# Multi-Head Attention-Based Spectrum Sensing for Cognitive Radio

Original Scientific Paper

**B.V. Ravisankar Devarakonda**
ANUCET, ANU
Guntur, India
dbvravisankar@gmail.com

**Venkateswararao Nandanavam**
Department of ECE, Bapatla Engineering College
Bapatla, India
nvrao68@gmail.com

***Abstract*** *– Spectrum sensing is one of the key tasks of cognitive radio to monitor the activity of the primary user. The sensing accuracy of the secondary user is dependent on the signal-to-noise ratio of the primary user signal. A novel Multi-head Attention-based spectrum sensing for Cognitive Radio is proposed through this work to increase the detection probability of the primary user at a low signal-to-noise ratio condition. A radio machine learning dataset with a variety of digital modulation schemes and varying signal-to-noise ratios served as a training source for the proposed model. Further, the performance metrics were evaluated to assess the performance of the proposed model. The experimental results indicate that the proposed model is optimized in terms of the amount of training time required which also has an increase of 27.6% in the probability of detection of the primary user under a low signal-to-noise ratio when compared to other related works that use deep learning.*

***Keywords****: cognitive radio, primary user, spectrum sensing, multi-head attention, additive attention, deep learning*

## 1. INTRODUCTION

Cognitive Radio (CR) [1] is introduced to mitigate the problems of spectrum scarcity and also to provide strategies for efficient communication owing to a huge increase in wireless traffic. Earlier studies on wireless traffic identified the underutilization of the available spectrum by licensed primary users (PUs). The idle time of PU transmission facilitates the unlicensed secondary users (SUs) to dynamically and opportunistically access the spectrum of PU without causing any interference to its transmission [2]. Through spectrum sensing the activity of the PU is monitored continuously by a SU to detect the spectral occupancy of the PU.

The diverse studies published in the area show that the accurate detection of the PU by the SU is highly impacted by the signal-to-noise ratio (SNR), fading, multipath, and shadowing effects [3]. Traditional spectrum sensing algorithms like Energy Detection [4-5], Cyclostationary-based detection [6], and Eigen-value-based detection [7] have been published earlier in the literature. The effects of various types of fading, multipath, and shadowing on spectrum sensing have been investigated in [8-10]. The capacity of fading channels and the data transmission rate through these channels are extremely important for reliable communication between SU and PU [11]. Reliable, efficient, and secured data transmission over wireless channels requires physical layer security and a controllable wireless propagation environment [12-13]. The requirement of an optimal threshold value for different channel conditions is the main drawback of traditional methods of spectrum sensing.

With the rapid advancement in technology and availability of a huge amount of data, Machine Learning (ML) based spectrum sensing algorithms are being implemented in place of traditional spectrum sensing techniques. In the ML-based approach, spectrum sensing is a binary classification to detect the presence or absence of PU. The accuracy of detection of PU presence or absence is significantly impacted by the range of SNR on which the spectrum sensing is performed. Some of the observations on early implementations of the ML models for spectrum sensing are presented in [14-19].

One of the limitations of the ML-based approach is the extraction of the appropriate features as test statistics for accurate decision making. While it is observed that Deep Learning (DL) based approaches help to overcome this limitation as they automatically learn features from the network. DL-based techniques are further used to detect patterns in various applications of natural language processing (NLP), computer vision, and signal processing-related tasks for wireless communications. In recent times, DL-based architectures have gained popularity in the implementation of spectrum sensing algorithms. The various studies that cite DL architectures employ Convolution Neural Networks (CNNs) for spectrum sensing.

Sandeep Kumar et al. [20] proposed a performance analysis of Cooperative Spectrum Sensing (CSS) over α–η–μ and α–κ–μ fading channels using a clustering-based technique. The interesting findings in this model show the use of an energy feature vector.

Dimpal Janu et al. [21] proposed a novel graph convolution network-based adaptive CSS in a cognitive radio network that handles dynamic channel conditions with multiple antennas experiencing different types of fading.

Dimpal Janu et al. [22] presented the performance comparison of machine learning-based multi-antenna CSS algorithms under a multi-path fading scenario. The CNN-based CSS method adopted in this model has obtained intriguing results.

Chang Liu et al. [23] proposed a Deep CM-CNN for spectrum sensing in cognitive radio that takes the covariance matrix as the input to CNN. The CM-CNN-based test statistics generated in this model achieved a higher detection probability.

Youheng Tan et al. [24] implemented CSS based on CNN. Though the model has resulted in better performance, its architecture is complex.

Surendra Solanki et al. [25] presented deep learning for spectrum sensing in cognitive radio and developed DLSenseNet architecture. The authors have not specified the threshold for detection probability, despite the model's interesting results.

Jiabao Gao et al. [26] proposed DLDetectNet architecture. The model was not able to achieve the desired probability of false alarm for various modulation schemes employed in this work.

Kai Yang et al. [27] proposed a blind spectrum sensing method based on deep learning to handle low SNR scenarios using a one-dimensional CNN and long short-term memory (LSTM). At a low SNR value, this method has a low detection probability percentage.

Jiandong Xie et al. [28] proposed a DL-based spectrum sensing in cognitive radio using the CNN-LSTM approach. The range of SNR on which the probability of detection is performed was not properly specified in this work.

Since spectrum sensing is a signal processing-related application that handles time series data, Neural network architectures that can handle and account for the chronological order of data are needed. DL-based architectures such as Recurrent Neural Networks (RNNs) and LSTM are generally used in these applications. However, the issue of vanishing and exploding gradients plagues RNNs. There is a need for DL based architecture that will be unaffected by these gradients.

Of late Transformer based DL architecture became popular and is used in many NLP applications. An architecture of a similar nature has been designed and deployed for this signal processing related spectrum sensing task. The proposed work is motivated by the self attention approach implemented in [29], which is based on an attention mechanism that primarily attends to those parts of the input that would significantly affect the model's prediction.

This paper proposes Multi-Head attention based Spectrum Sensing (MHASS) for CR. An attention function is applied to the input sequence multiple times in parallel in Multi-Head attention (MHA) based on the number of heads. The output of the multiple attention blocks is concatenated to get the overall attention function. The performance metrics considered in this work are the probability of detection (Pd) which is similar to the true positive rate, the probability of false alarm (Pf) which is similar to the false positive rate, precision (Pr), the area under the curve (AUC) and F1 score (F1). The proposed MHASS model is trained, validated, and tested on the signal sensed by SU over a wide range of SNR from -20 dB to 20 dB.

The main contributions of the proposed work are:

- Use of MHASS for the first time in the literature.
- Reducing the number of computations and the amount of time spent to train the model.
- Best performance metrics obtained at low SNR.

The forthcoming sections in this paper are organized as follows. Section 2 discusses the DL system model, section 3 describes the proposed MHASS model, section 4 presents the experimental setup, section 5 analyzes the results & discussions and conclusions are presented in section 6.

## 2. DL SYSTEM MODEL

The DL system model for spectrum sensing is depicted in Fig.1.It can be seen that the N observation vectors are utilized for spectrum sensing.

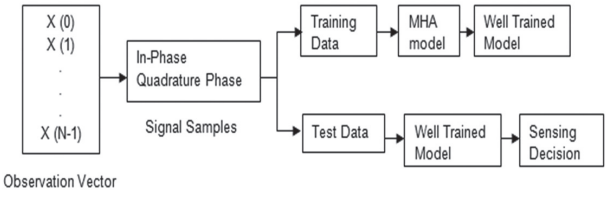


**Fig. 1.** DL System Model

Consider $X(n) = \{x_1(n), x_2(n), \ldots, xM(n)\}^T$, where $M$ represents the signal's sample length $l$, $n$=0,1,…, $N$-1,

denotes the $n^{th}$ received observation vector and $xi(n)$ denotes the $n^{th}$ discrete time sample.

The sensing decision in DL-based spectrum sensing has been formulated as a binary hypothesis as represented by (1).

$$X(n) = \begin{cases} R(n) + W(n) & : H_1 \\ W(n) & : H_0 \end{cases} \quad (1)$$

Here $R(n)$ denotes the PU signal samples vector, which also suffers from path loss and fading. $W(n)$ is the noise samples vector and $X(n)$ is the received SU observation vector. H1 denotes the PU presence and H0 denotes the PU absence.

The received observation vector consists of In-Phase (IP) and Quadrature Phase (QP) signal samples. The dataset of IP and QP samples is obtained from a Radio Machine Learning (RML) pickle file 'RML2016.10a_dict.pkl' [30] which is composed of eight digital modulations, multipath loss, Rician, and Rayleigh fading effects. A noise vector having a similar length of signal samples is generated using the Additive White Gaussian Noise (AWGN) scheme.

The real and imaginary parts of the SU observation vector are denoted as XI and XQ respectively. The received complex signal with both real and imaginary parts is given by (2).

$$\hat{X} = (X_I, X_Q) \quad (2)$$

The energy of this received complex signal is calculated as given by (3).

$$E = \sum_{i=1}^{M} (|\hat{X}_i|)^2 \quad (3)$$

To scale all the amplitude values of the signal samples to a similar magnitude, energy normalization is performed as given by (4).

$$\hat{X}_{norm} = \frac{\hat{X}}{E} \quad (4)$$

In the training phase, the proposed MHASS model is trained on annotated data of the energy-normalized signal samples and noise samples such that the occurrence of signal samples can be considered as PU presence (Label=1) and the occurrence of only noise samples can be considered as PU absence (Label=0). The implementation details of the MHASS model are discussed in section 3.

## 3. PROPOSED MHASS MODEL

In the proposed model for spectrum sensing, MHA has been used on the input layer consisting of IP and QP samples along with the noise samples. It has been observed that MHA attends to only those parts of the input tensor which are used to compute the decision of PU presence or absence. The use of MHA facilitated the subsequent convolution layer to have pre-computed attention over the input convolution volume. As a result of this, the network converged in less number of epochs. The main operation involved in MHA was the Scaled Dot Product Attention followed by concatenating the attention functions obtained from multiple attention blocks.

### 3.1. SCALED DOT PRODUCT ATTENTION

An attention function can be described as mapping a query ($Q$) and a set of key-value pairs $K$ and $V$ respectively to an output. $Q$, $K$, and $V$ are matrices representing the input sequence. The output is computed as a weighted sum of the values, where the weight assigned to each value is computed by a compatibility function of the query with the corresponding key.

Query: The query $Q$ is a feature vector of dimension m x n that describes the aspects being looked for in the input sequence.

Key: For each input element, a key $K$ of dimension $m\ x\ n$ is associated which is also a feature vector, that can identify the elements for which attention has to be paid based on the query.

Value: For each input element, value vector $V$ of dimension m $x$ $n$, the output can be computed as a weighted sum of the values. Each value can be accessed using the key $K$.

The scaled dot product attention is shown in Fig.2.

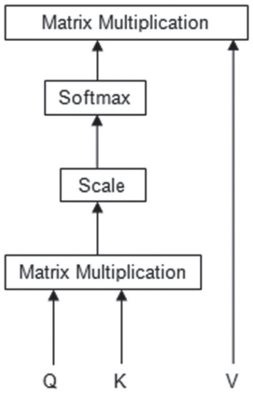


**Fig. 2.** Scaled dot product attention

**Table 1.** Parameters/ Hyperparameters of the MHA model

| Parameters / Hyper-parameters | Description | Value |
|---|---|---|
| m | Number of tokens in the input sequence | 2 (I, Q) |
| n | Dimensionality of hidden/ embedding layer | 128 |
| dk | Dimension of Q.K and V | 128 |
| Q | Query | 2x128 matrix |
| K | Key | 2x128 matrix |
| V | Value | 2x128 matrix |
| dmodel | Model dimension | 128 |
| h | Number of attention heads | 16 |
| wq | Query Transformation Matrix | 128x128 |
| wk | Key Transformation Matrix | 128x128 |
| wv | Value Transformation Matrix | 128x128 |

The attention function performed on three matrices $Q$, $K$, and $V$ of order $m\ x\ n$ is given by (5).

$$\text{Attention}(\mathrm{Q},\mathrm{K},\mathrm{V}) = \text{softmax}\left(\frac{\mathrm{QK^T}}{\sqrt{\mathrm{d_k}}}\right).\mathrm{V} \quad (5)$$

In the proposed model, instead of the original input sequence, a parameterized form of attention has been used as given in (6).

$$\text{Attention}\left(\mathrm{w_q Q}, \mathrm{w_k K}, \mathrm{w_v V}\right) = \text{softmax}\left(\frac{\mathrm{w_q Q}\ (\mathrm{w_k K})^{\mathrm{T}}}{\sqrt{\mathrm{d_k}}}\right).\mathrm{w_v V} \quad (6)$$

The steps involved in scaled dot product attention are:

- Initialize the three matrices $Q$, $K$, and $V$.

$$\mathrm{Q} = \begin{bmatrix}\mathrm{I_P}\\ \mathrm{Q_P}\end{bmatrix}_{2\mathrm{X}128} \quad \mathrm{K} = \begin{bmatrix}\mathrm{I_P}\\ \mathrm{Q_P}\end{bmatrix}_{2\mathrm{X}128} \quad \mathrm{V} = \begin{bmatrix}\mathrm{I_P}\\ \mathrm{Q_P}\end{bmatrix}_{2\mathrm{X}128}$$

- Compute the weight matrix by performing the dot product $Q.K^T$.

$$\begin{bmatrix}\mathrm{I_P}\\ \mathrm{Q_P}\end{bmatrix}_{2\mathrm{X}128} . [\mathrm{I_P Q_P}]_{128\mathrm{X}2} = \begin{bmatrix}\mathrm{w_1} & \mathrm{w_2}\\ \mathrm{w_3} & \mathrm{w_4}\end{bmatrix}_{2\mathrm{X}2}$$

- Apply the softmax on the scaled weight matrix and multiply it with V to calculate the attention function as (7).

$$\text{softmax}\left(\begin{bmatrix}\frac{\mathrm{w_1}}{\sqrt{128}} & \frac{\mathrm{w_2}}{\sqrt{128}}\\ \frac{\mathrm{w_3}}{\sqrt{128}} & \frac{\mathrm{w_4}}{\sqrt{128}}\end{bmatrix}_{2\mathrm{x}2}\right)\begin{bmatrix}\mathrm{I_P}\\ \mathrm{Q_P}\end{bmatrix}_{2\mathrm{X}128} = \text{Attention}\ (\mathrm{Q},\mathrm{K},\mathrm{V}) \quad (7)$$

### 3.2. MHA MECHANISM

In the MHA mechanism, instead of a single attention function, linearly project $Q$, $K$, and $V$ to a lower dimension and perform attention in parallel on the projected versions of $Q$, $K$, and $V$.Concatenate the output from h individual attention functions and perform final linear projection.

Fig. 3 illustrates the block diagram representation of the MHA mechanism with Attn-1, Attn-2, …., and Attn-h representing the attention functions of h individual heads respectively.

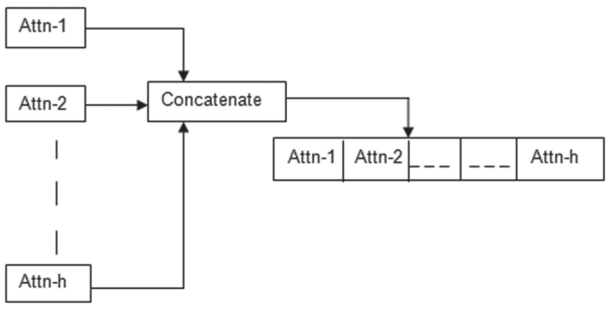


**Fig. 3.** MHA mechanism

The steps involved in the MHA mechanism are:

- Select the value of $h$ as a factor of model dimension. The value of $h$ in the proposed model is 16.
- Calculate the value of $dk = dmodel\ /\ h$.
- Perform the linear projection of $Q$, $K$, and $V$ each of dimension $2xdk$, $h$ number of times.
- Calculate the individual attention functions Attn-1, Attn-2,…, Attn-h.
- Compute the overall attention function by concatenating the individual attention functions.
- The concept of MHA for $h$=2 is illustrated as follows:
- Calculate $dk$= $dmodel$ / $h$ =128/2=64
- Let Attn-1 and Attn-2 will be the individual attention functions and they are computed as (8).

$$\mathrm{Attn-1} = \mathrm{Attn-2} = \text{softmax}\left(\frac{[Q]_{2X64}.\ [K^T]_{2X64}}{\sqrt{64}}\right) \quad (8)$$

- Concatenate Attn-1 and Attn-2 to compute the overall attention function.

### 3.3. ARCHITECTURE OF MHA

The architecture of MHA used for the experimental evaluation is shown in Fig. 4.

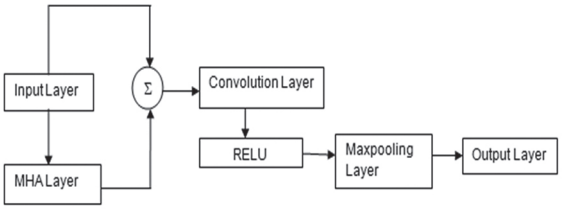


**Fig. 4.** Architecture of MHA

A 2 x 128 length sequence of the RML dataset serves as a training source for the input layer. Multi-Head attention is applied on the input layer followed by additive attention of the MHA layer with the input layer. The MHA attention input is then fed to the convolution layer, where the entire network will be trained end to end and validated on the dataset to build a well trained model as shown in the DL system model presented in section 2. The well trained model is tested on unseen samples of the dataset to predict the sensing decision of PU presence or absence at the output layer. The hyperparameters of the convolution layer are listed in Table 2.

**Table 2.** Hyperparameters of the convolution layer

| Hyperparameters | Value |
|---|---|
| Number of filters in the convolution layer | 64 |
| Kernel size of convolution layer | 5x5 |
| Dropout ratio | 0.25 |
| Max Pooling kernel size | 2x2 |
| Batch Size | 256 |
| Optimizer | Adam |

## 4. EXPERIMENTAL SETUP

The specifications of the dataset and the performance metrics used are discussed in this section. The proposed model makes use of the RML dataset, the parameters of which are tabulated in Table 3.

**Table 3.** RML Dataset Parameters

| Parameters | Value |
|---|---|
| Modulation scheme | 8PSK,BPSK, CPFSK, GFSK,PAM4,QAM16, QAM64, QPSK |
| Fading effects of the channel | Rician, Rayleigh |
| Sample Length | 128 |
| SNR Range | -20 dB to 20 dB in 2 dB increments |
| Training Samples | 112000 |
| Validation Samples | 32000 |
| Test Samples | 16000 |

### 4.1. PERFORMANCE METRICS

The proposed model is trained and validated on the RML dataset, and the performance metrics like $P_d$, $P_f$, AUC, and $F1$ are observed to evaluate its performance. Pd denotes the probability of PU presence when PU occupies the spectrum, and $P_f$ denotes the probability of PU presence when PU is not utilizing the spectrum. AUC is the overall area occupied by the receiver operating characteristics (ROC) and the quality of the model is indicated by $F1$ which is dependent on precision and recall.

### 4.2. TEST STATISTIC

The threshold (TH) value for various values of probability of false alarm (PFA) is calculated and it is compared with the test statistic $T$ as in (9).

$$\mathbf{T} = \frac{\mathbf{P(H_1)}}{\mathbf{P(H_0)}} \tag{9}$$

Where $\mathrm{P(H1)}$ and $\mathrm{P(H0)}$ represent the probability of PU presence and absence respectively. The proposed model predicts the likelihood of PU presence or absence using the two approaches listed below.

- A standard threshold value of 0.5 is used in all binary classification ML/DL algorithms.
- T>TH denotes the PU presence and T < TH denotes the PU absence.

## 5. RESULTS AND DISCUSSIONS

The proposed model's novelty and its effectiveness are discussed in this section The results of the validation and test evaluation metrics are used to assess the performance of the proposed MHASS model. The additive attention signal pattern of the MHA layer is discussed in section 5.1. In section 5.2 the proposed model's performance is compared to that of a DL CNN model. The ROC of the proposed model is shown in section 5.3. Analysis of the results and improvement in the performance over the previous work is discussed in sections 5.4 and 5.5 respectively.

### 5.1. ADDITIVE ATTENTION SIGNAL PATTERN

The pattern of the input sequence that is added to the attention weights which are obtained from the MHA layer for the various modulated signals at various SNR values is considered as an additive attention signal pattern. The novel nature of the proposed model is evident from this pattern. An additive attention signal pattern for QPSK modulation at SNR= 12 dB is depicted in Fig. 5.

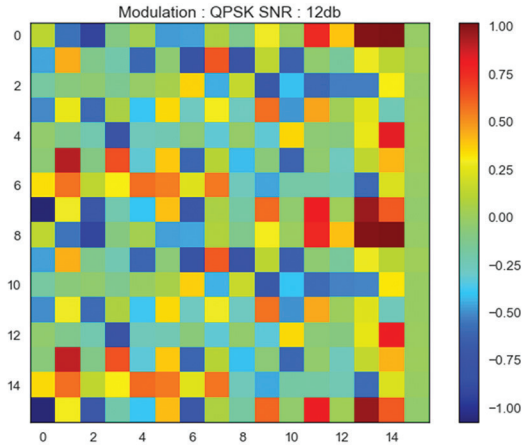


**Fig. 5.** Additive Attention Signal Pattern

In Fig. 5 the additive attention signal pattern is represented as a 16x16 image which is symmetric for both the IP and QP samples. Different colors in the image represent the signal intensities at each location of the image. The range of signal intensities represented by the vertical colour bar indicates the parts of the input sequence that should be considered for improved prediction.

The following inferences can be drawn from the additive attention signal pattern:

- A large part of the input has non-zero attention weights, which indicates that different parts of the input sequence are attended to increase the likelihood of PU presence.
- The peak levels of the signal at specific locations of the input have a very high magnitude, indicating the presence of PU.

### 5.2. PERFORMANCE COMPARISON

A DL CNN model without MHA is contrasted with the MHASS model's performance. Both models were compared for training, validation loss, and detection probability at low SNR.

The training and validation loss with respect to the number of epochs is shown in Fig. 6. MHA_TRAIN and MHA_VAL represent the training and validation loss of the MHASS model respectively. CNN_TRAIN and CNN_LOSS denote the training and validation loss of the CNN model respectively.

Fig. 6. depicts the fast convergence of the proposed MHASS model with optimized training time in comparison to CNN.

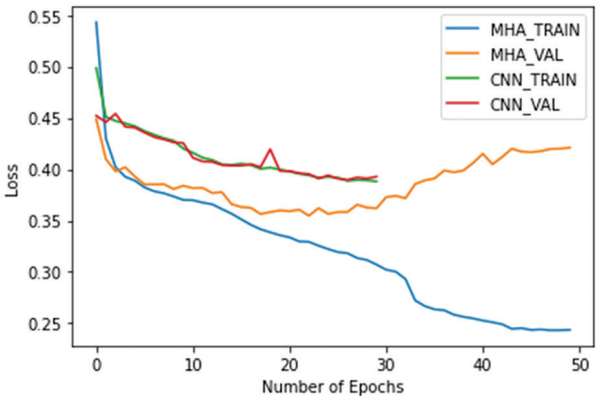


**Fig .6.** Loss VS Epochs for MHASS, CNN

The above figure depicts the fast convergence of the proposed MHASS model with optimized training time in comparison to CNN.

The plot of SNR and percentage of detection probability ($P_d$ %) of the MHA and CNN models is shown in Fig.7.

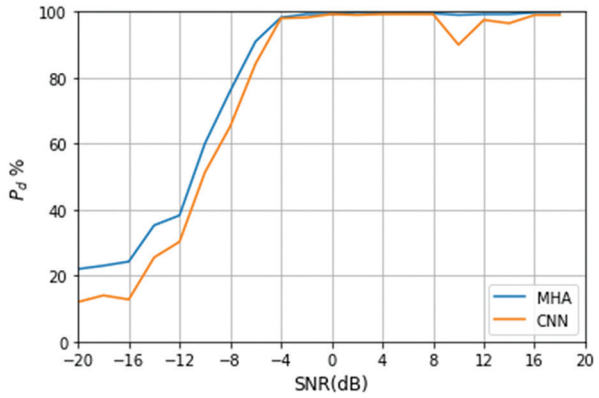


**Fig. 7.** SNR VS $P_d$

Fig.7 is an indication of the increased probability of detection of the proposed model when compared to CNN at low SNR.

### 5.3. ROC OF THE PROPOSED MHASS MODEL

The ROC of the MHASS model is given in Fig. 8.

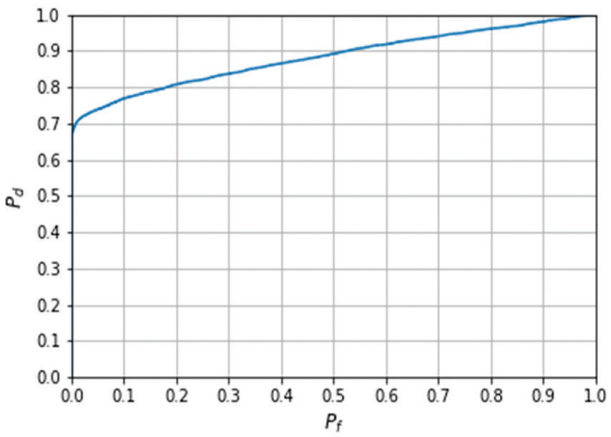

**Fig. 8.** ROC of MHASS model

The ROC of the MHASS model for SNR= - 20 dB and SNR = - 6 dB is shown in Fig. 9.

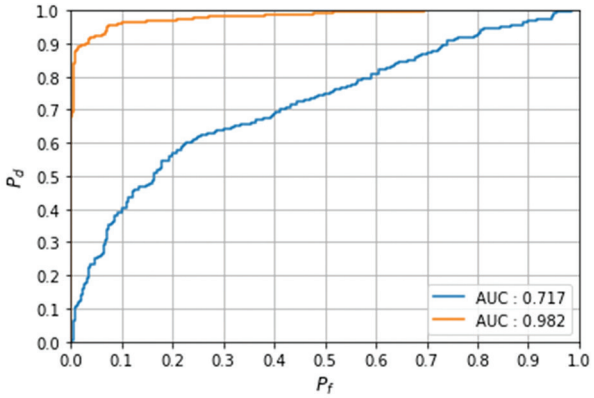


**Fig. 9.** ROC for SNR = -6 dB and SNR = -20 dB

Table 4 displays the proposed model's test evaluation metrics for different SNR values.

**Table 4.** Performance Metrics for Different SNR

| SNR(dB) | Pd | Pr | AUC | F1 |
|---|---|---|---|---|
| -20 | 0.22 | 0.863 | 0.717 | 0.351 |
| -18 | 0.23 | 0.868 | 0.744 | 0.364 |
| -16 | 0.243 | 0.874 | 0.745 | 0.38 |
| -14 | 0.353 | 0.91 | 0.795 | 0.51 |
| -12 | 0.383 | 0.916 | 0.814 | 0.54 |
| -10 | 0.6 | 0.945 | 0.89 | 0.734 |
| -8 | 0.76 | 0.956 | 0.937 | 0.847 |
| -6 | 0.91 | 0.963 | 0.982 | 0.936 |
| -4 | 0.9823 | 0.966 | 0.995 | 0.974 |
| -2 | 0.993 | 0.966 | 0.999 | 0.979 |
| 0 | 0.993 | 0.966 | 0.999 | 0.979 |
| 2 | 0.998 | 0.966 | 0.999 | 0.981 |
| 4 | 0.995 | 0.966 | 0.999 | 0.980 |
| 6 | 0.995 | 0.966 | 0.998 | 0.980 |
| 8 | 0.995 | 0.966 | 0.997 | 0.980 |
| 10 | 0.99 | 0.966 | 0.998 | 0.978 |
| 12 | 0.998 | 0.966 | 0.999 | 0.982 |
| 14 | 0.993 | 0.966 | 0.998 | 0.98 |
| 16 | 0.998 | 0.966 | 0.999 | 0.982 |
| 18 | 0.998 | 0.966 | 0.999 | 0.982 |

The impact of modulation schemes on percentages of detection probability (Pd %) and false alarm probability. (Pf %) at SNR = - 20 dB is displayed in Table. 5.

The Pd% of the proposed model to a varying range of SNR for different values of PFA based on the second test statistic mentioned in section 4.2 is shown in Fig. 10.

**Table 5.** Impact of modulation schemes

| Modulation Scheme | Pd (%) | Pf (%) |
|---|---|---|
| 8PSK | 19 | 0 |
| BPSK | 23 | 0 |
| CPFSK | 18 | 0 |
| GPSK | 14 | 0 |
| PAM4 | 24 | 0 |
| QAM16 | 23 | 0 |
| QAM64 | 28 | 0 |
| QPSK | 25 | 0 |
| All modulations | 22.5 | 4 |

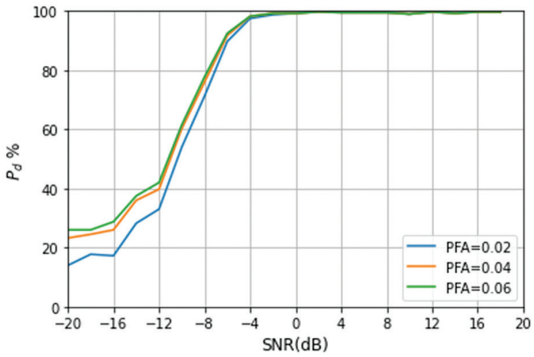


**Fig.10.** SNR VS Pd for different values of PFA

### 5. 4. ANALYSIS OF THE RESULTS

The following observations can be made after analyzing the results :

- From Table 4 the values obtained for Pd, AUC, and F1 are 0.91, 0.982, and 0.936 respectively at SNR= -6 dB. These values indicate that the proposed model has a high detection probability at a low SNR.
- Fig. 9 denotes the best AUC values obtained for low SNR of -20 dB and -6 dB respectively.
- Table 5 indicates that the best values of Pd are obtained with zero percentage of Pf for various modulation schemes at a low SNR = -20 dB.
- Fig.10 shows that for various values of PFA best performance metrics are obtained for the proposed model at a low SNR.

### 5. 5. IMPROVEMENT IN PERFORMANCE

The proposed MHASS model's performance in comparison to prior work is shown in Tables 6 and 7.

**Table 6.** Comparison of the $P_d$ and $P_f$ at SNR=-20 dB

| Model | Modulation Scheme | $P_d$ (%) | $P_f$ (%) |
|---|---|---|---|
| Gao et al. [26] | QPSK | <20 | 7.81 |
| | QAM16 | <20 | 6.54 |
| | QAM64 | <20 | 7.82 |
| Proposed MHASS | QPSK | 25 | 0 |
| | QAM16 | 23 | 0 |
| | QAM64 | 28 | 0 |

**Table 7.** Improvement in Pd % of the proposed model

| SNR (dB) | Model | $P_d$ (%) | % Improvement in $P_d$ |
|---|---|---|---|
| -6 | Kai Yang et al. [27] | 80 | 13.8 |
| | Proposed MHASS | 91 | |
| -12 | Kai Yang et al. [27] | 30 | 27.6 |
| | Proposed MHASS | 38.3 | |

From Table 6, it is observed that various modulation schemes used in the proposed model resulted in increased Pd % with 0% Pf when compared to the previously reported model.

Table 7 indicates that the proposed model has resulted in an improvement of 13.8 % and 27.6 % in the value of Pd for SNR values of -6 dB and -20 dB respectively when compared to the previously developed model.

## 6. CONCLUSION

Multi-Head attention based spectrum sensing for cognitive radio has been implemented in this work. The implemented model resulted in the best performance metrics such as Pd, Pf, AUC, and F1 over a wide range of SNR. The ROC and other plots obtained indicate that a higher value of detection probability is achieved at a low SNR for various modulations of the dataset used in this model. The use of multi-head attention has resulted in faster convergence of the proposed model with less number of computations. There is an improvement of 27.6 % in Pd (%) when compared to one of the previous works in deep learning. This work can be further extended by proposing a cooperative spectrum sensing scheme in which the secondary users are experiencing different levels of fading and other multipath effects.

## 7. REFERENCES


[1] J. Mitola, G. Q. Maguire, "Cognitive radio: making software radios more personal", IEEE Personal Communications, Vol. 6, No. 4, 1999, pp. 13-18.

[2] S. Haykin, "Cognitive Radio: Brain-empowered wireless communications", IEEE Journal on Selected Areas in Communications, Vol. 23, No. 2, 2005, pp. 201-220.

[3] K. Kockaya, I. Develi, "Spectrum sensing in cognitive radio networks: threshold optimization and analysis", EURASIP Journal on Wireless Communications and Networking, Vol. 255, 2020.

[4] J. Luo, G. Zhang, C. Yan, " An Energy Detection-Based Spectrum Sensing Method for Cognitive Radio", Wireless Communications and Mobile Computing, Vol. 2022, 2022, pp. 1-10.

[5] S. Srinu, S. L. Sabat, “Effective cooperative wideband sensing using energy detection under suspicious Cognitive Radio Network”, Computers & Electrical Engineering, Vol. 39, No. 4, 2013, pp. 1153-1163.

[6] G. R. George, S. C. Prema, “Cyclostationary Feature Detection Based Blind Approach for Spectrum Sensing and Classification”, Radioengineering, Vol. 28, No. 1, 2019, pp. 298-303.

[7] Y. Zeng, Y. C. Liang, “Eigenvalue-based Spectrum Sensing Algorithms for Cognitive Radio”, IEEE Transactions on Communications, Vol. 57, No. 6, 2009, pp. 1784-1793.

[8] S. Kumar, “Energy Detection in Hoyt/Gamma Fading Channel with Micro-Diversity Reception”, Wireless Personal Communications, Vol. 101, No. 2, 2018, pp. 723-734.

[9] S. Kumar, P. K. Verma, M. Kaur, P. Jain, S. K. Soni, "On the spectrum sensing of gamma shadowed Hoyt fading channel with MRC reception”, Journal of Electromagnetic Waves and Applications, Vol. 32, No. 16, 2018, pp. 2157-2166.

[10] S. Kumar, “Performance of ED Based Spectrum Sensing Over α–η–μ Fading Channel”, Wireless Personal Communications, Vol. 100, No. 4, 2018, pp. 1845-1857.

[11] P. Yadav, S. Kumar, R. Kumar, "A Review of Transmission Rate over Wireless Fading Channels: Classifications, Applications, and Challenges", Wireless Personal Communications, Vol. 122, No. 2, 2022, pp. 1709-1765.

[12] P. Yadav, S. Kumar, R. Kumar, “A comprehensive survey of physical layer security over fading channels: Classifications, applications, and challenges”, Transactions on Emerging Telecommunications Technologies, Vol. 32, No. 9, 2021, pp. 1-41.

[13] S. Kumar, P. Yadav, M. Kaur, R. Kumar, “A survey on IRS NOMA integrated communication networks”, Telecommunication Systems, Vol. 80, No. 2, 2022, pp. 277-302.

[14] C. Clancy, J. Hecker, E. Stuntebeck, T. O'Shea, “Applications of Machine Learning to Cognitive Radio Networks”, IEEE Wireless Communications, Vol. 14, No. 4, 2007, pp. 47-52.

[15] D. Janu, K. Singh, S. Kumar, "Machine learning for cooperative spectrum sensing and sharing: A survey”, Transactions on Emerging Telecommunications Technologies, Vol. 33, No. 1, 2022, pp. 1-28.

[16] R. Sarikhani, F. Keynia, “Cooperative Spectrum Sensing Meets Machine Learning: Deep Reinforcement Learning Approach”, IEEE Communications Letters, Vol. 24, No. 7, 2020, pp. 1459-1462.

[17] N. Abbas, Y. Nasser, K. El Ahmad, “Recent advances on artificial intelligence and learning techniques in cognitive radio networks”, EURASIP Journal on Wireless Communications and Networking, 2015, pp. 1-20.

[18] C. H. A. Tavares, J. C. Marinello, M. L. Proenca Jr., T. Abrao, “Machine learning-based models for spectrum sensing in cooperative radio networks”, IET Communications, Vol. 14, No. 18, 2020, pp. 3102-3109.

[19] D. B. V. Ravisankar, N. Venkateswararao, “Ensemble Classifier with Heterogenous Fusion Center for Cooperative Spectrum Sensing in Cognitive Radio”, Journal of Interconnection Networks, Vol. 22, No. Supp01, 2022, pp. 1-20.

[20] S. Kumar, P. S. Chauhan, R. Bansal, M. Kaur, R. K. Yadav, “Performance Analysis of CSS Over α–η–μ and α-k-μ Fading Channel Using Clustering-Based Technique”, Wireless Personal Communications, Vol. 126, No. 4, 2022, pp. 3595-3610.

[21] D. Janu, S. Kumar, K. Singh, “A Graph Convolution Network Based Adaptive Cooperative Spectrum Sensing in Cognitive Radio Network”, IEEE Transactions on Vehicular Technology, Vol. 72, 2022, pp. 1-11.

[22] D. Janu, K. Singh, S. Kumar, “Performance Comparison of Machine Learning based Multi-Antenna Cooperative Spectrum Sensing algorithms under Multi-Path Fading Scenario”, Proceedings of the IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications, Goa, India, 8-9 October 2022.

[23] C. Liu, J. Wang, X. Liu, Y.-C. Liang, “Deep CM-CNN for Spectrum Sensing in Cognitive Radio”, IEEE Journal on Selected Areas in Communications, Vol. 37, No. 10, 2019, pp. 2306-2321.

[24] Y. Tan, X. Jing, "Cooperative Spectrum Sensing Based on Convolutional Neural Networks", Applied Sciences, Vol. 11, No. 10, 2021, pp.1-13.

[25] S. Solanki, V. Dehalwar, J. Choudhary, "Deep Learning for Spectrum Sensing in Cognitive Radio", Symmetry, Vol. 13, No. 1, 2021.

[26] J. Gao, X. Yi, C. Zhong, X. Chen, Z. Zhang, "Deep Learning for Spectrum Sensing", IEEE Wireless Communications Letters, Vol. 8, No. 6, 2019, pp. 1727-1730.

[27] K. Yang, Z. Huang, X. Wang, X. Li, "A Blind Spectrum Sensing Method Based on Deep Learning", Sensors, Vol. 19, No. 10, 2019, pp. 1-17.

[28] J. Xie, J. Fang, C. Liu, X. Li, "Deep Learning-Based Spectrum Sensing in Cognitive Radio: A CNN-LSTM Approach", IEEE Communications Letters, Vol. 24, No. 10, 2020, pp. 2196-2200.

[29] A. Vaswani et al. "Attention Is All You Need", Proceedings of 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4-9 December 2017, pp.6000-6010.

[30] T. O'Shea, N. West, "Radio Machine learning Dataset Generation with GNU Radio", Proceedings of the 6th GNU Radio Conference, University of Colorado, Boulder, USA, 12-16 September 2016.