A Hybrid Deep Learning and Optimization Approach for Accurate Channel Estimation in 5G MIMO-OFDM Systems

Mohammed Fakhreldin
Department of Computer Science-Faculty of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia

***Abstract*—Channel estimation plays a pivotal role in enhancing the reliability and efficiency of 5G wireless communication systems, particularly in MIMO-OFDM (Multiple Input Multiple Output - Orthogonal Frequency Division Multiplexing) architectures under multipath and Doppler-affected conditions. Conventional methods such as Least Squares (LS) are widely used due to their low computational complexity and lack of requirement for prior channel statistics. However, these approaches often result in poor estimation accuracy, especially in dynamic environments. To overcome these limitations, this study introduces a hybrid deep learning-based channel estimation framework that integrates Harris Hawks Optimization (HHO), Sparrow Search Algorithm (SSA), and Long Short-Term Memory (LSTM) networks—referred to as HHO-SSA-LSTM. The proposed method is designed to optimize the LSTM parameters using HHO and SSA, enhancing learning efficiency and estimation accuracy. Additionally, the model employs hybrid pre-coding aligned with codebook modeling strategies to preserve angle characteristics without disrupting azimuthal distributions. The system is evaluated in a 5G MIMO-OFDM setting under realistic conditions simulated using Doppler frequency and multipath propagation. Performance is assessed using key metrics including Bit Error Rate (BER), Mean Square Error (MSE), Symbol Error Rate (SER), efficiency, and execution time across different Pilot Lengths (PL = 128, 136, and 160). Simulation results demonstrate that the HHO-SSA-LSTM framework outperforms LS, LMMSE (Linear Minimum Mean Square Error), CNN (Convolutional Neural Network), FDNN (Forest Deep Neural Network), and standalone LSTM models. Notably, at PL = 160, BER is reduced by up to 91% and MSE by 86%, with an efficiency improvement exceeding 12% compared to traditional methods. Although the model exhibits a slightly higher execution time due to its hybrid design, the substantial accuracy gains justify the trade-off. The findings validate the effectiveness of the proposed hybrid model for robust and efficient channel estimation in 5G networks.**

Keywords*—Multiple-input multiple-output; channel estimation; orthogonal frequency division multiplexing; long short-term memory; pilot length

I. INTRODUCTION

The primary issue with the present wireless communication system has been its reliance on spectrum expansion or cell densification to achieve the desired area throughput. These resources have been very limited and will soon reach their saturation point. Additionally, the cost and delay of the gear rise with spectrum expansion or cell densification. The spectral effectiveness, which could increase the throughput in such a space, has largely remained unaltered throughout the rapid evolution of wireless networks. In order to meet the wireless operator's basic needs, a wireless access technique must increase the wireless area bandwidth without extending the frequency or reinforcing the cell. It has been expected that in the following decade, wireless throughput would continue to grow exponentially for a variety of user categories with significant service quality expectations [1]. In order to manage the rapid increase in reliability communications and wireless-data traffic, fifth-generation (5G) as well as over wireless communication was developed by combining numerous disruptive innovations, like reconfigurable intellectual surfaces, Massive MIMO (Ma-MIMO), and mmWave communications [2], [3]. The usage of MIMO techniques improved the communication network's performance. Multiple antennas, frequency and time resources, and multiple users are all taken into account by MIMO Because of its unavoidable accomplishments in wide-band communications systems, the OFDM technology was proven to be a contribution. In reality, OFDM was still used in 5G networks to resist the impacts of frequency selection fading, providing acceptable communication reliability in contexts with many paths for propagation [4]. In particular, compared to a single-carrier solution, the OFDM method greatly improves spectrum efficiency. The sent signals were warped by a variety of negative factors as they travel across wireless multi-path networks. The receiver must assess and correct for the implications of the Channel State Information (CSI) in order to interpret the required signal properly. Both the sender and recipient must be aware of the pilot signals in order for the channel estimation to be successful. Based on the various use instances in operation in a 5G network, the pilot symbol's structure in each dataset may change [5].

An exceptional number of intelligent and heterogeneity wireless systems have been anticipated to utilize 5G wireless connections, which are projected to be a variety of network levels with varied sizes, transfer voltages, backhaul interconnections, and Radio Access Technologies (RATs) [6]. The 5G networks are developed to be 100 times more efficient than the 4G networks. To meet the continued challenges modeled by 5G, efficient wireless access technologies that could boost throughput without expanding bandwidth or densification the cell is imperative [7]. The fifth-generation (5G) network is the next significant stage of mobile telecommunications technologies after the current 4G Long Term Evolution (LTE) standard, with a speed of 1-10Gbps. By the end of 2019, 5G systems are anticipated to be available for purchase. Within the most recent smartphone operating systems, 5G technologies

offer remarkable data capacities and infinite data broadcasting. Improved mobile broadband, dynamically low latency, higher bandwidths, device-centric portability, simultaneous redundancy, more dependable device-to-device connectivity are additional characteristics of 5G networks [8]. 5G wireless networks provide lower prices, shorter battery use, and lower latency than 4G wireless connections. Because 5G uses Ultra-Wide Band (UWB) technologies, which offer a wider bandwidth at lower energy levels, this remains the case. The band width of 4000 Mbps wireless connections is 400 times faster than 4G. Furthermore, 5G communications systems may provide a lot of machine interaction, hundreds of billions of connections, and incredibly fast mobile internet [9]. Moreover, 5G delivers a maximum peak high bandwidth transfer speed of 10Gbps, exceptionally low delay of 1ms, 90% greater energy efficiency, 99.9% ultra-reliability, and mobile data quantities of 10Tb [10]. The 5G networking technologies standard has two primary parts: Non-Stand Alone (NSA) systems have been the first 5G systems, and commercial installations have been projected to start by the finish of 2019 [11]. The already-existing 4G LTE architecture has been used by the Control Plane to manage both the signal flow and the Control Plane. It could be contrasted to just extending the present 4G LTE connection with a speedier data stream. Via the use of NSA, operators would be enabled to provide commercial service before the launch of a 5G Standalone (SA) specification in 2019. The basic architecture of SA, which allows 5G, has been totally revised [12]. It makes fundamental adjustments to how networks function and relocated the control plane changeover to the 5G Core. 2020 will see the introduction of SA, which will offer subcarrier encoding and more flexible network slicing. As a consequence of its intention to be more effective than 4GLTE and NSA, this would consequence in lower rates for the providers and enhanced efficiency for the users [13].

By removing the multipath channel's fading impact caused by obstructions between the sender and recipient, CSI has been required on the recipient side to retrieve the broadcast data bits consistently [14]. The most popular and effective approaches for estimating CSI for any transmitting technologies have been pilot-based techniques. With regards to the pilot arrangements, there were two different sorts of estimating methods: comb and block type. Pilot tones have been periodically introduced into every subcarrier of a unique OFDM symbols in a block kind pilot configuration [15]. Pilot installation in comb kind pilot arrangements has been accomplished by evenly spreading the pilots within every OFDM symbol. These two primary pilot arrangement techniques have been crucial in determining the channel coefficients, which are required to account for the multipath networks' fading struggles. Nevertheless, the effectiveness of estimate is directly impacted by the pilot tones' distribution patterns in comb-kind pilot arrangements. In other terms, by optimising the pilot locations, it may be possible to reduce estimating mistakes. Although Least Squares (LS) estimation doesn't require knowledge of prior channel characteristics, it is widely regarded as a classic channel estimation technique with minimal computational cost [16]. Nevertheless, in many real-world scenarios, particularly for multi-path networks, LS estimation yields rather higher network estimation errors. Through reducing the average channel estimate mistakes, mean square error MSE estimation, an alternate approach, produces substantially higher channel estimation reliability than LS estimation. Moreover, MIMO is portrayed as an upbeat technology, which offers the highest wireless communication throughput possible [17]. Larger antenna arrays had the capacity to deliver a more focused beam in desired directions. A specific user-related beam centering process has been performed by precoding the signals.

Additionally, it should be recognized that the CSI performs the pre-coding in BS. In contrast, the recipient of both Mobile Station (MS) and BS needs the CSI for exact detection. Therefore, it is crucial for the most accurate assessment of CSI [18]. Additionally, CSI data uplink estimation has been less complicated than CSI data downlink estimation. Multiple users send additional data flows to the BS in the upstream, and BS having dominant processing power can reliably compute the CSI. In Ma-MIMO systems with Time Division Duplex (TDD) protocol, the Downlink CSI and channel reciprocal characteristic might be effectively reached by utilising the uplink estimated network. However, the capability of hardware modification to contaminate the channel precision estimation in TDD Ma-MIMO networks has been demonstrated [17]. Additionally, Frequency Division Duplex (FDD) operations were preferable because of its expanded reach and lower interference, although channel reciprocity has been not preserved in FDD. Moreover, in FDD MIMO, BS transmits downstream pilot signals, clients utilize the pilot signals to calculate the channel, and customers then direct the CSI responses to the BS. Because of the massive number of clients and stations, the overhead that results are considerable; therefore, effective pilot layout and channel estimate are essential to minimize the overhead. Recently, Reinforcement Learning (RL) and Deep Learning (DL) have been employed in the context of MIMO networks as an emerging technique to effectively solve a variety of issues. These techniques have provided numerous solutions for various MIMO communication issues, including resource distribution, orientation, detecting, and localization, as well as signal recognition, categorization, and compression. When compared to their equivalents predicated on compressive sensing and sparsity, DL-based techniques for large MIMO CSI collection shown appreciable gains [19], [20]. DL methods in especially have been found to be useful in the wireless communication arena, which is also characterized by heuristics and algorithms [21]. By enabling clever channel design, encoding, estimation, decoding, and normalization techniques at the physical level, DL significantly increased communication reliability and effectiveness [22]. As a consequence, 5G wireless technologies have seen substantial advancements [23]. Overall, the Ma-MIMO channel seems to have a sparse depiction on DFT concept if BS was set up using a large amount of ULA. Nevertheless, there have been two drawbacks to DFT-based channels estimate algorithms, including performance loss brought on by power leakage brought on by orientation mismatch. Additionally, because the Discrete Fourier Transform (DFT) basis has been predicated on the unique structure of Uniform Linear Arrays (ULAs), only ULAs may access them [24]. Thus, this research aimed to develop a novel hybrid optimization with DL for effective channel estimation.

A. Key Contributions

1) Proposed a novel hybrid channel estimation framework (HHO-SSA-LSTM) that combines HHO, Sparrow Search Algorithm, and Long Short-Term Memory networks for accurate channel estimation in 5G MIMO-OFDM systems.

2) Achieved significantly lower BER, SER, and MSE compared to traditional and deep learning-based estimation methods (LS, LMMSE, CNN, FDNN, and LSTM), especially under higher pilot lengths and Doppler effects.

3) Integrated a hybrid metaheuristic-based pre-coding strategy aligned with codebook modeling to enhance beamforming accuracy without conflicting with azimuth angle characteristics.

4) Demonstrated higher estimation efficiency (up to 96.92%) across different pilot lengths, showing better performance than baseline models even under limited pilot overhead.

5) Conducted detailed complexity and execution time analysis against recent methods (e.g., CS, OCEAN, PSO-Adam-LSTM), showing a justified trade-off between increased computation and improved estimation performance.

B. Rest of the Section

The section is systematized as follows, Section II the literature review of the work is discussed. Section III explains the proposed approach. The result and discussion of the study is given in Section IV. Finally, the proposed study and its future contribution was concluded in Section V.

II. Related Works

Xisuo Ma et al. [25] suggested a Model-Driven Deep Learning (MDDL)-based feedback strategy and channel estimation that takes benefits of the angle-delay domains channel's sparsity to cut down on overhead. The uplink-based channel estimation for TDD systems came first. Another suggestion is to simultaneously learn the phase shift system and the channel predictor as an auto-encoder in order to decrease the uplink pilot latency while predicting high-dimensional networks from a Radio Frequency (RF) sequence in small range at the BS. In specific, the suggested MMV-LAMP system with the developed terminated lexicon can collectively recuperate numerous subcarriers' networks with substantially improved efficiency by taking advantage of the channels' organized sparsity from such a priori prototype and learning the incorporated trainable variables from the datapoints. Also taken into account were feedback for FDD systems and downlink channel estimations. Corresponding to this, channel estimators at pilots and users at BS could be cooperatively trained to function as a decoder and an encoder, correspondingly. Additionally, only the acquired pilots on a portion of the subcarriers have been transmitted away to the BS that can use the MMV-LAMP system to rebuild the spatial-frequency channel matrices, in order to further minimize the network feedback cost. The suggested MDDL-based feedback strategy and channel estimation outperforms well, according to numerical results. However, the model was not tested on fixed-scattering environments.

Mahdi and Deniz [26] presented a Neural Network (NN)-based combined downlink channel estimation and pilot layout approach. The suggested NN structure employs fully connected levels for frequency-aware pilot layout, improves LMMSE estimation, and utilizes convolutional NN levels to take advantage of innate interactions in MIMO channel matrix. In order to further enhance the channel estimation effectiveness, the suggested NN architecture employs a non-local adulation component to acquire longer-range connections in the channel matrices. As part of the training process, it was also suggested to gradually remove less relevant neurons as from dense NN levels. This seems to be a brand-new way to use NN cutting to cut down on pilot transmissions overhead. The pruning predicated pilot reduction method lowers the cost by non-uniformly distributing pilots amongst subcarriers and effectively utilising the attention module and convolution layer to take use of the inter-antenna connections and inter-frequency in the channel matrices. However, the computational time is high.

Omar et al. [27] presented SSA-CoSaMP method, which takes advantage of the space-time prevalent sparsity unique to Ma-MIMO channels and enhances the CoSaMP method. Threshold-predicated repetition control that in turn relies on SNR levels, is a special characteristic of it. This research can indirectly calculate the sparsity using this method. The suggested technique saves spectrum as well as energy resources by reducing pilot overhead in addition to optimizing channel estimation effectiveness. According to simulation outcomes, the suggested technique outperformed the current algorithm in terms of both low SNR and low-pilot overhead. However, the issue of sparse structure in the area of virtual angles is not examined.

Lijun et al. [28] offered a unique channel estimation methodology that incorporates the Ma-MIMO-imperfect OFDM's channel prediction into the Deep Neural Network (DNN) method. Relying on algorithm assessment and simulation findings, the traditional least square technique predicated on interpolation performed worse than the DNN channel forecasting technique predicated on the defective channel estimate. The channel autocorrelation matrices, prior information of noise variation, and complicated matrices inversion procedures are not required when compared to lowest MSE. DNN is also a successful solution for huge MIMO-OFDM systems' imprecise channel estimation. However, the design complexity is high. Anughna and Ramesha [29] outline the MIMO OFDM system's fundamentals. Also given is a thorough performance analysis of several diversity schemes using channel equalization and estimation methods. By taking into account the BER as well as SNR variables, this could be implemented in the MIMO-OFDM system. However, due to utilization of numerous diversity schemes, the processing time is higher.

Lijun Ge et al. [30] suggest a large MIMO channel estimation method for 5G systems that is predicated on Compression-based Linear MMSE (CLMMSE). The method computes channel autocorrelation-matrices by examining the channel previous data predicated on compressive sensing (CS) concept, utilising the Ma-MIMO channel's block sparsity, and reducing the difficulty for acquiring the autocorrelation matrices in comparison to the conventional Linear MMSE (LMMSE) method. To further minimize the computational burden, it then

replaces single value decomposition for matrices inverse function. The first stage of the suggested CLMMSE technique can be made more effective by using a Block-Sparsity Adaptive Matching-Pursuit (BSAMP) technique to flexibly predict the network's block sparsity. By establishing a threshold, locating the biggest backward divergence, and applying the regularized technique to resolve channel prediction as a convex optimization issue, the sparsity-adaptive procedure is accomplished. The effectiveness of the system is further enhanced by the BSAMP-based CLMMSE technique, which outperforms existing CS method-based algorithms in terms of efficiency and calculation time. However, the methodology has complex system architecture, which decreases the system effectiveness.

Yanfeng Zhang et al. [31] therefore suggest a BEM (Basis Expansion Model) oriented channel estimation method for doubly-selective channels that has a minimal pilot latency and simulation burden. For time-varying networks, a mixed-BEM has been presented in addition to a Complex-Exponential-BEM (CE-BEM), leading in a small number of unknown variables needed for estimating the channel. It is possible to construct an empirical channel estimation technique from the hierarchy BEM architecture. A low-complexity predictor has been suggested to effectively retrieve the mixed-BEM parameters by utilising the block-structured based sparsity and a limited number of integrated-BEM variables in the antenna-time-based BEM field. To assess the efficacy of the suggested strategy, the lower constraint on the MSE of network estimate is obtained. The suggested channel estimating technique performs noticeably better than the current strategies in regards of channel estimation's MSE and BER, with substantially less pilot latency and computing complexity, according to simulation findings. However, the developed model has higher computational time.

Wei Ji et al. [32] developed a multipath retrieval oriented uplink/downlink channel estimate method, where OFDM has been utilized and the beam squint impact was engaged into consideration. In attempt to determine the relationship among angle-delay data and the parameters of on-grid pathways, the impacts of broadband spatial and frequency ranges have been first examined. Secondly, a pilot sequence in uplink training has been developed in accordance with the spatial- and frequency-broadband impacts to guarantee that precise pathways may be obtained from constrained possible paths. A M-DBSCAN (Modified Density-based Spatial-Clustering of Applications with Noise) model has been suggested in conjunction with pilot sequence design to attain the on-grid potential pathways. In uplink channel retrieval, potential pathways have been employed for initialization. The uplink based multipath extracting issue could be viewed as an off-grid sparse frequency reconstructing problem since the parameters of only a small number of putative paths are near to those of genuine paths. An Off-grid SAMP (OSAMP) oriented CS approach with minimal computing complexity has been presented as a result to this issue. In addition to precisely estimating angle-delay data and path benefits for exact pathways, OSAMP also takes advantage of the angle-delay congruence among downlink and uplink networks and uses uplink multipath retrieval to predict the downlink network. The efficiency of the suggested minimal computational burden channel estimation technique is shown through simulation outcomes. However, the model has higher processing time.

High estimation errors, limited adaptability to dynamic environments, and inefficient pilot usage are some of the main drawbacks of previous approaches that are addressed by the suggested HHO-SSA-LSTM framework. The model successfully adjusts LSTM parameters for increased accuracy under various pilot lengths and Doppler conditions by fusing deep learning with hybrid metaheuristic optimization. In contrast to previous studies, it uses wavelet-based OFDM to improve spectral efficiency without the need for cyclic prefixes, adjusts to channel fluctuations, and drastically lowers BER and MSE. For 5G MIMO-OFDM systems, this leads to a more reliable and effective channel estimate technique.

III. Proposed Methodology for Hybrid Channel Estimation

A. System Model

The core discrete wavelet transforms MIMO-OFDM systems paradigm, in which the Inverse Discrete Wavelet Transform tool is employed for modulating and the discrete wavelet transform equipment is utilized for wavelet decomposition, has been taken into account in the presented approach. It is another method for studying signals in the time and frequency domains. Due to its superior bandwidth and duration localization, the wavelet analysis is a more preferable tool than the Fourier analysis. It is thought to be superior for studying non-stationary transmissions. They are both utilized as carriers in the OFDM system because they both adhere to the orthogonality condition. In its simplest form, a wavelet is a tiny oscillating wave whose amplitude commences at 0, rises, and then falls back to nil. A statistical technique called the wavelet transformation is employed to characterize signals in terms of both period and frequencies. The analysis of the signal's harmonic module's timeframe is helpful. Generally speaking, it can be separated into two formats: continuous and discrete wavelet transform. Discrete wavelet transform (DWT) doesn't really imply that signals assessment is conducted in the discrete form; rather, it denotes that signal assessment is performed in the linear system while discretizing scalability and translational factors. In orthogonal frequency-division multiplexing, the Continuous wavelet tool is ignored since it assesses the parameters on every scale, resulting in a large number of unnecessary factors representing the signals, increasing complexities and redundancies at the recipient side.

The wavelet form functions are defined as follows:

$$D(v,\delta) = \left(\left(\frac{1}{\sqrt{v}}\right)\right)\varphi\left(\frac{(e-\delta)}{v}\right) \quad (1)$$

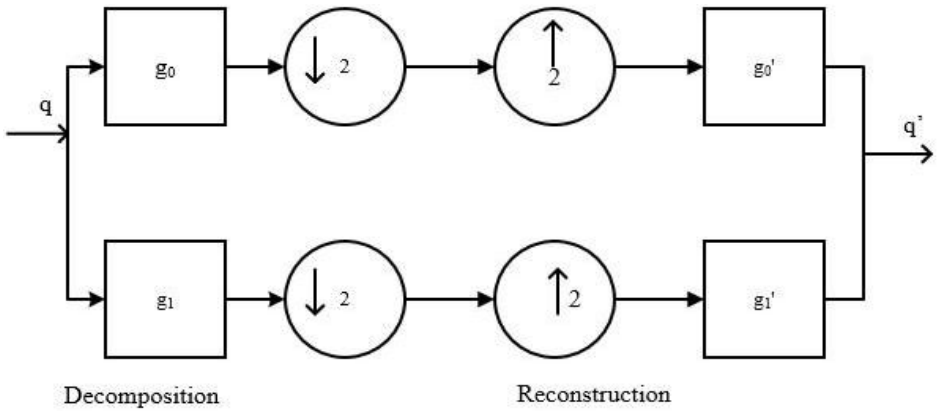


Fig. 1. Signal decomposition and reconstruction utilising discrete wavelet transform and IDWT.

It enlarges, transversely translations, and scales the wavelet transformation $\varphi(e)$. Here, v is a scalability factor that establishes the wavelet's amplitude, and δ is the length of the wavelet was determined by the translational variable. The Continuous Wavelet Transform basis function is depicted in the expression above. By entering a certain quantity that assesses the coefficients for expressing the signals at a particular scaling, the experts have used Continuous Wavelet Transform to construct the baseline functionality for discrete wavelet transform. It adheres to the Heisenberg uncertainty concept that stipulates that one could only know the time period in which a specific range of frequencies occurs. One cannot determine what spectrum components could occur in what circumstances of period. In order to examine signals at various scales, discrete wavelet transform primarily decomposes the evidence into a collection of wavelet basis functions that are perpendicular to one another. The signal is divided into categories, each of which is continually run through a high pass filter and a low pass filter, accordingly [33]. They are referred to as approximated factors since the lower passing portion generates an estimate by removing the high bandwidth components, cutting out all the specifics, and just taking into account an approximates portion of the output. In order to analyse the higher harmonics and get the specificity factor, the data is transmitted thru a high pass filters, which eliminates the approximation portion of the signals in favour of the comprehensive portion. Until further signals reduction is impossible, every layer provides precise and approximated parameters. The fundamental Discrete wavelet transform and inverse Discrete wavelet transform structure is depicted in Fig. 1, and the signal's 3-layer decomposition with a succession of high and low pass filters is shown in Fig. 2.

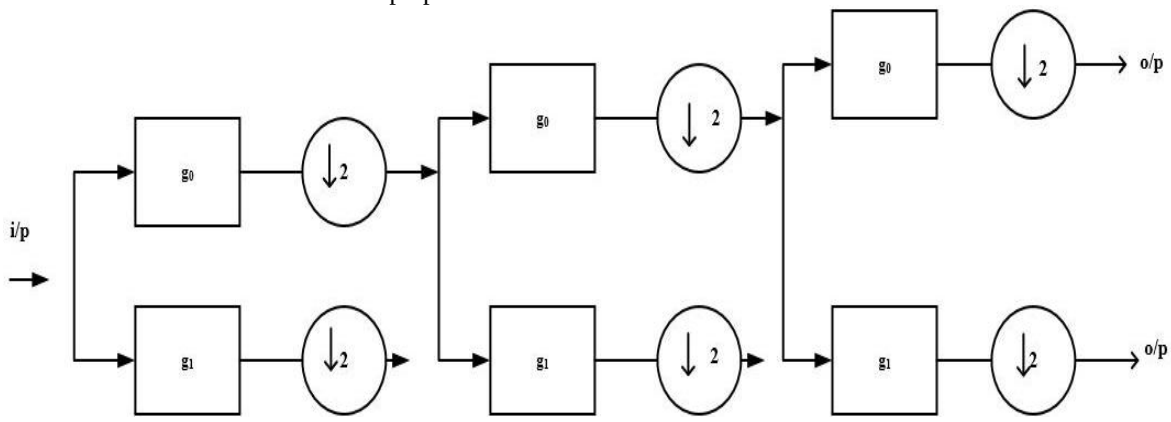


Fig. 2. Workflow of discrete wavelet transform.

Here is the expression for the Inverse Discrete Wavelet Transform of $q(z)$ is given by,

$$q(z) = \sum_{u=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} Q(u,l) 2^{\frac{u}{2}} \varphi\,(2^u z - l) \qquad (2)$$

The following expression represents discrete wavelet transform of $q(z)$ is given by

$$Q(u,l) = \sum_z q\,(z) 2^{\frac{u}{2}} \varphi\,(2^u z - l) \qquad (3)$$

Up-sampling and down-sampling techniques alter the dimension. It scales back a portion of the signal's sampling. Here u Stands for the translational scaling, and $\varphi(u,l)$ is the wavelet function component. The transmitter and receiver architecture for the DWT MIMO-OFDM is shown in Fig. 3. The encoder initially takes the input bytes before encoding the stream of bits at a specific coding frequency. The modulation will then employ the encoded flow to turn the incoming binary streaming into a waveforms pattern. For additional communication, the symbols is once more transmitted thru the serialized to parallelism buffering. Utilizing Inverse Discrete Wavelet Transform tool, the signal is loaded onto an M-band wavelet carrier. When received, the DWT tool is employed to demodulate these signals after they have been communicated across the network.

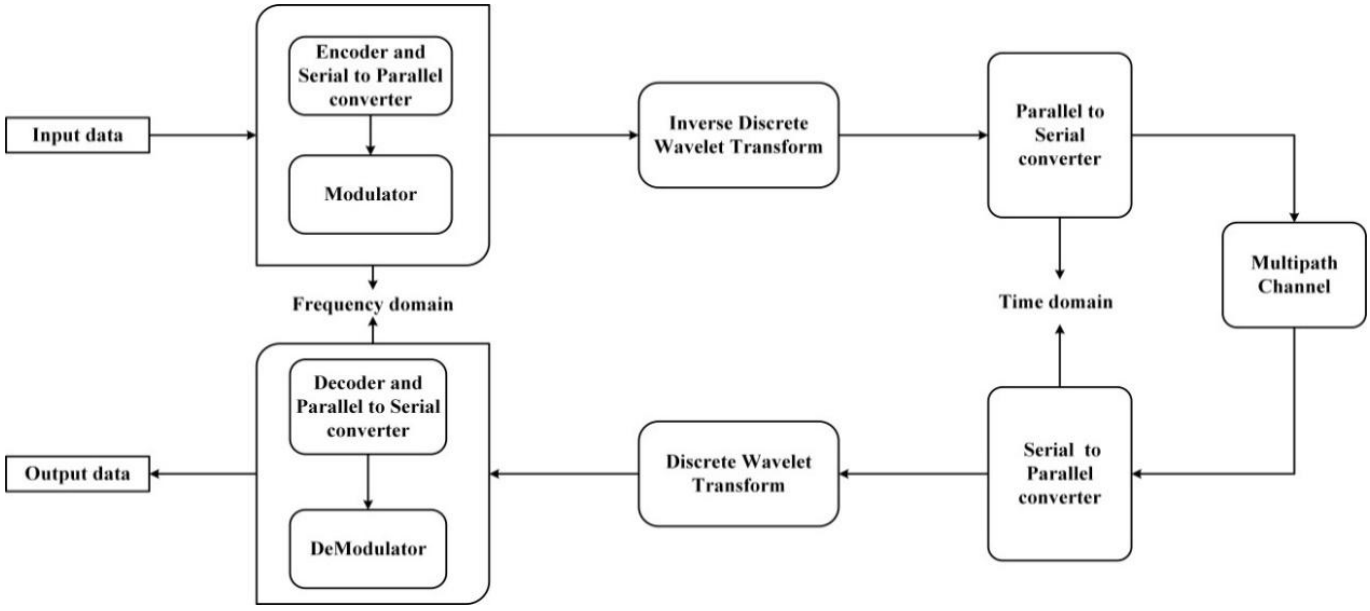


Fig. 3. Framework for an OFDM transmitter system based on DWT.

B. Cyclic Prefix

Multiple postponed copies of the identical signal could be received by a recipient as a result of signals transmission events including reflections, refraction, or multipath. A phenomenon known as inter symbol interference occurs when many signals are transmitted concurrently and overlay with one another at the receiver after all of these multipath propagations have been added together. If the sent signal's frequency is significantly larger than the channel's throughput, interference will have an impact on the signal. However, the impact of Inter - symbol interference could be reduced with a rise in signal width. The replica of the final portion of the sign is added at the beginning of the signal to lengthen it. Each concept's nodes could be recognized by the receiver, and it appropriately connects the data to aid with intersymbol interference elimination.

M^{dw} indicates the cyclical prefix's duration, M^{dwt} represents the size of the Inverse Fast-Fourier transform frame, and M demonstrate the amount of subordinate carriers. Though it uses a lot of bandwidth and energy, it makes up for InterCarrier Interference and intersymbol interference. Add a feedback path with a duration z higher than the channels impulsive responder u, presuming the network has a maximal latency of u. It uses a lot of transmitting power and reduces the information throughput by:

$$C_P(z) \times 100\% \Big/ C_P(z) + No.of\ subcarriers(Z) \quad (4)$$

Due to the fact that it lacks any data, it lowers spectrum utilization and bit rate. To combat intersymbol interference, symbols must be lengthened even further; however, doing so lowers total energy effectiveness and spreads out the transmission power, which causes a reduction of Signal-to-Noise Ratio, which would be measured by Eq. (5).

$$-10\log_{10}\frac{Overall\ OFDM\ symbol\ length - C_pLength}{Overall\ OFDM\ symbol\ length} \quad (5)$$

Since wavelet transforms are utilized, which have lengthier waveforms than standard transforms, discrete wavelet transform -OFDM signals overlap in period, which improves frequencies localization. As a consequence, extending the signal period to reduce intersymbol interference does not need to use a cyclic prefix. Additionally, whereas discrete wavelet transform-OFDM avoids this, Fourier transform signals pass via a rectangular waveguide, resulting in wider side lobes. Additionally, it lowers the system's difficulty from fast Fourier transforms $A\,(Zlog_2 Z)\ to\ A\,(Z)$. In addition to saving frequency by just not inserting Cyclic prefix to the Orthogonal Frequency Division multiplexing system signal, DWT-OFDM also reduces transmitting energy usage and increases system capacity [34].

1) Hybrid Pre-coding: It provides a quick overview of Fourier transforms and the CP decomposed to keep the article self-contained. Users can get additional information on the annotation and the fundamentals of Fourier transforms. A vector sometimes referred to as a manner or modal, is just a generalization of matrices to higher-order dimensions. It is possible to think of Fourier transforms including one phase and 2 modes as being represented by both vectors and matrices, accordingly. The Kronecker, outer, and Khatri-Rao products are each indicated in °,⊙,⊗ this article by the characters, and, correspondingly.

Let $Y \in D_{l1 \times l2 \times \ldots \times lu}$ indicate a vector of u−th rank with its $(l_1, l_2, \ldots l_u)$th admittance represented by$Y_{l1}, \ldots l_n$. The amount of dimensionality is represented by a tensor's order u in this case. The column and row of matrices are analogous to fibres at a higher form. By setting all indexes other than in, the $mode - u$ fibres of Y are produced as In-dimensional matrices. A vector could be divided into 2-dimensional segments by setting all but 2 of its indexes. The process that converts a vector into matrices is known as unfolding or matricization. The column of the resultant matrices is organized by the phase unfolding of a

vector$Y$, designated as $Y(n)$ and it is given in Eq. (6) and Eq. (7).

$$Y = \sum_{h=1}^{s} \mu_h d_h^{(1)} \circ d_h^{(2)} \circ \ldots \circ d_h^{(u)} \quad (6)$$

$$Y_{l1,l2,..lu} = \sum_{h=1}^{s} \mu_h d_{hl1}^{(1)} \circ d_{hl2}^{(2)} \circ \ldots \circ d_{hlu}^{(u)} \quad (7)$$

Following Eq. (8) is an expression for Y's mode-u unfolding:

$$Y_u = D_{(u)} \propto (D_{(u)} \odot \ldots . D_{(u+1)} \odot D_{(u-1)} \ldots . D_{(1)})^q \quad (8)$$

Where $\propto \cong diag(\mu_1, \ldots \mu_h)$. According to [35], the hybrid pre-coding method was implemented and the process is shown in algorithm.1. In the process, BS = Base Station, and MS = Mobile Station.

C. Hybrid Sparrow-Harris Hawks Optimization

The suggested HHO was based on a simulation of Harris hawks' foraging habits. The HHO replicates its cooperative foraging in two phases, i.e., exploration and exploitation phase—using a variety of strategies. Based on the parameter H, which represents the prey's escape energy, HHO carries out the change from the exploration to the exploitation phase. The mathematical formation is given in Eq. (9):

$$H = 2H_0(1 - i/I) \quad (9)$$

The early energy state of the prey is denoted as H_0, fluctuates at random among (1,1) for every iteration. The computation is $H_0 = 2 * ran - 1$ where I and i are the maximal and current iterations, and rand is the random integer of (0,1).

1) Searching phase: In this research, the searching phase is accomplished by the sparrow search optimization algorithm. The sparrow is highly clever and has a good memory, unlike numerous other small birds. Keep in mind that there are two different breeds of house sparrows which are kept as pets: the producer and the scrounger. While the scroungers rely on the producers to provide them with food, the producers constantly pursue for sources of food. The findings also demonstrate that the birds frequently transition between generating and scrounging behaviour patterns. Additionally, it may be noted that sparrows typically combine their producer and scrounger strategies to get food. The computational formula is developed to create the sparrow search method based on the earlier description of the sparrows. The study created equivalent rules based on the following idealized behaviour of the sparrows for convenience.

To search for food in the simulation study, virtual sparrows must be developed. The following Eq. (10) may be used to illustrate where sparrows are located:

$$A = \begin{bmatrix} A_{11} & A_{12} \ldots & A_{1m} \\ A_{21} & A_{22\ldots} & A_{2m} \\ A_{n1} & A_{n2} & A_{nm} \end{bmatrix} \quad (10)$$

Here, m is the dimensionality of the parameters that need to be optimized and n denotes the number of sparrows. The corresponding vector may therefore be used to indicate the fitness value of every sparrows as given in Eq. (11):

$$F_A = \begin{bmatrix} F([A_{11} & A_{12} \ldots & A_{1m}]) \\ F([A_{21} & A_{22\ldots} & A_{2m}]) \\ F([A_{n1} & A_{n2\ldots} & A_{nm}]) \end{bmatrix} \quad (11)$$

"Here, the amount of every row in F_A denotes the individual's fitness value and n is the number of sparrows. In the SSA, food is prioritised for producers with higher fitness ratings during the procedure of search. The producers are also responsible for charge of locating food and guiding the entire community's circulation. The producers may thus look for food in a wider variety of locations than the scavengers. The mathematic formulation is given in Eq. (12).

$$A_{i+1}^{xy} = \begin{cases} A_i^{xy} . exp\left(\frac{-t}{\beta . i_{max}}\right) \\ A_i^{xy} + R.M \end{cases} \begin{matrix} if\ D < S \\ if\ D \geq S \end{matrix} \quad (12)$$

Here, y = 1, 2,..., m, and i denotes the current iteration. The frequency of the y th dimensions of the ith sparrow at iteration i is represented by A_{xy}^{i}. The constant with the most iteration is called i_{max}. 0 and 1 are two random numbers. The alert value and the safety threshold are denoted by D and S respectively. R is a chance quantity that follows the normal distribution. M displays a $1 * m$ matrix with 1 as the value of each member. The producer switches to broad search mode at $D < S$, that indicates that there are no nearby predators. Many sparrows must immediately fly to other safe regions if $D \geq S$, which indicates that certain sparrows have identified the predator."

2) Exploitation phase: The procedure moves into the stage of development when$|H| < 1$. The prey that were followed and monitored during the exploration phase would be the subject of a raid hunt by the hawks throughout this stage. Nevertheless, when hunting in nature, the prey would often attempt to flee from the chase, hence hawks would also employ varied pursuing strategies for the various prey escape behaviour. HHO suggested four methods, which are all outlined below, to mimic this pursuing and hunting behaviour. HHO chooses a strategy based on the escape energy H and the probability of escape e.

a) Soft besiege: When $|H| \geq 0.5$ and ≥ 0.5, the prey had sufficient energy to make a random jump to safety, but hawks have encircled it. The hawk then decides to effectively hunt by using gentle besiege to deplete the prey's physical ability. Eq. (13) to Eq. (15) illustrate the computational equations.

$$A_{i+1}^{xy} = \Delta X_i^y - H * |J A_i^{p,y} - A_i^{xy}| \quad (13)$$

$$\Delta X_i^y = A_i^{p,y} - A_i^{xy} \quad (14)$$

$$J = 2(1 - e_5) \quad (15)$$

Here, e_5 is a random value between (0,1) and y is the distance between the prey's current location in the y th dimension as well as the current location of the xh hawk; where Jump (J) denotes a randomly changing jump intensity that alternates between (0,2) for each repetition.

b) Hard besiege: When the prey lacks the energy to flee and the hawks have encircled it, $|H| < 0.5$ and e <0.5, the hawks would decide to engage in a hard besiege swift raid hunt. Eq. (16) displays the behaviour model in question.

$$A_{i+1}^{xy} = A_i^{p,y} - H * |\Delta X_i^y| \quad (16)$$

c) Soft besiege with progressive rapid dives: The prey has had enough energy to break free of the siege as well as perform a zigzag motion in the situation of $|H| \geq 0.5$and $e < 0.5$, and the hawks haven't yet fully organized an surrounding denounce on the prey. The hawks now decide to concurrent using prey energy as well as eventually create a full surround. Eq. (17) to Eq. (20) illustrate the steps of this method.

$$A_{i+1}^{xy} \begin{cases} B_{i+1}^{xy} \; if \; f(B_{i+1}^{xy}) < f(B_{i+1}^{xy}) \\ C_{i+1}^{xy} if \; f(C_{i+1}^{xy}) < f(C_{i+1}^{xy}) \end{cases} \quad (17)$$

$$B_{i+1}^{xy} = A_i^{p,y} - H * |JA_i^{p,y} - A_i^{xy}| \quad (18)$$

$$C_{i+1}^{xy} = B_{i+1}^{xy} + T_y * L_y \quad (19)$$

$$L_y = L(r_y, s_y, \alpha_y) \quad (20)$$

Here T_y represents random number; L indicates Lévy flight function.

d) Hard besiege with progressive rapid dives: Hawks utilize this tactic to speed up and reduce the mean location distance among themselves and the prey in order to build a hard encircling circle earlier the raid in the situation of $|H| < 0.5$ and $e < 0.5$, when the prey lacks the energy to flee but is not entirely surrounded. Eq. (21) through Eq. (23) provide the models for this tactic.

$$\begin{cases} B_{i+1}^{xy} \; if \; f(B_{i+1}^{xy}) < f(B_{i+1}^{xy}) \\ C_{i+1}^{xy} if \; f(C_{i+1}^{xy}) < f(C_{i+1}^{xy}) \end{cases} \quad (21)$$

$$B_{i+1}^{xy} = A_i^{p,y} - H * |JA_i^{p,y} - A_i^{xy}| \quad (22)$$

$$C_{i+1}^{xy} = B_{i+1}^{xy} + T_y * L_y \quad (23)$$

The Sparrow Search Optimization (SSO) algorithm emulates the foraging and anti-predator actions of sparrows to solve optimization issues. In this algorithm, population members are separated into two primary roles: producers and scroungers. Producers are the sparrows with more energy, usually the ones that get assigned higher fitness values. They are responsible for actively searching the search space for regions with high resources (i.e., improved solutions). As soon as they find such a region, they alert other members of the group. In the event that there is danger (e.g., in the presence of a predator)—as expressed by a risk level above a particular threshold value—these producers lead the others to a safe area in the search space directly. Conversely, scroungers are less fit sparrows. Rather than searching on their own, they tail producers with the expectation of gaining access to good resources. Certain scroungers tactically monitor producers, and if they observe a producer discovering a high fitness solution, they move rapidly to that spot in a bid to upgrade their own position. If they succeed, they directly take advantage of the producer's solution; otherwise, they fall back to a rule-based updating mechanism. Population behavior demonstrates a balance natural to the system: while sparrows towards the center of the group meander randomly to preserve close proximity, those at the edges are more sensitive to threats and move quickly to less risky or more advantageous areas. The scrounger positions get updated according to certain mathematical laws in Eq. (24), that determine how they move compared to producers so as to gain opportunities for better solutions.

$$A_{i+1}^{xy} = \begin{cases} R.ep\left(\frac{A_i^{worst} - A_i^{xy}}{t^2}\right) \\ A_o^{i+1} + \left|A_i^{xy} - A_{i+1}^{xy}\right| \end{cases} \quad if \; T > n/2 \quad (24)$$

Here A_o denotes the ideal position. A_{worst} designates the world's worst location at the moment. Every member in the matrix R is given a random number between 1 and 1. The tth scrounger with the lowest fitness score is most certain to be starved when $T > n/2$.

These sparrows, which are aware of the threat, make about 10% to 20% of the overall population for the simulation study. Those sparrows were distributed at random in their starting places. The mathematical model may be depicted as follows using rules as given in Eq. (25):

$$A_{i+1}^{xy} = \begin{cases} A_i^{best}.\alpha\left|A_i^{xy} - A_i^{best}\right| & if \; f_t > f_g \\ A_{xy}^{i} + K.\left(\frac{A_i^{xy} - A_i^{worst}}{(f_t - f_g) + \varepsilon}\right) & if \; f_t = f_g \end{cases} \quad (25)$$

Here, the current global ideal location is A_i^{best}. With an average value and a variance of 0 and 1, respectively, the step size process parameters is a normal distribution of random integers. The random integer K is between [1, 1]. f_t is the current sparrow's fitness value in this instance. "The top and worst fitness values on the planet right now are f_g and f_t, respectively. To prevent a zero-division mistake, is the lowest constant.

For ease of use, when the sparrow was at the group's edge if $f_t > f_g$. A_i^{best} stands for where the population's centre is safe in that area. The sparrows inside the centre of the community are conscious of the threat and need to migrate closer to the others, as shown by the equation $f_t = f_g$. The step-size control coefficient and the sparrow's movement direction are both indicated by the letter K.

D. LSTM Based Channel Estimation

"The dispersion channels between the receiver and the transmitter in wireless communications networks must be known in order to perform coherent identification; these channels could be estimated by applying traditional estimation methods. This section introduces LSTM-based channel estimation strategies that encourage us to use DL techniques to reduce channel estimation errors. A recurrent neural network or LSTM neural network was used to examine the behaviours of the channel correlations (RNN). Fig. 4 shows the one-layer RNN's basic construction. This illustration shows that the RNN cell's input for the present time-step is its output for the previous time-step. By functioning in this way, the RNN has been able to recall previous input statistics. The computing unit, or fundamental RNN cell, executes the following computation in Eq. (26) and Eq. (27)."

$$d_t = f(R_{id}s_t + R_{dd}d_{t-1} + g_{id} + g_{dd}) \quad (26)$$

$$U_t = f(g_{do} + d_t R_{do}) \quad (27)$$

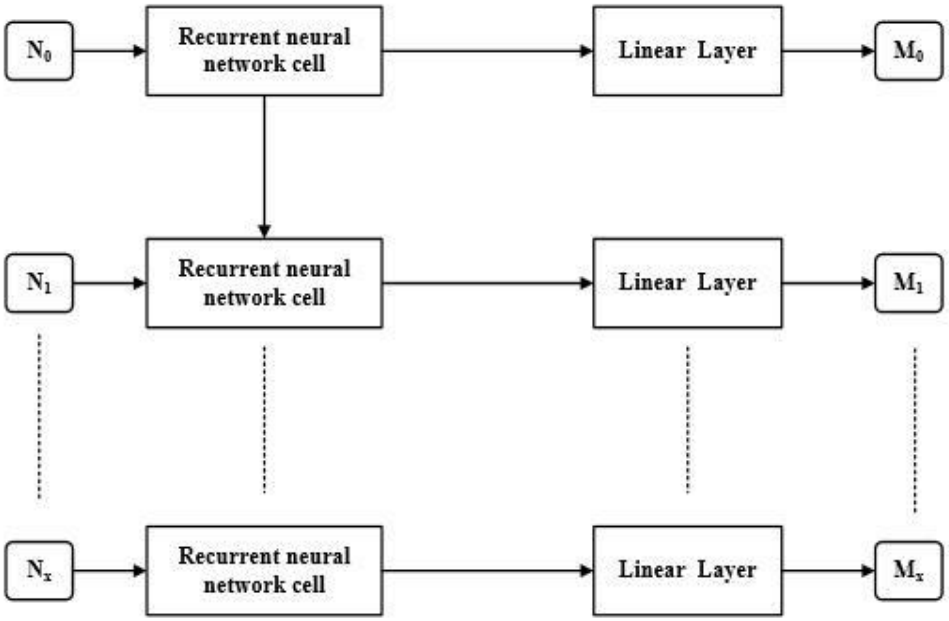


Fig. 4. Basic Structure of RNN layer.

where $f(\cdot)$ represents activation function; s_t is the input and U_t is the output; $d_t$and d_{t-1} are the hidden states at the time step t and $t-1$, correspondingly, R_{id}, R_{dd}, and R_{do} are the weights for input-hidden layer; hidden-next hidden layer, and hidden-output layer, respectively; and g_{id}, g_{dd}, and g_{do} were the corresponding biases.

"However, the basic RNN cell has some shortcomings. For starters, even if the channel at time step t has a connection to both the past and the future, it is unable to exploit the data's future information. In this case, the bidirectional network must be used to achieve higher efficiency. Second, a disadvantage of a simple RNN cell is that it cannot record long-term data. One solution to this problem is to use LSTM as an alternative. In this study, we propose a bidirectional LSTM (bi-LSTM) network for 5G channel estimation in order to overcome these problems."

The outcome of the LSTM cell's computation is indicated in the following formulae from Eq. (28) to Eq. (33).

$$f_t = f\left(V_f S_t + g_f + R_f d_{t-1}\right) \quad (28)$$

$$j_t = f\left(V_j S_t + g_j + R_j d_{t-1}\right) \quad (29)$$

$$b'_t = tand(V_b S_t + g_b + R_b d_{t-1}) \quad (30)$$

$$b_t = f_t \odot b_{t-1} + j_t \odot b'_t \quad (31)$$

$$p_t = f(V_o S_t + g_o + R_o d_{t-1}) \quad (32)$$

$$d_t = o_t \odot tand(b_t) \quad (33)$$

Where, R_f, R_j, R_b, R_o, V_f, V_j, V_b, V_o, g_f, g_j, g_b, and g_o are the respective weights of matrices and biases, as well as $tand$ is the hyperbolic tangent functional. The cell state b_t holds the critical data from the past, the forget function f_t specifies which statistics would be elapsed by the LSTM cell, and the new candidate values b'_t specifies which data would be upgraded to the cell state. The LSTM cell's hidden state function is represented by b_t and d_t. The LSTM cell has a stronger capacity to capture information than a basic RNN cell since it may avoid duplicate information by operating in this manner, allowing it to catch the vital data from the past. The bi-LSTM network's structure is depicted at the lowest part of Fig. 8. Moreover, the bi-LSTM technique combines two LSTM networks with two distinct orientations. The bi-LSTM's output considers the two LSTM cell's outputs via the linear layer as given in Eq. (34).

$$U_t = RD_t + g \quad (34)$$

Where, R and g are really the linear layer's weights and biases, and D_t is the hidden phase created by concatenating the forward-hidden phase ht with the backward-hidden phase d'_t. As a result, the bi-LSTM technique may take advantage of the relationship between the information in the present time step and both the past and the future. This research first collects the LS-estimated channels from across all antennas in order to apply the bi-LSTM framework to the network. "Next, it specifies a realization of the given data (input) for the procedure of training in Eq. (35).

$$X_{n-bi-LSTM} = \left\{\left[Re\{\hat{d}^n_{LS}(0)\}; Im\{\hat{d}^n_{LS}(0)\}\right], \ldots\ldots, \left[Re\{\hat{d}^n_{LS}(l-1)\}; Im\{\hat{d}^n_{LS}(l-1)\}\right]\right\} \quad (35)$$

l represents the sequence length taken into account for the bi-LSTM network. The number of characteristics for the input is 2NT×NR since the bi-LSTM's input, $\hat{d}^n_{LS}$ indicates the LS-estimated channel for entire NT×NR channel streams. The equivalent true channel is the bi-LSTM network's output as shown in Eq. (36).

$$O_{n-bi-LSTM} = \left\{\left[Re\{\hat{d}^n(0)\}; Im\{\hat{d}^n(0)\}\right], \ldots\ldots, \left[Re\{\hat{d}^n(l-1)\}; Im\{\hat{d}^n(l-1)\}\right]\right\} \quad (36)$$

The MSE loss function is taken into account since the goal of utilising a bi-LSTM network was to reduce the MSE between

the anticipated channel and the real channel. The bi-LSTM network's goal function is Eq. (37).

$$F_{bi-LSTM}(\mathcal{R}, A) = {}^{1}/_{Nl} \sum_{n=1}^{N} \sum_{j=0}^{l-1} \left\| \hat{d}^{n}(i) - d^{n}(i) \right\|_{2}^{2} \quad (37)$$

Where, N represents the training sample's total number, $\hat{d}^{n}(i)$ represents the actual channel conforming to $\hat{d}^{n}(i)$; $\mathcal{R}$ and A were all the bi-LSTM weights and biases, and the superscript n signifies the nth trained model. By employing gradient descent techniques to update $\mathcal{R}$ and A, the loss function could be reduced. The inadequate channel state data as a potential future expansion of the approach because this research assumes that the ideal instantaneous channels are accessible for the training stage.

The novelty of the proposed approach is integrating a hybrid metaheuristic optimization framework with deep learning overcoming the channel estimation of 5G MIMO-OFDM systems with great success. This method differs with traditional ones which involve only Deep learning or traditional estimators like LS, LMMSE and picks out the best of the two HHO and Sparrow Search Algorithm technologies to tune the parameters of a Long Short-Term Memory (LSTM). This HHO-SSA-LSTM hybrid model has been successful in adjusting itself to channel conditions dynamically in response to environment high-mobility and Doppler effects. Furthermore, OFDM modulation which is based on discrete wavelet transform (DWT) is used to enhance spectral efficiency and eliminate the effects of inter-symbol interference which needs not to use any cyclic prefixes. The metaheuristic layer globally optimizes the learning parameters in LSTM, resulting at such, the parameters capture better temporal relationships of the wireless channel, as well as nonlinear attributes of the wireless channel. This coupled architecture enables more efficient and robust channel estimation mechanism to have a lower BER, MSE, and SER than the existing approaches. In general, the suggested hybrid structure overcomes the weakness of the current estimation methods due to the smart process of the parameter's optimization and dynamic learning in a real-time communication case."

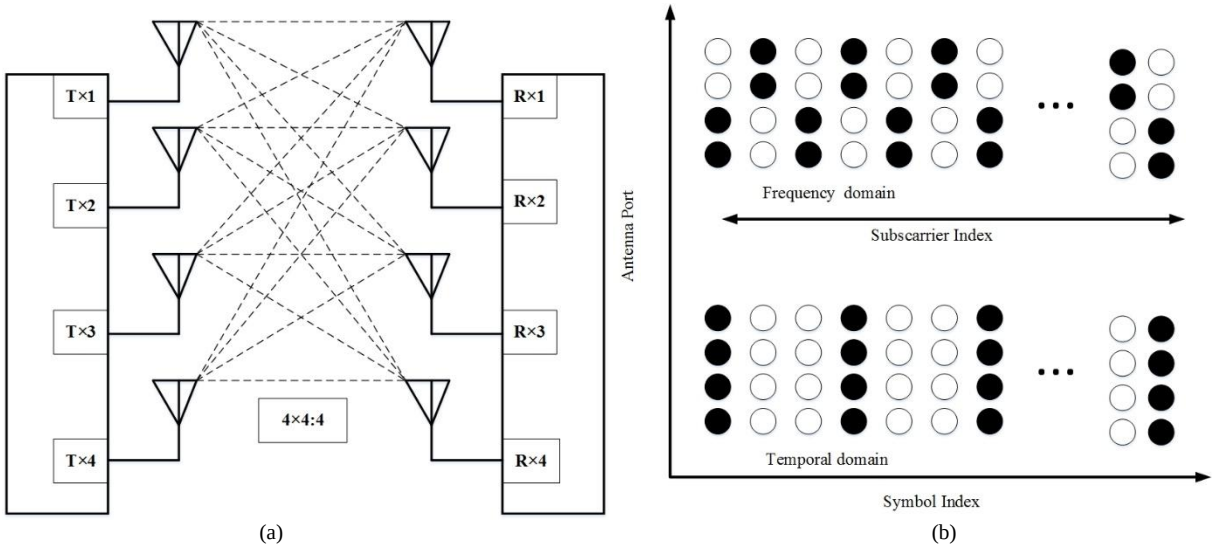
IV. Result and Discussion

This section evaluates the suggested hybrid optimization with DL estimates over the 5G channel profile's performance and compares with conventional approaches, such as LS, CNN, LSTM, FDNN, and LMMSE. Additionally, give a justification for each received outcome. The simulation's settings have been stated first, and the outcomes have been then shown and discussed. The parameters employed for MIMO-OFDM system is shown in Table I.

TABLE I. Parameter Setup

Parameters	Values
Modulation type	16-QAM
Model of noise	Gaussian Noise
FFT size	256
Sample frequency	3.84MHz
Cyclic prefix	24
Subcarrier spacing	15kHz
MIMO	4 × 4

Four antennas are used in 4x4 MIMO, also referred as 4T4R, to connect up to 4-data streams to the recipient device. 4x4 networks offer up to a 400% throughput boost over standard single antenna (SISO) networks. Construction complexity significantly rises as compared to earlier 2x2 antennas. Since the majority of manufacturers cannot offer four unique polarizations, nominal MIMO arrangements are the most typical. Additionally, 4x4 MIMO is up to 160% quicker than 2x2 MIMO in weak signal settings and roughly 90% faster in good signal circumstances. The employed 4x4 MIMO and the used pilot structure for the presented MIMO-OFDM construction are indicated in Fig. 5 (a) and (b), correspondingly. In Fig. 5 (a), T×1 and T×4 seems to be the transmitted antenna-indices and R×1 and R×4 seems to be the receiver antenna-indices.



(a) (b)

Fig. 5. Presented MIMO-OFDM (a) 4×4 MIMO (b) Pilot structure.

The scenario relating to the mobile velocity has been employed to evaluate the effectiveness of all the possible channel estimations. The outcomes of the simulation have been achieved for different models using different performance measures in order to evaluate the suggested DL model-based estimation's effectiveness. The suggested technique has been applied in MATLAB, and the experimental outcomes were obtained. The suggested algorithm evaluates the efficiency of the channel by adjusting the network's SNR in relation to the quantity of sender and recipient arrays. Moreover, the simulation was being run to assess the effectiveness of the HHO-SSA-LSTM-aided estimation, and the outcomes have been contrasted with the traditional LS, CNN, LSTM, FDNN, and LMMSE-based estimation using the BER, SER, and MSE vs SNR, efficiency, and execution time.

A. Performance Analysis

The presented channel estimation strategy's performance is assessed in regards to BER, SER, and MSE for SNR, efficiency, and execution time. The HHO-SSA-LSTM's measures have been contrasted with those of the conventional LS, CNN, LSTM, FDNN, and LMMSE. Additionally, the effectiveness of alternative channel estimation algorithms is assessed across a range of pilot lengths. The description of SNR, BER, SER, and MSE are described as follows:

- SNR: The noise effectiveness, and consequently the susceptibility of radio receivers, could be examined in a variety of methods. Comparing the signal as well as noise rates for a particular signal level, or SNR, seems to be the most straightforward technique. It goes without saying that the bigger the SNR—the difference among the signals and the undesired noise—the greater the radio recipient's sensitivity effectiveness.
- BER and SER: The chances of obtaining a symbol and bit in errors were denoted by SER and BER, correspondingly. By recreating a whole system with a large number of bits as well as comparing the proportion of bits or symbols obtained in error to the overall number-of-bits, BER and SER may be estimated. Hence SER is computed by Eq. (38).

$$BER = {N_S(E)}/{T_N(TS)} \quad (38)$$

Where, $T_N(TS)$ indicates total no. of transferred symbols and $N_B(E)$ indicates no. of symbols in error.

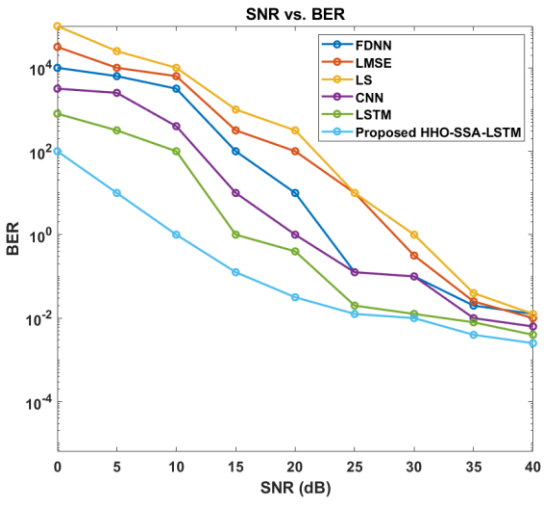
Moreover, BER is computed by Eq. (39).

$$BER = {N_B(E)}/{T_N(TB)} \quad (39)$$

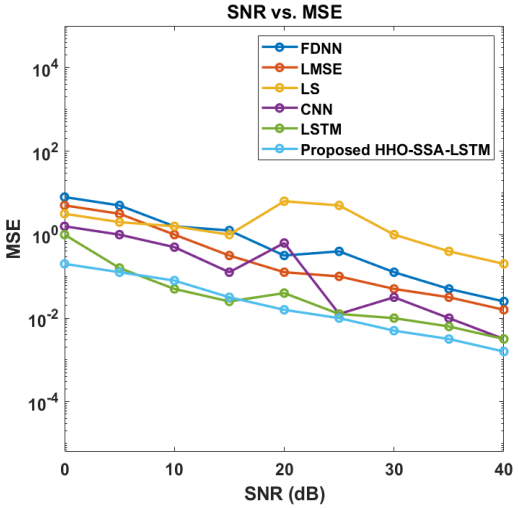
Where, $T_N(TB)$ indicates total no. of transferred bits and $N_B(E)$ indicates no. of bits in error.

- MSE: In statistics, MSE seems to be a measurement of the mean square error. It is an expected value-corresponding risk function. It is random, hence it could hasn't ever be null. It has been always employed in a non-negative context when assessing estimator performance. MSE seems to be a second moment that considers the sources of error as well as estimation variation. An unbiased estimation with invariance of estimate is the MSE. The mean squared variances between the predicted and real values have been used to compute it.

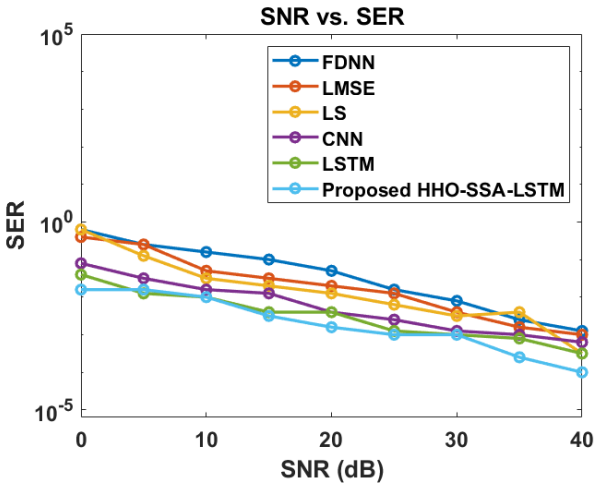
1) Performance evaluation using different pilot lengths: Under several pilot lengths (PL) (i.e., PL =128, 136, and160), the assessment metrics of several channel estimation algorithms are analyzed. Fig. 6 compares the BER, SER, and MSE of several channel estimate algorithms when the PL has been set to 128. As may be seen in Fig. 6 (a), BER (Bit Error Rate) decreases as SNR (Signal to Noise Ration) rises. The LSTM's BER is lowered to 30% less compared to the traditional CNN as long as the weight variables have been chosen for it in the best way possible. However, FDNN has been computationally more efficient than LMSE, and as a result, its BER was 50% lower contrasted to LMSE. In this method, HHO-SSA has been introduced to adjust the weight variables of LSTM in order to produce improved prediction outcome. As a result, the HHO-SSA-LSTM's BER is lower than those of the FDNN, LMSE, LS, CNN, and LSTM by 50%, 75%, 90%, 88%, and 90%, correspondingly.



(a)



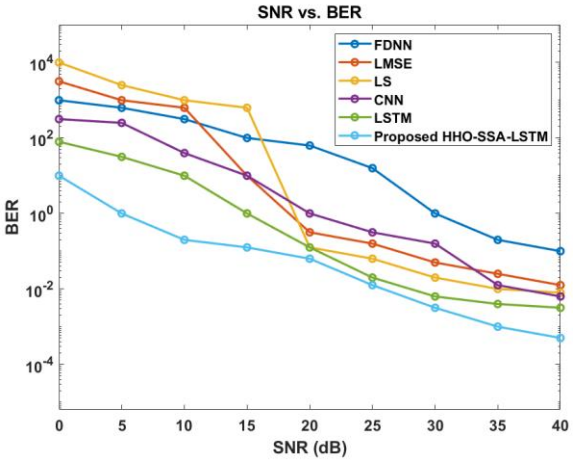
(b)



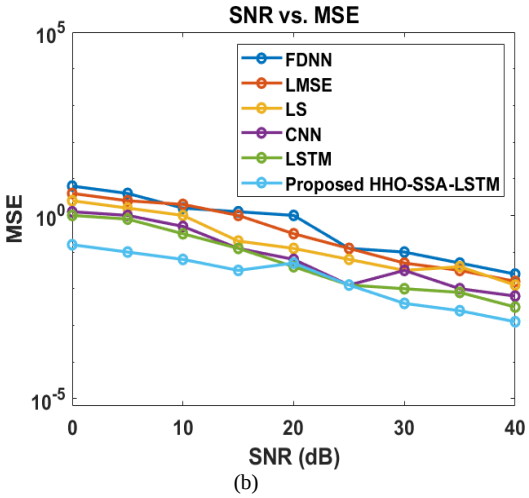
(c)

Fig. 6. "Performance of different channel estimation algorithms when PL=128 (a) BER vs SNR (b) MSE vs SNR (c) SER vs SNR"

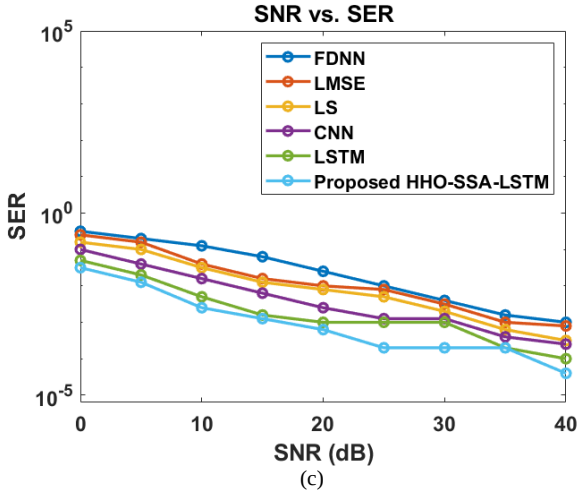
The HHO-SSA-LSTM's MSE (Mean Square Error) is lowered to 40%, 55%, 70%, 85%, and 92% less than that of the FDNN, LMSE, LS, CNN, and LSTM, correspondingly, when PL=128 as demonstrated in Fig. 6 (b). Additionally, as illustrated in Fig. 6 (c), the SER (Symbol Error Rate) of the HHO-SSA-LSTM was lower at PL=128 than it is for the FDNN, LMSE, LS, CNN, and LSTM at 50%, 53%, 64%, 79%, and 82%, correspondingly.



(a)



(b)



(c)

Fig. 7. "Performance of different channel estimation algorithms when PL=136 (a) BER vs. SNR (b) MSE vs. SNR (c) SER vs. SNR"

Fig.7 compares the BER, SER, and MSE of several channel estimate algorithms when PL =136. As may be seen in Fig. 7 (a), BER decreases as SNR rises. The CNN's BER is lowered to 75% less compared to the FDNN; moreover, LMSE's BER is mitigated to 60% than that of CNN. However, LSTM has been computationally more efficient than LS, and as a result, its BER was 25% lower contrasted to LS. In this method, HHO-SSA-LSTM has been introduced to offer improved prediction outcome. As a result, the HHO-SSA-LSTM's BER is lower than those of the all-other methods like FDNN, LMSE, LS, CNN, and LSTM by 20%, 53%, 72%, 80%, and 88%, correspondingly. In addition, HHO-SSA-LSTM's MSE is lowered to 53%, 65%, 72%, 78%, and 85% less than that of the FDNN, LMSE, LS, CNN, and LSTM, respectively, when PL=136 as depicted in Fig. 7 (b). Additionally, as represented in Fig. 7 (c), the SER of the HHO-SSA-LSTM was lower at PL=136 than it is for the FDNN, LMSE, LS, CNN, and LSTM at 65%, 70%, 75%, 89%, and 94%, respectively.

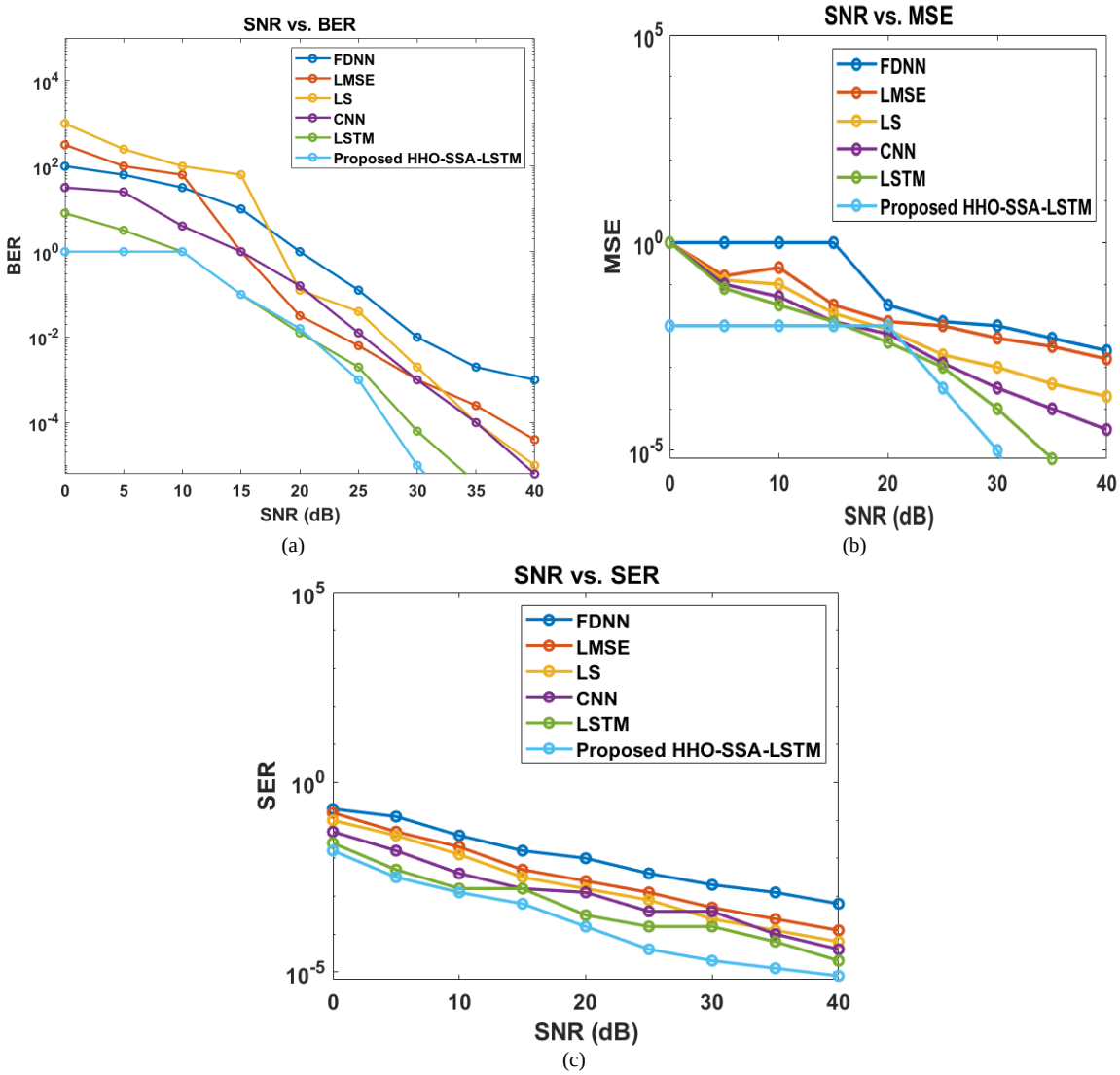


Fig. 8. "Performance of different channel estimation algorithms when PL=160 (a) BER vs. SNR (b) MSE vs. SNR (c) SER vs. SNR".

Fig. 8 compares the BER, SER, and MSE of several channel estimate techniques under the PL=160 condition. As shown in Fig. 8 (a), the BER of HHO-SSA-LSTM was lower than that of FDNN, LS, CNN, LMSE, and LSTM, falling to 76%, 80%, 42%, 83%, and 91%, accordingly. Additionally, as shown in Fig. 8 (b), the MSE of the HHO-SSA-LSTM has been lowered to 73%, 78%, 75%, 80%, and 86%, as opposed to the FDNN, LS, CNN, LMSE, and LSTM, correspondingly. Additionally, as shown in Fig. 8 (c), the SER of the HHO-SSA-LSTM significantly lowered to 58%, 69%, 76%, 84%, and 92%, as opposed to the FDNN, LS, CNN, LMSE, and LSTM, respectively. In addition, the efficiency of the different channel estimation algorithms is also computed for different pilot lengths. Efficiency seems to be the maximum degree of effectiveness that needs the fewest inputs and creates the greatest number of outputs. Efficiency appeals for cutting back on the amount of resources, particularly energy, that are utilized inefficiently to generate a given outcome. Additionally, efficiency seems to be a quantifiable notion that may be assessed by comparing the usable outcome to the entire input.

TABLE II. EFFICIENCY OF HHO-SSA-LSTM OVER FDNN, LSME, LS, CNN, AND LSTM WHEN NUMBER OF PL VARIED

Methods	Efficiency (%)		
	PL = 128	PL = 136	PL = 160
FDNN [36]	85.8	87.64	88.5
LSME [37]	87.7	88.9	90.2
LS [38]	81.1	84.7	87.9
CNN [39]	79.5	84.1	85
LSTM [40]	92.3	94.1	94.91
Proposed HHO-SSA-LSTM	96.41	96.75	96.92

The obtained efficiency is represented in Table II and Fig. 9. The result indicated that the presented HHO-SSA-LSTM has attained higher efficiency under PL = 128, 136, and 160 over other methods. The efficiency of LSTM is increased to 9%, 10%, and 11% than CNN under PL = 128, 136, and 160, respectively. Moreover, when PL = 128, the presented model is enhanced to 11%, 11%, 10%, 9%, and 4% as compared to FDNN, LSME, LS, CNN, and LSTM, respectively. Also, when PL = 136, the presented model is enhanced to 11%, 12%, 10%, 11%, and 2% as compared to FDNN, LSME, LS, CNN, and LSTM, respectively. In addition, when PL = 160, the presented model is enhanced to 12%, 15%, 12%, 11%, and 2% as compared to FDNN, LSME, LS, CNN, and LSTM, correspondingly.

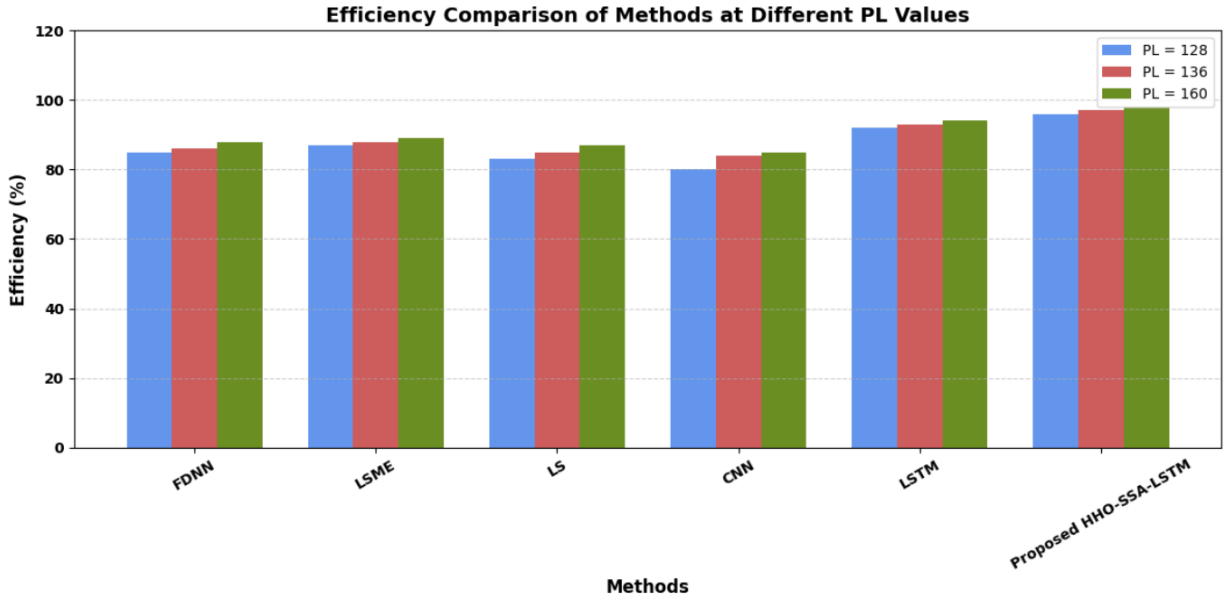


Fig. 9. Efficiency of HHO-SSA-LSTM over other methods.

2) Complexity analysis: The computational assessment of various channel estimate strategies in regards to execution time was shown in Fig. 10 and Table III. Every simulation and execution time measurement was carried out on a system that has an NVIDIA GeForce GTX 1660 Ti GPU, 16 GB of RAM, and an Intel Core i7-10750H CPU running at 2.60GHz. MATLAB R2021b was used for the implementation, and Windows 10 Pro (64-bit) was used. In order to ensure uniformity across all compared models, no GPU acceleration was used during execution time testing. As seen in the table, traditional channel estimation algorithms take longer to execute than DL-based ones due of their computational complexity. However, contrasted to FDNN, LSME, LS, CNN, and LSTM, the given HHO-SSA-LSTM takes a long time to execute. MATLAB's tic and toc functions were used to measure the execution time on a Windows 10 computer running an Intel Core i7-11700 CPU with 32 GB of RAM. After ten runs of each method, the average execution time was noted. Because of its hybrid optimization structure, the suggested HHO-SSA-LSTM model took 22 seconds, which is a little longer than other models. However, this is justified by the fact that it estimates channels with much greater accuracy and efficiency. The computational complexity has grown as a result of the integrated HHO and SSA approaches.

TABLE III. COMPLEXITY ASSESSMENT IN REGARDS TO EXECUTION TIME

Methods	Execution time (sec)
FDNN	21
LSME	18
LS	20
CNN	16
LSTM	10
Proposed HHO-SSA-LSTM	22

Contrasted to the proposed method, the other methods like FDNN, LMSE, LS, CNN, and LSTM has attained less execution time as 21s, 18s, 20s, 16s, and 10s, respectively. Those methods have less complexity due to implementation of single models. However, the presented model was a hybridized method; thus, the execution time was increased than other methods. Moreover, the execution time of the presented HHO-SSA-LSTM is 22s.

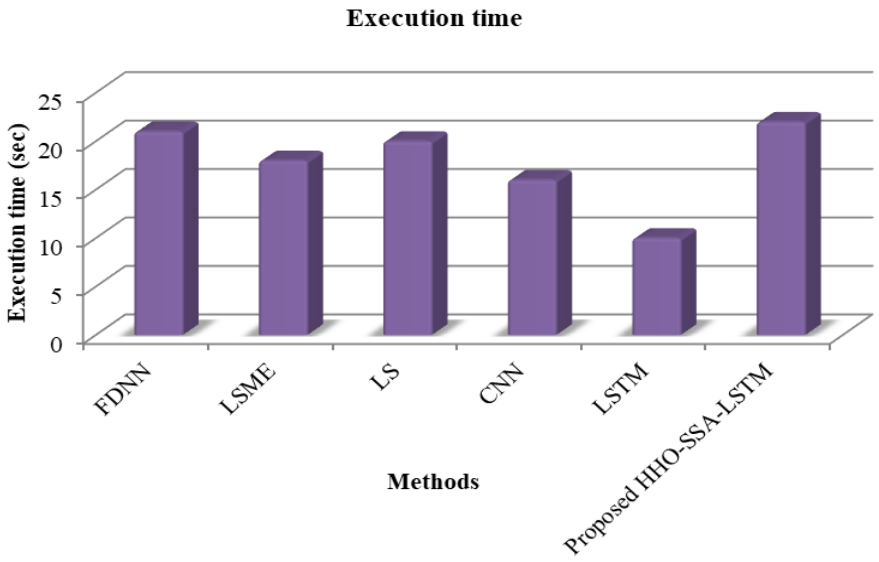


Fig. 10. Execution time.

B. Execution Time Comparison

"The comparative study of the suggested technique with the methods in the prior literature is illustrated in Table IV and Fig. 11. The presented HHO-SSA-LSTM method was compared with other existing methods like Compressive Sensing (CS) [24], OCEAN [41], and PSO-Adam-LSTM [42] in terms of execution time.

TABLE IV. COMPARATIVE ASSESSMENT

Reference	Method	Execution time (s)
[24]	CS	9
[41]	OCEAN	4
[42]	PSO-Adam-LSTM	20
Proposed	HHO-SSA-LSTM	22

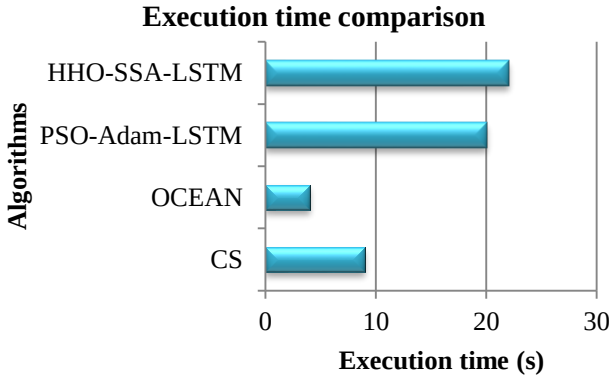


Fig. 11. Execution time comparison.

Haythem Bany Salameh et al. [24] achieved an execution time of 9s, as shown in the table. The researchers obtained highest MSE although execution time was less than the recommended technique. The method put out by Changqing Luo et al. [41] has attained execution time as 4s. Unfortunately, this takes the least amount of time to execute contrasted to the other techniques in the chart. In contrast to the suggested plan, Lipsa and Anand [42] were able to achieve an execution time of 20s. Furthermore, the provided work seems to have a longer execution time than these studies. However, the presented model has attained maximum BER and MSE as less, which is less than the other works.

C. Discussion

The proposed HHO-SSA-LSTM model outperformed traditional and deep learning-based channel estimation techniques—such as LS, LMMSE, CNN, LSTM, and FDNN—on various pilot lengths and signal-to-noise ratio scenarios consistently. The experimental findings showed considerable decreases in BER, SER, and MSE, especially at larger pilot lengths (PL=160), in which the suggested model attained as much as 91% decreased BER and 86% decreased MSE against isolated LSTM. Additionally, the model attained the most efficient estimation, with 96.92%, outperforming even sophisticated DL approaches. While the hybrid model added computational complexity, resulting in the longest execution time (22s), this comes at the expense of the significant improvements in estimation reliability and accuracy [43]. These results have been supported by Mahdi and Deniz [26], who showed that deep learning-based channel estimation models typically have higher execution times because of architectural complexity, comparable to the trade-offs seen in our suggested HHO-SSA-LSTM model, even though they improve accuracy over traditional methods.

The suggested approach is superior despite having a longer execution time because it provides significantly higher accuracy and efficiency, both of which are essential in high-reliability 5G MIMO-OFDM situations. The research presents a compelling argument for the necessity and justification of the trade-off, particularly for applications where accuracy is more important than speed. In comparison with other current benchmark solutions such as CS, OCEAN, and PSO-Adam-LSTM, the proposed scheme represents a reliable solution for channel estimation in 5G MIMO-OFDM systems [44], ideally applicable

for high-reliability and high-capacity communication scenarios. There are several restrictions to take into account, even with the HHO-SSA-LSTM framework's greater accuracy and efficiency. When compared to single-model approaches, the hybrid structure results in longer execution times due to increased computational overhead. Its direct application in real-time or energy-constrained applications may be limited as a result. Additionally, in low-pilot or rapidly changing settings, the model may not always have access to the full channel state information needed for training. The following studies will concentrate on lowering computational complexity to optimize the model for real-time deployment. Additionally, transfer learning and adaptive techniques will be explored to enhance performance in dynamic or low-pilot environments. There will also be consideration of integration with new 6G technologies including THz-band systems and RIS.

V. Conclusion and Future Works

This paper suggested a hybrid deep learning-based channel estimation scheme, HHO-SSA-LSTM, for 5G MIMO-OFDM systems under multi-path fading and Doppler shift. By combining Harris Hawks Optimization and Sparrow Search Algorithm to adjust LSTM parameters, the model showed notably reduced BER, SER, and MSE for various pilot lengths. The performance measurement showed up to 91% BER improvement and 86% MSE reduction in comparison with traditional and deep learning benchmarks like LS, LMMSE, CNN, FDNN, and LSTM. In addition, the novel approach attained maximum efficiency (96.92%) at a pilot size of 160, outperforming the available methods in all instances. The results prove the accuracy and resilience of the HHO-SSA-LSTM technique for practical application scenarios in real-world 5G channel estimation.

Although it has higher accuracy, the proposed model has two primary drawbacks: (1) higher computational complexity leading to longer runtime time (22s) than basic models, and (2) its requirement of a completely known channel profile at training time, which could restrict usability in fast-changing or low-pilot environments. Its computational overhead should be minimized in future studies by developing lightweight variants of the hybrid model for real-time embedded applications. Moreover, being able to extend the model to enable transfer learning and adaptive generalization across various channel conditions will make it more scalable for the next-generation wireless systems, such as 6G and beyond."

References

[1] J. G. Andrews et al., "What Will 5G Be?," IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, June 2014, doi: 10.1109/JSAC.2014.2328098.

[2] X. Wang et al., "Millimeter Wave Communication: A Comprehensive Survey," IEEE Commun. Surv. Tutor., vol. 20, no. 3, pp. 1616–1653, 2018, doi: 10.1109/COMST.2018.2844322.

[3] T. Van Chien, H. Q. Ngo, S. Chatzinotas, M. Di Renzo, and B. Ottersten, "Reconfigurable Intelligent Surface-Assisted Cell-Free Massive MIMO Systems Over Spatially-Correlated Channels," 2021, doi: 10.48550/ARXIV.2104.08648.

[4] J. Tanveer, A. Haider, R. Ali, and A. Kim, "Machine learning for physical layer in 5G and beyond wireless networks: A survey," Electronics, vol. 11, no. 1, p. 121, 2022.

[5] A. L. Ha, T. Van Chien, T. H. Nguyen, W. Choi, and V. D. Nguyen, "Deep Learning-Aided 5G Channel Estimation," in 2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM), Seoul, Korea (South): IEEE, Jan. 2021, pp. 1–7. doi: 10.1109/IMCOM51814.2021.9377351.

[6] C. Tong, "5G Technology Components and Material Solutions for Hardware System Integration," in Advanced Materials and Components for 5G and Beyond, C. Tong, Ed., in Springer Series in Materials Science. , Cham: Springer Nature Switzerland, 2022, pp. 1–32. doi: 10.1007/978-3-031-17207-6_1.

[7] R. Chataut and R. Akl, "Massive MIMO Systems for 5G and beyond Networks—Overview, Recent Trends, Challenges, and Future Research Direction," Sensors, vol. 20, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/s20102753.

[8] A. R. Mishra, Fundamentals of network planning and optimisation 2G/3G/4G: evolution to 5G. John Wiley & Sons, 2018.

[9] S. Hossain, "5G wireless communication systems," Am. J. Eng. Res. AJER, vol. 2, no. 10, pp. 344–353, 2013.

[10] K. S. Kim et al., "Ultrareliable and Low-Latency Communication Techniques for Tactile Internet Services," Proc. IEEE, vol. 107, no. 2, pp. 376–393, Feb. 2019, doi: 10.1109/JPROC.2018.2868995.

[11] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang, and N. Li, "5G Deployment: Standalone vs. Non-Standalone from the Operator Perspective," IEEE Commun. Mag., vol. 58, no. 11, pp. 83–89, Nov. 2020, doi: 10.1109/MCOM.001.2000230.

[12] M. Attaran, "The impact of 5G on the evolution of intelligent automation and industry digitization," J. Ambient Intell. Humaniz. Comput., Feb. 2021, doi: 10.1007/s12652-020-02521-x.

[13] Z. Zhang, J. Chen, and J. Hou, "Analysis of the Evolution Scheme of NSA RAN Sharing to SA RAN Sharing," in 2021 International Wireless Communications and Mobile Computing (IWCMC), June 2021, pp. 204–211. doi: 10.1109/IWCMC51323.2021.9498959.

[14] I. Makhdoom, M. Abolhasan, and J. Lipman, "A comprehensive survey of covert communication techniques, limitations and future challenges," Comput. Secur., vol. 120, p. 102784, Sept. 2022, doi: 10.1016/j.cose.2022.102784.

[15] Aida Zaier and R. Bouallegue, "Channel Estimation Study for Block - Pilot Insertion in OFDM Systems Under Slowly Time Varying Conditions," Int. J. Comput. Netw. Commun., vol. 3, no. 6, pp. 39–54, Nov. 2011, doi: 10.5121/ijcnc.2011.3603.

[16] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency," Found. Trends® Signal Process., vol. 11, no. 3–4, pp. 154–655, 2017, doi: 10.1561/2000000093.

[17] R. Dai, Y. Liu, Q. Wang, Y. Yu, and X. Guo, "Channel estimation by reduced dimension decomposition for millimeter wave massive MIMO system," Phys. Commun., vol. 44, p. 101241, Feb. 2021, doi: 10.1016/j.phycom.2020.101241.

[18] N. Shalavi, M. Atashbar, and M. Mohassel Feghhi, "Downlink channel estimation of FDD based massive MIMO using spatial partial-common sparsity modeling," Phys. Commun., vol. 42, p. 101138, Oct. 2020, doi: 10.1016/j.phycom.2020.101138.

[19] Z. Liu, L. Zhang, and Z. Ding, "Overcoming the Channel Estimation Barrier in Massive MIMO Communication Systems," 2019, doi: 10.48550/ARXIV.1912.10573.

[20] M. Boloursaz Mashhadi and D. Gündüz, "Deep Learning for Massive MIMO Channel State Acquisition and Feedback," J. Indian Inst. Sci., vol. 100, no. 2, pp. 369–382, Apr. 2020, doi: 10.1007/s41745-020-00169-2.

[21] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?," IEEE Trans. Commun., vol. 67, no. 10, pp. 7331–7376, Oct. 2019, doi: 10.1109/TCOMM.2019.2924010.

[22] T. Wang, C.-K. Wen, H. Wang, F. Gao, T. Jiang, and S. Jin, "Deep learning for wireless physical layer: Opportunities and challenges," China Commun., vol. 14, no. 11, pp. 92–111, Nov. 2017, doi: 10.1109/CC.2017.8233654.

[23] H. Huang et al., "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions," IEEE Wirel. Commun., vol. 27, no. 1, pp. 214–222, Feb. 2020, doi: 10.1109/MWC.2019.1900027.

[24] Z. Albataineh, K. Hayajneh, H. Bany Salameh, C. Dang, and A. Dagmseh, "Robust massive MIMO channel estimation for 5G networks using compressive sensing technique," AEU - Int. J. Electron. Commun., vol. 120, p. 153197, June 2020, doi: 10.1016/j.aeue.2020.153197.

[25] X. Ma, Z. Gao, F. Gao, and M. Di Renzo, "Model-Driven Deep Learning Based Channel Estimation and Feedback for Millimeter-Wave Massive Hybrid MIMO Systems," IEEE J. Sel. Areas Commun., vol. 39, no. 8, pp. 2388–2406, Aug. 2021, doi: 10.1109/JSAC.2021.3087269.

[26] M. B. Mashhadi and D. Gunduz, "Pruning the Pilots: Deep Learning-Based Pilot Design and Channel Estimation for MIMO-OFDM Systems," IEEE Trans. Wirel. Commun., vol. 20, no. 10, pp. 6315–6328, Oct. 2021, doi: 10.1109/TWC.2021.3073309.

[27] O. A. Saraereh, I. Khan, Q. Alsafasfeh, S. Alemaishat, and S. Kim, "Low-Complexity Channel Estimation in 5G Massive MIMO-OFDM Systems," Symmetry, vol. 11, no. 5, p. 713, May 2019, doi: 10.3390/sym11050713.

[28] L. Ge et al., "Deep Neural Network Based Channel Estimation for Massive MIMO-OFDM Systems With Imperfect Channel State Information," IEEE Syst. J., vol. 16, no. 3, pp. 4675–4685, Sept. 2022, doi: 10.1109/JSYST.2021.3114229.

[29] N. Anughna and M. Ramesha, "Performance Analysis on various Diversity Schemes with Channel Equalization and Estimation Techniques in MIMO OFDM system," in 2022 3rd International Conference for Emerging Technology (INCET), Belgaum, India: IEEE, May 2022, pp. 1–5. doi: 10.1109/INCET54531.2022.9825175.

[30] L. Ge, Y. Zhang, G. Chen, and J. Tong, "Compression-Based LMMSE Channel Estimation With Adaptive Sparsity for Massive MIMO in 5G Systems," IEEE Syst. J., vol. 13, no. 4, pp. 3847–3857, Dec. 2019, doi: 10.1109/JSYST.2019.2897862.

[31] Y. Zhang, X. Zhu, Y. Liu, Y. Jiang, Y. L. Guan, and V. K. N. Lau, "Hierarchical BEM Based Channel Estimation With Very Low Pilot Overhead for High Mobility MIMO-OFDM Systems," IEEE Trans. Veh. Technol., vol. 71, no. 10, pp. 10543–10558, Oct. 2022, doi: 10.1109/TVT.2022.3184361.

[32] W. Ji, F. Zhang, and L. Qiu, "Multipath Extraction Based UL/DL Channel Estimation for FDD Massive MIMO-OFDM Systems," IEEE Access, vol. 9, pp. 75349–75361, 2021, doi: 10.1109/ACCESS.2021.3081497.

[33] M. A. Gupta, M. P. Nigam, and V. Chaurasia, "A Review on wavelet transform as a substitute to cyclic prefix removal in FFT in OFDM," Int. J. Eng. Sci. Res. Technol., 2014.

[34] A. Al-Jzari and K. Iviva, "Cyclic prefix length determination for orthogonal frequency division multiplexing system over different wireless channel models based on the maximum excess delay spread," Am. J. Eng. Appl. Sci., vol. 8, no. 1, p. 82, 2015.

[35] M. G. Kallur, "Hybrid PSO-GSA Algorithm for Channel Estimation in Massive MIMO System," J. Netw. Commun. Syst., vol. 4, no. 3, Art. no. 3, July 2021, Accessed: Feb. 17, 2023. [Online]. Available: https://publisher.resbee.org/admin/index.php/jnacs/article/view/17

[36] L. Ge et al., "Deep neural network-based channel estimation for massive MIMO-OFDM systems with imperfect channel state information," IEEE Syst. J., vol. 16, no. 3, pp. 4675–4685, 2021.

[37] L. Ge, Y. Zhang, G. Chen, and J. Tong, "Compression-based LMMSE channel estimation with adaptive sparsity for massive MIMO in 5G systems," IEEE Syst. J., vol. 13, no. 4, pp. 3847–3857, 2019.

[38] E. Björnson, J. Hoydis, L. Sanguinetti, and others, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Found. Trends® Signal Process., vol. 11, no. 3–4, pp. 154–655, 2017.

[39] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for MIMO-OFDM systems," IEEE Trans. Wirel. Commun., vol. 20, no. 10, pp. 6315–6328, 2021.

[40] O. A. Saraereh, I. Khan, Q. Alsafasfeh, S. Alemaishat, and S. Kim, "Low-complexity channel estimation in 5G massive MIMO-OFDM systems," Symmetry, vol. 11, no. 5, p. 713, 2019.

[41] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach," IEEE Trans. Netw. Sci. Eng., vol. 7, no. 1, pp. 227–236, Jan. 2020, doi: 10.1109/TNSE.2018.2848960.

[42] L. Dash and A. S. Thampy, "Channel estimation using hybrid optimizer based recurrent neural network long short-term memory for MIMO communications in 5G network," SN Appl. Sci., vol. 5, no. 2, p. 60, Feb. 2023, doi: 10.1007/s42452-022-05253-z.

[43] C. A. ul Hassan et al., "Optimizing Deep Learning Model for Software Cost Estimation Using Hybrid Meta-Heuristic Algorithmic Approach," Comput. Intell. Neurosci., vol. 2022, no. 1, p. 3145956, 2022.

[44] L. Dash and A. S. Thampy, "Channel estimation using hybrid optimizer based recurrent neural network long short term memory for MIMO communications in 5G network," SN Appl. Sci., vol. 5, no. 2, p. 60, 2023.