 2880 multi-channel feature maps; each dataset was made of the treated signals from the eight sensors on a MyoArmband through power spectral density maps. This enables architecture to be taught, achieving 98.4% accuracy validation and 99% accuracy in testing. In another paper, the authors employ a forward neural network with three layers in their work: the input layer, the hidden layer, and the output layer. A sigmoid transfer function was used in the model. It was trained using complete batch gradient descent with a cost function based on cross-entropy. The model responds in 227.76 ms from the beginning of the gesture, which was lower than the limit defined for real-time (300 ms). At the same time, the model shows a recognition accuracy of 98.7% [27]. The method of [28] employs a camera to track the region of interest (ROI), namely the hand region, in the picture range and distinguish hand gestures in real- time. The background was then removed using the skin color (Y–Cb–Cr color space) technique and morphology. Furthermore, kernel correlation filters (KCF) were utilized to track ROI. The final image was sent into a deep convolutional neural network (CNN). The CNN model was used to compare the performance of two modified Alex Net and VGG Net models. The recognition rate for training and testing data was 99.90% and 95.61%, respectively.

A Convolutional Neural Network (CNN)-based system for distinguishing hand gestures was demonstrated in [29]. For hand classification, feature extraction, and an adapted deep convolutional neural network (ADCNN) were used. The experiment assesses the training data at 100% and the testing data at 99%, with an execution duration of 15,598s. [30] proposes a real-time hand gesture recognition model based on both a shallow feedforward neural network with 3 layers and electromyography (EMG) of the forearm. The model achieves 90.1% accuracy in recognizing 5 gesture categories (fist, wave-in, wave-out, open, and pinch) and an average response time of 11 ms. Other authors used a convolutional neural network (CNN) to extract feature attributes from raw EMG input images. The gestures were then identified using a Support Vector Machine (SVM) classifier. The trials revealed that the suggested technique achieved an accuracy of 68.2%, which was approximately 2.5% higher than using solely a CNN and approximately 9.7% better than using the classical time-domain feature and an SVM-based classifier [11].

One author proposes a novel technique based on a CNN-RNN to better capture the temporal properties of EMG signals. A new sEMG image representation method was presented, based on a traditional feature vector. After evaluating five sEMG benchmark databases (NinaProDB1, NinaProDB2, BioPatRec26MOV, CapgMyo-DBa, and csl-hdemg), the results show that the method outperforms all the existing methods for both sparse multichannel and high-density sEMG databases. The recognition accuracies for the five databases are 87.0%, 82.2%, 94.1%, 99.7%, and 94.5%, respectively. The results show that the method's performance was 9.2%, 3.5%, 1.2%, 0.2%, and 5.2% higher than the state-of-the-art, respectively [31]. Another author proposes a simultaneous, multiple-scale convolutional architecture for limb motion recognition. Compared with existing approaches, the proposed architecture considers all properties of the sEMG signal. They have considered kernel filter sizes larger than those typically used in other CNN-based hand recognition algorithms. Besides the features of the sEMG signal, namely, muscle independence, it has been considered in this case. This method was examined across all categorization algorithms in the NinaPro database. As the results came in, the classification algorithm achieved 78.86% accuracy, which was higher than that of the state-of-the-art methods [32]. Some other authors developed a novel deep learning-based model to improve the accuracy of EMG-based hand gesture detection. They have used five convolution layers in a parallel design. When utilizing a convolutional neural network, the findings show a small improvement. The CNN-based technique achieved classification accuracies of 83.23% with preprocessing and 82.16% without preprocessing, both higher than those of traditional methods. In all the experiments, the SVM approach had the lowest average classification accuracy. The results showed that preprocessing can improve the average classification accuracy. The test results with and without preprocessing suggest that the CNN-based technique is more reliable than the traditional method [33].

Another study employed a convolutional neural network (CNN) to automatically extract properties from raw EMG images, without using traditional extractors. The hand motions were identified using a Support Vector Machine (SVM) classifier. This technique achieved an accuracy 2.5% higher than CNN and 9.7% higher than the time-domain feature and an SVM classifier. This synthetic technique outperformed any single classifier in the recognition of myoelectric signals-based motions [11]. Another author leveraged the complementarity between Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs) by integrating them into an architecture we refer to as LSTM-CNN (LCNN). This model can dynamically recognize gestures by directly feeding preprocessed EMG signals into the network. To extract information from signals, the LSTM model was utilized. The CNN model may be used to extract secondary features and classify signals. The suggested model achieved an average recognition accuracy of 98.14% [34]. Compared to similar works evaluated on a similar Ninapro dataset that includes EMG data related to fifty-three hand movements of 78 subjects (11 trans-radial amputees, 67 intact subjects) divided into three datasets, the projected model of [35] outperforms the initial network of, whereas it's inferior to the additional complicated approaches using ML classifiers, such as k-Nearest Neighbors (kNN), Support Vector Machines (SVM), Multi-Layered Per- ceptron (MLP), Linear Discriminant Analysis (LDA), Random Forests (RF) Deep Belief Network (DBN) and finally gained state-of-the-art accuracy of 85% on the Ninapro DB-1 with 53 gestures [35].

In [36], the authors employed a ConvNet with a fully connected G-way layer and a softmax function, which comprises the Deep Convolutional network, where G was the number of gestures. ReLU nonlinearity after each hidden layer, batch normalization after the input and before each ReLU nonlinearity, and dropout with a probability of 0.5 after the fourth, fifth, and sixth layers are all included in the model. The deep-learning framework was built on MxNet, a multi-language machine learning library designed to make it easier to build ML algorithms, particularly deep-learning algorithms. The identification accuracy on the CSL-HDEMG dataset was 85% and 67.4% when identifying 52 finger gestures from NinaPro. In another study, the authors used two datasets of 18 and 17 able-bodied subjects recorded with a Myo wristband, a low-cost, low-sampling-rate (200Hz), 8-channel, consumer-grade, dry-electrode sEMG device (Thalmic Labs). To exploit inter-user data from the first dataset and ease the data creation load imposed on a single individual, a convolutional neural network (CNN) was enhanced in the work, utilizing transfer learning techniques. The findings indicate that the suggested classifier was robust and accurate enough to guide a 6DoF robotic arm with the same speed and precision as a joystick (when combined with orientation data). For the 17 participants in the second dataset, after four training cycles, the target network achieved an average accuracy of 97.81% across seven hand/wrist movements [37].

For operator control of UAVs, another author proposes a robust wearable sensor suite that combines arm motion and hand gesture recognition. To record arm movement and hand gesture signals, the sensor suite combines mechanomyography (MMG) and an inertial measurement unit (IMU). The IMU generates world-referenced orientation and acceleration data, while the MMG monitors muscle activity via surface vibration. Surface muscle vibration for gesture identification eliminates the need for electrical contact with the skin, which has hampered the application of other muscle-measurement techniques in the field. In their study, hardware design, inertial arm movement recognition, and the full architecture of a convolutional neural network (CNN) were used for real-time hand gesture detection based on MMG data. With easy calibration for each user, the system achieved 94% accuracy for five gestures, resulting in an intuitive gesture- based UAV control system. According to the authors, this was the first wearable technology to provide multimodal control of UAVs using intuitive gestures without electricity [38]. Using a convolutional neural network, other authors have developed a technique for calculating finger motion based on frequency-domain conversion of electromyogram (EMG) inputs and image identification (CNN). By categorizing pictures obtained from a wavelet transformation of EMG data, a basic CNN model was utilized to estimate finger motion. The model has two pairs of convolution and pooling layers, as well as two fully linked layers, and was initially designed for document recognition. To gather EMG data and record finger motion, a prototype system comprising low-cost sensor devices was built. The test accuracy in identifying EMG signals into four categories reached 83% when a thumb opened or closed, and when fingers, excluding the thumb, opened or closed, according to the data [39].

Another study developed a robust approach to hand gesture detection using convolutional neural networks. This algorithm was used to dynamically obtain the physical and semantic features of hand gestures. This method utilizes a revised Convolutional Neural Network architecture and data pre-processing with the Canny operator, a multi-level optimization descriptor for enhancing and filtering the hand gesture portrait and detecting edges. This method can determine the actual edge as accurately as possible and ensure that the identified edge points are as similar to the true edge points as possible, thereby improving hand gesture recognition accuracy to 98.2% with author-constructed data and 94.1% with CHGD data [40]. Additionally, several researchers explored using a convolutional neural network (CNN) to distinguish hand motions associated with human task activities from images captured by a camera. Given the significant effect of light on skin color, they used a Gaussian Mixture Model (GMM) to train the skin model, which was then used to reliably reject non-skin elements in an image. They validated the recommended approach for identifying human gestures in the study, demonstrating strong results across a range of hand postures and orientations, as well as lighting conditions. They evaluated seven patients performing seven hand gestures, achieving a recognition accuracy of approximately 95.96% [41].

Others suggested a robust approach for 3D gesture detection using inertial MEMS (Microelectromechanical systems). The linear acceleration and rotational velocity, provided by the accelerometer and gyroscope, respectively, are sampled over time, yielding 6D values (no additional pre-processing was conducted) at each time step, which were used as inputs to the gesture detection system. They developed a system for gesture categorization from raw MEM data using Bidirectional Long Short-Term Memory Recurrent Neural Networks (BLSTM-RNN). The recommended methodology, applied to 22 persons executing 14 motions, surpasses standard classification approaches, achieving a recognition rate of 95.57% [42]. Another model employed a cutting-edge deep neural network (NN) merging convolution with max pooling (MPCNN) for supervising feature learning and categorization of hand gestures made by mobile robots wearing colored gloves. Color segmentation was used to obtain the hand contour, which was subsequently smoothed using morphological image processing to remove noisy edges. With 96%, this large and deep MPCNN model detects 6 gesture types [43].

III. DISCUSSION

Deep Neural Networks (DNNs) have a significant impact on hand gesture recognition. Fig. 1 and Fig. 2 illustrate some of the successful works of gesture recognition based on the predictive number of classes of gestures. We can see that the accuracy of the maximum models exceeds 95% when the average number of recognizable gestures is only 6-8; however, when the number of gestures is increased to 50 or more, the accuracy of gesture classification is only 75-82%. This can be explained by noting that a lower number of gestures provides the model with fewer trainable features, thereby decreasing its complexity. When a deep learning model is trained with fewer features, it finds it easier to classify only the selected gestures and achieves very high accuracy. As a result, authors have regularly discussed this limitation in their research.

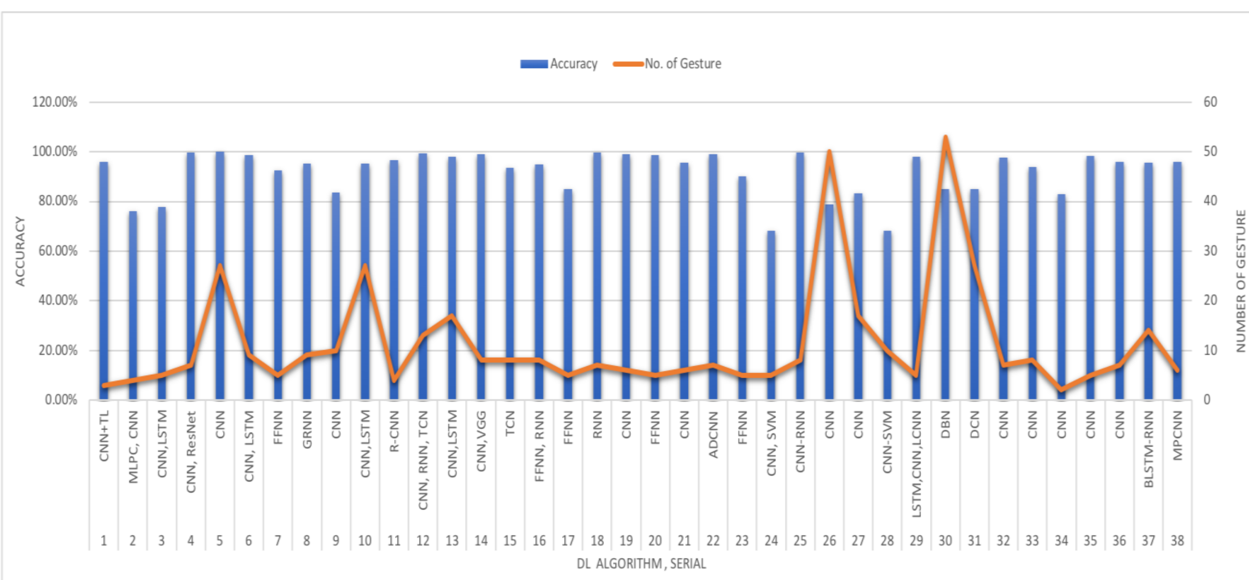

Fig. 1. Comparative performance of deep learning algorithms based on the number of recognized gestures.

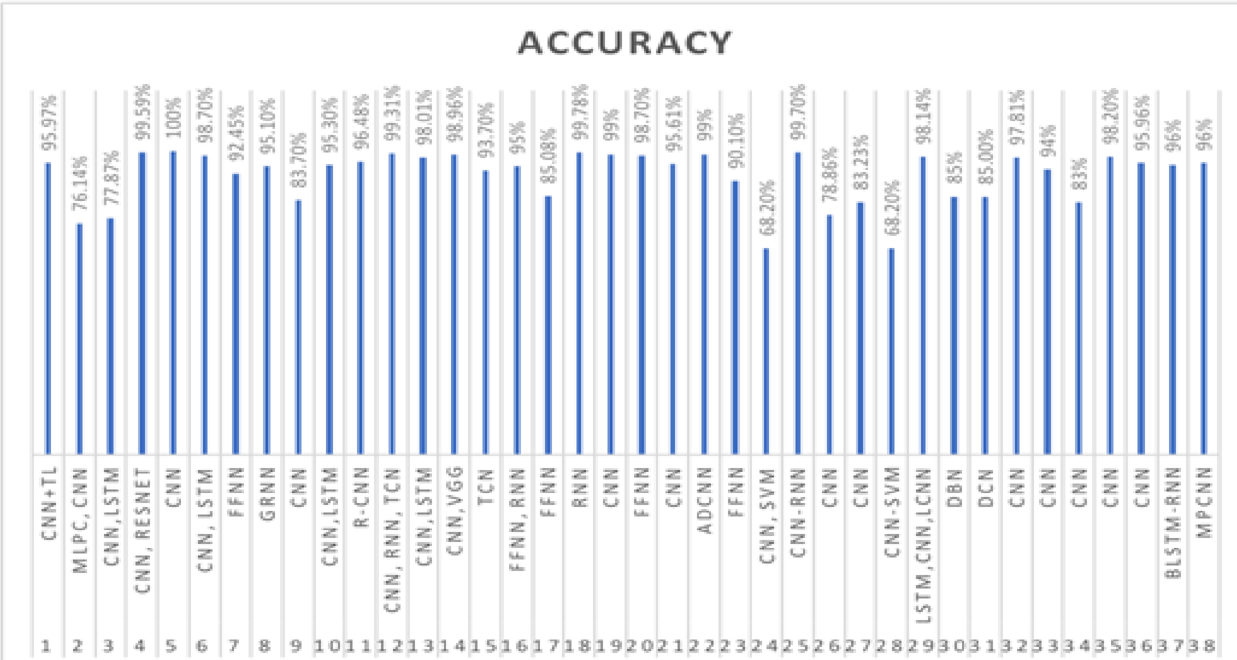

Fig. 2. Recognition accuracy of deep learning models for gesture recognition.

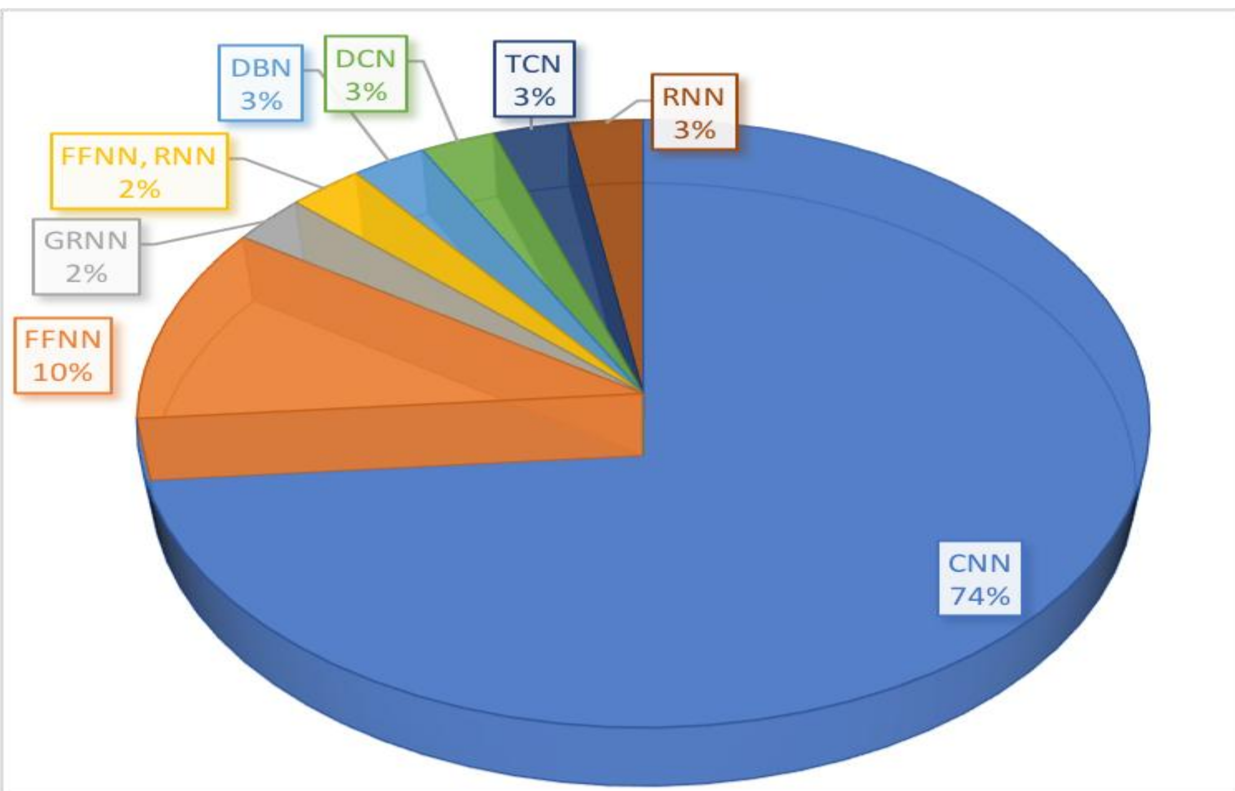

Fig. 3. Distribution of deep learning algorithms used for gesture recognition.

Fig. 3 shows that around 74% of gesture-detection deep learning models used a CNN architecture. The convolutional layer is the fundamental building block of CNN. A CNN convolutional layer learns features from the input data, making it well-suited for processing 2D data, such as photographs of gestures or

EMG signals in our example. Also, CNNs are extremely popular for their architectural style because they require no feature extraction. The system is trained to extract features, and the key notion of a CNN is that it combines image convolution with filters to produce invariant features that are passed to the next layer. Convoluting the features in the subsequent layer with various filters generates additional invariant and abstract elements, and the process is repeated until the final output is obtained. That is, it is precisely because of autonomous feature extraction that CNNs are so well-suited and efficient for classification tasks such as gesture recognition [44].

From Fig. 1 and Fig. 3, 26% of the other DL algorithms, from out selected papers, which led to better accuracies include- TL (Transfer Learning), LSTM (Long Short-Term Memory), FFNN (Feed Forward Neural Network), GRNN (General Regression Neural Network), RNN (Recurrent Neural Network), TCN (Temporal Convolution Networks), ADCNN (Adaptive Distributed Convolution Network), DBN (Deep Belief Network), DCN (Deep and Cross Network), BiLSTM (Bidirectional Long Short-Term Memory), MPCNN (Max-Pooling Convolutional Neural Network).

IV. APPLICATION AREAS OF HAND GESTURE RECOGNITION SYSTEMS

Hand gesture research has become an intriguing and vital topic; it provides a natural means of contact while reducing the cost of using sensors in data gloves. Conventional interactive techniques rely on devices such as a mouse, keyboard, touchscreen, joystick, and console for machine control.

A. MEDICAL CARE

A surgeon may require details about the patient's full body anatomy or a detailed organ model during clinical procedures to reduce operating time or improve the accuracy of the results. This is accomplished by employing a medical imaging system, such as an MRI, CT, or X-ray system [45], [46], which takes data from the patient's body and shows it as a detailed image on the screen. The surgeon can assist with the seen pictures by performing hand gestures in front of the camera. Some activities, such as zooming, rotating, image cropping, and switching to the next or previous slide, can be performed with these motions without the assistance of a peripheral device such as a mouse, keyboard, or touchscreen. Sterilization is required for any extra equipment, which can be challenging for keyboards and touchscreens. Hand gestures can also be used for assistance, such as wheelchair control [47]. Sign language is an alternative means of communication used by people who are unable to communicate verbally. It is made up of a series of motions, each of which indicates a different letter, number, or expression. Many research publications have advocated the use of a glove-attached sensor worn on the hand that generates responses based on hand movement to recognize sign language for deaf-mute persons. This might also incorporate uncovered hand interaction with the camera, with computer vision algorithms used to recognize the gesture. The dataset utilized for gesture categorization in both algorithms discussed above matches a real-time gesture made by the user [48], [49], [50].

B. BIOMEDICAL FIELD

Now, sEMG is widely employed in the medical field, biomedical engineering, and other sectors, most notably in the use of smart prostheses, which patients and rehabilitation engineering scientists have warmly embraced. Additionally, it serves as the optimal control signal for muscle-powered robotics and effective electrical stimulation. Additionally, several mature, commercially available myoelectric prostheses are available, such as the Otto Bock prosthetic hand (manufactured in Germany by Otto Bock) and the i-Limb multi-degree bionic hand (designed by David Grow). All these devices make life easier for people with disabilities [11]. In the case of identical network architecture, 3D convolutions can achieve higher performance than a combination of the 2nd convolution and majority voting, particularly for sEMG data, which contains dynamic processes [20].

C. REHABILITATION

Several applications of EMG gesture identification include developing a human-computer interface (HCI) or controlling prosthetic equipment [11]. Human-machine interfaces (HMIs) may enhance the quality of life for individuals with neuromuscular disorders. For these types of faults, interactions between a machine and a person must be natural, healthy, and high in communication. Typically, an HMI framework implements bidirectional interaction between a machine and a human through the exchange of information via response- and control-based interactions. Numerous biological signals, such as MMG, ultrasound, and EEG, are used to accomplish this type of control. Among these biological signals, EMG is most often employed for HMI [21], [29].

D. ROBOTICS

Robot technology is applied in a variety of industries, including industrial, assistive services [51], retail, sports, and entertainment. Robotic control systems employ machine learning, artificial intelligence, and complex algorithms to perform a given task, allowing the robotic system to interact naturally with its surroundings and make independent decisions. Some studies suggest combining computer vision technology with a robot to create helpful technologies for the elderly. Another study employs computer vision to allow a robot to ask a person for the optimum path within a specific building [52]. Vision-based gesture recognition can be used effectively for human-robot interaction, even on tiny, low-power robots such as household mobility robots (e.g., Roomba and Scooba) and swarm robotic systems [3].

E. VIRTUAL REALITY

Virtual worlds are built on a 3D model that requires a 3D gesture detection system to interact in real-time as an HCI. These motions can be used for modification and viewing, as well as for leisure activities, such as playing a virtual piano. The gesture recognition system uses a dataset to match an acquired gesture in real time [53]. The unique application of LSTMs and RNNs, normally used for online gesture classification, can offer benefits by simplifying the information chain from sensors to the classification model and the online output. Additionally, the methodology may be valid for larger datasets and can even be applied to alternative types of signals [26].

F. HOME AUTOMATION

Hand gestures may be effectively employed for home automation. Controlling lighting, fans, television, radio, and other devices is as simple as shaking a hand or making a gesture. They can be utilized to improve the quality of life of older people [54].

G. PERSONALIZED DEVICES

Hand gestures can serve as an alternative input device for communicating with a computer without a mouse or keyboard, such as dragging and dropping files across the desktop and performing cut-and-paste operations [55]. They may also be used to control slide show presentations [56]. Furthermore, they are used in conjunction with a tablet to allow deaf people to engage with others by moving their hands in front of the tablet's camera. This requires installing an application that converts sign language to text and displays it on the screen. This is like converting acquired voice to text. TCNM has considerably lower memory requirements for coaching than HRM during the implementation of novel expanded causal convolutions that gently increase the network's receptive field and utilize shared filter parameters [22].

H. GAMING GESTURES

The Microsoft Kinect for Xbox, which has a camera positioned above the screen and connects to the Xbox via a cable, is the best example of gesture-based gaming. The user can interact with the game using hand and body movements detected by the Kinect camera [57].

I. CRIME DETECTION

Hand gesture recognition has been a complex concept used in a wide variety of practical applications. For instance, surveillance cameras detect and classify hand gestures to prevent illegal activity. Additionally, several experiments have been conducted on hand gesture recognition. such as sign language recognition, deception detection, and automatic control [2].

V. RESEARCH GAPS AND CHALLENGES

Many research publications focus on improving mechanisms for hand gesture identification or on developing new technologies rather than on establishing direct implementation in health care. The researcher's primary problem is to develop a strong structure that addresses the most prevalent challenges with the fewest possible constraints and produces consistent, accurate outcomes. Deep learning and machine learning algorithms are used to identify interactive gestures in real time using datasets of gestures that already contain specified postures or gestures. While this technique can identify a vast number of movements, it has certain limitations in specific circumstances, including the possibility of missing some motions due to differences in the performance of the categorization algorithms. Additionally, it takes longer than the first strategy due to the fitting of a dataset involving many datasets [56]. In addition, the gestures dataset cannot be utilized by other frameworks. In this paper [21], the ReLU function is used, which sets negative values to zero, leading to significant information loss, as evidenced by the network's performance. The difficulties encountered when taking EMG measurements are discussed in [29], which finds that short- and long-term muscular tension, as well as electrode dislocation throughout the tests, are the primary variables affecting categorization. Different individuals express the same hand gesture uniquely, and even the same individual signals it uniquely each time. Additionally, vision-based hand gesture detection is affected by illumination, perspective, and other factors, as discussed in [1]. Due to the large number of variables in image space, it is critical to extract key image characteristics for an image-based human gesture identification system. To create an effective identification system, a substantial training set is often necessary, as are models of diverse motions [2].

VI. CONCLUSION

To summarize, we provided an in-depth analysis of deep- learning-based models for gesture recognition. We created a classification system that encompasses most of the fundamental and critical information on human gesture interpretation, and then discussed contemporary Deep Neural Network (DNN)- based approaches. The following are key subjects outlined: architecture, fusion techniques, datasets, and problems. Performance is based on the amount of data, regardless of the approach. The community is now focusing on developing larger data sets capable of supporting massive parametric deep models and on organizing challenges (using novel data sets and well-defined assessment protocols) that will advance the field's state-of-the-art and facilitate comparisons among various deep learning architectures. Most papers use several strategies for data augmentation and pre-training. Another technique for improving deep learning model performance is data fusion. Individual networks can be trained on various subsets of input data, primary characteristics, information cues, and so on, and then combined. We conclude that the paper provides a brief overview of the promise of deep learning for large-scale, real-time gesture recognition. A comparison of experimental results between the proposed novel hand gesture recognition methods and alternative similar approaches indicates the superior effectiveness of the new design. From previous sections, it is clear that there is a significant need to invest in developing credible, robust deep learning algorithms to achieve reliable results and higher recognition accuracy. Although the techniques mentioned above show the highest number of preferences among authors, other techniques also have their advantages and disadvantages. They may perform well in some challenging circumstances while being inferior in others.

REFERENCES

[1] W. Ke, Y. Xing, G. Di Caterina, L. Petropoulakis, and J. Soraghan, "Intersected EMG heatmaps and deep learning based gesture recognition," in Proceedings of the 2020 12th International Conference on Machine Learning and Computing, Shenzhen, China: ACM, Feb. 2020, pp. 73–78. doi: 10.1145/3383972.3383982.

[2] M. Ison, I. Vujaklija, B. Whitsell, D. Farina, and P. Artemiadis, "High-Density electromyography and motor skill learning for robust long-term control of a 7-DoF robot arm," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 24, no. 4, pp. 424–433, Apr. 2016, doi: 10.1109/TNSRE.2015.2417775.

[3] Y. Fang, D. Zhou, K. Li, and H. Liu, "Interface prostheses with classifier-feedback-based user training," IEEE Trans. Biomed. Eng., vol. 64, no. 11, pp. 2575–2583, Nov. 2017, doi: 10.1109/TBME.2016.2641584.

[4] W. Ding, G. Li, Y. Sun, G. Jiang, J. Kong, and H. Liu, "D-S evidential theory on sEMG signal recognition," Int. J. Computing Science and Mathematics, vol. 8, no. 2, pp. 138–145, Aug. 2017, doi: 10.1504/IJCSM.2017.083747.

[5] M. Yasen and S. Jusoh, "A systematic review on hand gesture recognition techniques, challenges

and applications," PeerJ Comput. Sci., vol. 5, p. e218, Sep. 2019, doi: 10.7717/peerj-cs.218.

[6] M. Asadi-Aghbolaghi et al., "A survey on deep learning based approaches for action and gesture recognition in image sequences," in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, DC, USA: IEEE, May 2017, pp. 476–483. doi: 10.1109/FG.2017.150.

[7] J. S. Sonkusare, N. B. Chopade, R. Sor, and S. L. Tade, "A review on hand gesture recognition system," in 2015 International Conference on Computing Communication Control and Automation, Pune, India: IEEE, Feb. 2015, pp. 790–794. doi: 10.1109/ICCUBEA.2015.158.

[8] S. Yang, P. Premaratne, and P. Vial, "Hand gesture recognition: an overview," in 2013 5th IEEE International Conference on Broadband Network & Multimedia Technology, Guilin, China: IEEE, Nov. 2013, pp. 63–69. doi: 10.1109/ICBNMT.2013.6823916.

[9] Z. Yu, J. Zhao, Y. Wang, L. He, and S. Wang, "Surface EMG-Based instantaneous hand gesture recognition using convolutional neural network with the transfer learning method," Sensors, vol. 21, no. 7, p. 2540, Apr. 2021, doi: 10.3390/s21072540.

[10] J. M. Fajardo, O. Gomez, and F. Prieto, "EMG hand gesture classification using handcrafted and deep features," Biomed. Signal Process. Control, vol. 63, p. 102210, Jan. 2021, doi: 10.1016/j.bspc.2020.102210.

[11] H. Chen, R. Tong, M. Chen, Y. Fang, and H. Liu, "A hybrid cnn-svm classifier for hand gesture recognition with surface emg signals," in 2018 International Conference on Machine Learning and Cybernetics (ICMLC), Chengdu: IEEE, Jul. 2018, pp. 619–624. doi: 10.1109/ICMLC.2018.8526976.

[12] M. A. Ozdemir, D. H. Kisa, O. Guren, A. Onan, and A. Akan, "EMG based hand gesture recognition using deep learning," in 2020 Medical Technologies Congress (TIPTEKNO), Antalya, Turkey: IEEE, Nov. 2020, pp. 1–4. doi: 10.1109/TIPTEKNO50054.2020.9299264.

[13] [13] A. Bahador, M. Yousefi, M. Marashi, and O. Bahador, "High accurate lightweight deep learning method for gesture recognition based on surface electromyography," Comput. Methods Programs Biomed., vol. 195, p. 105643, Oct. 2020, doi: 10.1016/j.cmpb.2020.105643.

[14] E. Rahimian, S. Zabihi, A. Asif, and A. Mohammadi, "Hybrid deep neural networks for sparse surface EMG-Based hand gesture recognition," in 2020 54th Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA: IEEE, Nov. 2020, pp. 371–374. doi: 10.1109/IEEECONF51394.2020.9443400.

[15] M. E. Benalcazar, J. Gonzalez, A. Jaramillo-Yanez, C. E. Anchundia, P. Zambrano, and M. Segura, "A model for real-time hand gesture recognition using electromyography (EMG), covariances and feed-forward artificial neural networks," in 2020 IEEE ANDESCON, Quito, Ecuador: IEEE, Oct. 2020, pp. 1–6. doi: 10.1109/ANDESCON50619.2020.9271979.

[16] J. Qi, G. Jiang, G. Li, Y. Sun, and B. Tao, "Surface EMG hand gesture recognition system based on PCA and GRNN," Neural Comput. Appl., vol. 32, no. 10, pp. 6343–6351, May 2020, doi: 10.1007/s00521-019-04142-8.

[17] A. R. Asif et al., "Performance evaluation of convolutional neural network for hand gesture recognition using EMG," Sensors, vol. 20, no. 6, p. 1642, Mar. 2020, doi: 10.3390/s20061642.

[18] J. Chen, S. Bi, G. Zhang, and G. Cao, "High-Density surface EMG-Based gesture recognition using a 3D convolutional neural network," Sensors, vol. 20, no. 4, p. 1201, Feb. 2020, doi: 10.3390/s20041201.

[19] V. Shanmuganathan, H. R. Yesudhas, M. S. Khan, M. Khari, and A. H. Gandomi, "R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals," Neural Comput. Appl., vol. 32, no. 21, pp. 16723–16736, Nov. 2020, doi: 10.1007/s00521-020-05349-w.

[20] C. Andronache, M. Negru, A. Neacsu, G. Cioroiu, A. Radoi, and C. Burileanu, "Towards extending real-time EMG-based gesture recognition system," in 2020 43rd International Conference on Telecommunications and Signal Processing (TSP), Milan, Italy: IEEE, Jul. 2020, pp. 301–304. doi: 10.1109/TSP49548.2020.9163481.

[21] E. Rahimian, S. Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi, "Surface EMG-Based hand gesture recognition via hybrid and dilated deep neural network architectures for neurorobotic prostheses," J. Med. Robot. Res., vol. 05, no.

01n02, p. 2041001, Mar. 2020, doi: 10.1142/S2424905X20410019.

[22] M. Zanghieri, S. Benatti, A. Burrello, V. Kartsch, F. Conti, and L. Benini, "Robust real-time embedded EMG recognition framework using temporal convolutional networks on a multicore IoT processor," IEEE Trans. Biomed. Circuits Syst., vol. 14, no. 2, pp. 244–256, Apr. 2020, doi: 10.1109/TBCAS.2019.2959160.

[23] M. Simão, P. Neto, and O. Gibaru, "EMG-based online classification of gestures with recurrent neural networks," Pattern Recognit. Lett., vol. 128, pp. 45–51, Dec. 2019, doi: 10.1016/j.patrec.2019.07.021.

[24] E. A. Chung and M. E. Benalcazar, "Real-Time hand gesture recognition model using deep learning techniques and EMG signals," in 2019 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain: IEEE, Sep. 2019, pp. 1–5. doi: 10.23919/EUSIPCO.2019.8903136.

[25] A. A. Neacsu, G. Cioroiu, A. Radoi, and C. Burileanu, "Automatic EMG-based hand gesture recognition system using time-domain descriptors and fully-connected neural networks," in 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary: IEEE, Jul. 2019, pp. 232–235. doi: 10.1109/TSP.2019.8768831.

[26] J. O. Pinzon-Arenas, R. Jimenez-Moreno, and J. E. Herrera-Benavides, "Convolutional neural network for hand gesture recognition using 8 different EMG signals," in 2019 XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), Bucaramanga, Colombia: IEEE, Apr. 2019, pp. 1–5. doi: 10.1109/STSIVA.2019.8730272.

[27] Z. Zhang, K. Yang, J. Qian, and L. Zhang, "Real-Time surface EMG pattern recognition for hand gestures based on an artificial neural network," Sensors, vol. 19, no. 14, p. 3170, Jul. 2019, doi: 10.3390/s19143170.

[28] H.-Y. Chung, Y.-L. Chung, and W.-F. Tsai, "An efficient hand gesture recognition system based on deep CNN," in 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia: IEEE, Feb. 2019, pp. 853–858. doi: 10.1109/ICIT.2019.8755038.

[29] N. Alnaim, M. Abbod, and A. Albar, "Hand gesture recognition using convolutional neural network for people who have experienced a stroke," in 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey: IEEE, Oct. 2019, pp. 1–6. doi: 10.1109/ISMSIT.2019.8932739.

[30] M. E. Benalcazar, C. E. Anchundia, J. A. Zea, P. Zambrano, A. G. Jaramillo, and M. Segura, "Real-Time hand gesture recognition based on artificial feed-forward neural networks and EMG," in 2018 26th European Signal Processing Conference (EUSIPCO), Rome: IEEE, Sep. 2018, pp. 1492–1496. doi: 10.23919/EUSIPCO.2018.8553126.

[31] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," PLOS ONE, vol. 13, no. 10, p. e0206049, Oct. 2018, doi: 10.1371/journal.pone.0206049.

[32] Z. Ding, C. Yang, Z. Tian, C. Yi, Y. Fu, and F. Jiang, "SEMG-Based gesture recognition with convolution neural networks," Sustainability, vol. 10, no. 6, p. 1865, Jun. 2018, doi: 10.3390/su10061865.

[33] K. Xing et al., "Hand gesture recognition based on deep learning method," in 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, China: IEEE, Jun. 2018, pp. 542–546. doi: 10.1109/DSC.2018.00087.

[34] Y. Wu, B. Zheng, and Y. Zhao, "Dynamic gesture recognition based on LSTM-CNN," in 2018 Chinese Automation Congress (CAC), Xi'an, China: IEEE, Nov. 2018, pp. 2446–2450. doi: 10.1109/CAC.2018.8623035.

[35] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Deep learning in EMG-based gesture recognition:," in Proceedings of the 5th International Conference on Physiological Computing Systems, Seville, Spain: SCITEPRESS - Science and Technology Publications, 2018, pp. 107–114. doi: 10.5220/0006960201070114.

[36] Y. Du, W. Jin, W. Wei, Y. Hu, and W. Geng, "Surface EMG-Based inter-session gesture recognition enhanced by deep domain adaptation," Sensors, vol. 17, no. 3, p. 458, Feb. 2017, doi: 10.3390/s17030458.

[37] U. Cote-Allard, C. L. Fall, A. Campeau-Lecours, C. Gosselin, F. Laviolette, and B. Gosselin, “Transfer learning for sEMG hand gestures recognition using convolutional neural networks,” in 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB: IEEE, Oct. 2017, pp. 1663–1668. doi: 10.1109/SMC.2017.8122854.

[38] Y. Ma et al., “Hand gesture recognition with convolutional neural networks for the multimodal UAV control,” in 2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS), Linkoping: IEEE, Oct. 2017, pp. 198–203. doi: 10.1109/RED-UAS.2017.8101666.

[39] K. Asai and N. Takase, “Finger motion estimation based on frequency conversion of EMG signals and image recognition using convolutional neural network,” in 2017 17th International Conference on Control, Automation and Systems (ICCAS), Jeju: IEEE, Oct. 2017, pp. 1366–1371. doi: 10.23919/ICCAS.2017.8204206.

[40] X. Yingxin, L. Jinghua, W. Lichun, and K. Dehui, “A robust hand gesture recognition method via convolutional neural network,” in 2016 6th International Conference on Digital Home (ICDH), Guangzhou, China: IEEE, Dec. 2016, pp. 64–67. doi: 10.1109/ICDH.2016.023.

[41] H.-I. Lin, M.-H. Hsu, and W.-K. Chen, “Human hand gesture recognition using a convolution neural network,” in 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei: IEEE, Aug. 2014, pp. 1038–1043. doi: 10.1109/CoASE.2014.6899454.

[42] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia, “BLSTM-RNN based 3D gesture classification,” in Artificial Neural Networks and Machine Learning – ICANN 2013, vol. 8131, V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. P. Villa, B. Appollini, and N. Kasabov, Eds., in Lecture Notes in Computer Science, vol. 8131. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 381–388. doi: 10.1007/978-3-642-40728-4_48.

[43] J. Nagi et al., “Max-pooling convolutional neural networks for vision-based hand gesture recognition,” in 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia: IEEE, Nov. 2011, pp. 342–347. doi: 10.1109/ICSIPA.2011.6144164.

[44] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. Faughnan, “Smart surveillance as an edge network service: from harr-cascade, SVM to a lightweight CNN,” in 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA: IEEE, Oct. 2018, pp. 256–265. doi: 10.1109/CIC.2018.00042.

[45] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, “Vision-based hand-gesture applications,” Commun. ACM, vol. 54, no. 2, pp. 60–71, Feb. 2011, doi: 10.1145/1897816.1897838.

[46] L. Gallo, A. P. Placitelli, and M. Ciampi, “Controller-free exploration of medical image data: experiencing the Kinect,” in 2011 24th International Symposium on Computer-Based Medical Systems (CBMS), Bristol, UK: IEEE, Jun. 2011, pp. 1–6. doi: 10.1109/CBMS.2011.5999138.

[47] J. Zeng, Y. Sun, and F. Wang, “A natural hand gesture system for intelligent human-computer interaction and medical assistance,” in 2012 Third Global Congress on Intelligent Systems, Wuhan, China: IEEE, Nov. 2012, pp. 382–385. doi: 10.1109/GCIS.2012.60.

[48] J. R. Pansare, S. H. Gawande, and M. Ingle, “Real-Time static hand gesture recognition for American sign language (ASL) in complex background,” J. Signal Inf. Process., vol. 03, no. 03, pp. 364–367, 2012, doi: 10.4236/jsip.2012.33047.

[49] M. Karabasi, Z. Bhatti, and A. Shah, “A model for real-time recognition and textual representation of Malaysian sign language through image processing,” in 2013 International Conference on Advanced Computer Science Applications and Technologies, Kuching, Malaysia: IEEE, Dec. 2013, pp. 195–200. doi: 10.1109/ACSAT.2013.46.

[50] J. T. Belter, J. L. Segil, A. M. Dollar, and R. F. Weir, “Mechanical design and performance specifications of anthropomorphic prosthetic hands: a review,” J. Rehabil. Res. Dev., vol. 50, no. 5, p. 599, 2013, doi: 10.1682/JRRD.2011.10.0188.

[51] X. Zhao, A. M. Naguib, and S. Lee, "Kinect based calling gesture recognition for taking order service of elderly care robot," in The 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK: IEEE, Aug. 2014, pp. 525–530. doi: 10.1109/ROMAN.2014.6926306.
[52] M. Van Den Bergh et al., "Real-time 3D hand gesture interaction with a robot for understanding directions from humans," in 2011 RO-MAN, Atlanta, GA, USA: IEEE, Jul. 2011, pp. 357–362. doi: 10.1109/ROMAN.2011.6005195.
[53] J. Taylor et al., "Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences," ACM Trans. Graph., vol. 35, no. 4, pp. 1–12, Jul. 2016, doi: 10.1145/2897824.2925965.
[54] N. Modi, P. Verma, and B. Trivedi, Eds., Proceedings of International Conference on Communication and Networks: ComNet 2016, vol. 508. in Advances in Intelligent Systems and Computing, vol. 508. Singapore: Springer Singapore, 2017. doi: 10.1007/978-981-10-2750-5.
[55] R. Suriya and V. Vijayachamundeeswari, "A survey on hand gesture recognition for simple mouse control," in International Conference on Information Communication and Embedded Systems (ICICES2014), Chennai, India: IEEE, Feb. 2014, pp. 1–5. doi: 10.1109/ICICES.2014.7033762.
[56] M. Oudah, A. Al-Naji, and J. Chahl, "Hand gesture recognition based on computer vision: a review of techniques," J. Imaging, vol. 6, no. 8, p. 73, Jul. 2020, doi: 10.3390/jimaging6080073.
[57] H. Kaur and J. Rani, "A review: study of various techniques of hand gesture recognition," in 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Delhi, India: IEEE, Jul. 2016, pp. 1–5. doi: 10.1109/ICPEICES.2016.7853514.