MACHINE LEARNING BASED SPECTRUM SENSING FOR INTERFERENCE REDUCTION IN 5G COGNITIVE RADIO NETWORKS

Judith Lihemo Ajega

A thesis submitted in partial fulfillment of the requirements for the award of the degree of Master of Science in Electrical Engineering, Telecommunication Option at Masinde Muliro University of Science and Technology.

September, 2025

DECLARATION

This thesis is my original work prepared with no other than the indicated sources and support and has not been presented elsewhere for a degree or any other award.

Sign …………………………. Date………………………….

Name: Judith Lihemo Ajega Reg. No: ECE/G/01-70420/2022

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance of Masinde Muliro University of Science and Technology a dissertation entitled ***'Machine learning based spectrum sensing for interference reduction in 5g cognitive radio networks.'***

Sign ………………………………….…. Date……………………….

Dr. Raymond Wekesa

School of Engineering and Built Environment

Masinde Muliro University of Science and Technology

Sign ………………………………….…. Date……………………….

Dr. Filbert Ombongi

School of Engineering and Built Environment

Masinde Muliro University of Science and Technology

COPYRIGHT

ACKNOWLEDGEMENT

I am profoundly thankful to God Almighty for all His blessings. I would like to express my gratitude to my parents and my son, Brylee, for their support of my schooling. Their unwavering support, valuable counsel, and prayers during my studies were immensely beneficial. Their affection inspired me to exert much effort.

I am deeply appreciative of my thesis supervisors, Dr. Wekesa Raymond, for his unwavering support and encouragement, as well as for providing me with the opportunity to learn and grow, and Dr. Ombongi Filbert for the time he dedicated to assessing my contributions, identifying issues, and engaging in discussions that facilitated my advancement in both my work and outcomes.

I would like to thank Dr. Okinda Cedrick for his guidance during research methodology.

I wish to express my gratitude to Prof. James Kulubi for his insightful remarks, conversations, and support. I express my gratitude for his constant availability and assistance.

I would like to extend my sincere gratitude to Dr. Owuor George, the chair of the Department of Electrical and Communications Engineering, for his unwavering support and for allowing me the opportunity to complete my research work.

I would like to express my gratitude to my colleagues for the productive discussions we had while preparing my thesis.

ABSTRACT

The rapid increase in global population has driven a surge in users of radio technology, leading to a shortage of available frequency spectrum for wireless systems. To optimize the use of limited spectrum, secondary (unlicensed) users can access the spectrum of primary (licensed) users when it is temporarily unused. These unused portions of spectrum are called spectrum holes or white spaces. Cognitive radios play a key role by performing spectrum sensing to detect when the spectrum is available for secondary users. Real-time spectrum detection is essential for allowing secondary users to access the spectrum without interfering with primary users. However, existing spectrum sensing methods often suffer from poor detection accuracy due to channel fading and noise. The work addressed the creation of machine learning-related algorithms that perform effective user classifications and spectrum sensing in 5G wireless cognitive radio networks to maximize spectrum usage and minimize interference. To classify users according to the patterns of activities and spectral behaviour, a hybrid sequential algorithm that merges Particle Swarm Optimization (PSO) and K-Means clustering were created to classify users. To perform spectrum sensing, a PSO-K Means-based algorithm has also been used to identify the spectrum holes through clustering sensed data in occupied and unoccupied frequency bands. This method employed PSO a population-based optimization method to compute the initial centroids and give an optimal starting point in the clustering. K-means then grouped the sensed spectrum into two occupied and unoccupied. The hybrid PSO-K algorithm increased the primary user detection and allowed cognitive radios the opportunity to access the spectrum without resulting in interruptions. Extensive simulations in Python were conducted to evaluate the performance of the PSO-K algorithm in various 5G network scenarios. Results showed that PSO-K significantly outperformed traditional energy detection methods in terms of both performance and detection accuracy. The algorithm notably enhanced the probability of detection while reducing the probability of false alarms and missed detections. Analysis of detection accuracy demonstrated a 9.3% improvement compared to traditional energy detection techniques.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

CR	Cognitive Radio
CRN	Cognitive radio network
SU	Secondary user
PU	Primary user
PSO	Particle swarm optimization
PSO-K	Particle swarm optimization-k means
5G	Fifth generation
SCR	Software –controlled radio
SDR	Software defined radio
SNR	Signal to noise ratio
ML	Machine learning
SVM	Support vector machine
KNN K-	Nearest neighbour
GMM	Gaussian mixture model
LR	Logistic regression
DT	Decision tree
RF	Random forest
TP	True positive
TPR	True positive rate

FN	False negative
FNR	False negative rate
FP	False positive
TN	True negative
CBS	Cognitive radio base station
ROC	Received operating characteristics
AUC	Area under the curve
ROC-AUC	Receiver operating characteristic-area under the curve
WCSS	Within-cluster sum of squares
OFDM	Orthogonal frequency-division multiplexing
GHz	Giga hertz

CHAPTER 1: INTRODUCTION

1.1: Background

The global population is rising daily, resulting in a growing number of wireless communication users (Cisco, 2020). As our society becomes further reliant on wireless communication networks, there is a necessity for more efficient and dependable technology to accommodate the escalating demand for bandwidth. The valuable resource of the electromagnetic spectrum is encountering scarcity issues. The emergence of accelerated telecommunication platforms has resulted in numerous users vying for this limited resource. This has led to a scarcity of accessible frequency slots for all users, necessitating innovative technology to address spectrum utilization issues and facilitate optimal usage by multiple users.

1.1.1 Software Defined Radio

Software-defined radios are defined as radio systems in which software specifies some or all physical layer functionalities. It processes and converts digital signals using reconfigurable software-based components. The adaptability and versatility of these radio devices set them apart from conventional radio communication systems.

Prior to Software Defined Radio (SDR), radio engineers constructed customized radio systems in which a single radio platform was designed to operate on a specific frequency. The SDR platform is versatile, allowing it to support many signals across various frequencies. Consequently, the number of platforms necessary for communication is significantly diminished. In Traditional Radio, functionality is entirely contingent upon gear. In Software Controlled Radio (SCR), only a restricted set of functions can be modified via software, whereas in software defined radio, a wider array of capabilities is contingent upon software-configurable components. Software Radio achieves its adaptability via software, utilizing fixed hardware (Jondral, 2005).

1.1.2. Fifth Generation Wireless Technology

Fifth Generation wireless technology is a major technology in mobile communication systems, whose aim is to provide ultra-fast speed of data, ultra-low network latency, a higher network capacity and a high degree of reliability. It offers three main service types: Enhanced Mobile Broadband to use the high-speed internet and stream media; Ultra-Reliable Low-Latency Communication to support mission-critical services such as autonomous vehicles and remote surgery; and Massive Machine-Type Communication to enable the connections between large numbers of IoT devices, like smart meters and sensors (Ali and Yigang, 2019).

5G is an evolution in mobile communication, designed to address the growing demand for data-intensive, latency-sensitive, and highly dependable applications. It introduces a paradigm shift in network architecture, leveraging advanced technologies such as Massive MIMO, beamforming, network slicing, and software-defined networking to ensure more efficient and flexible use of network resources.

Operating across a broad spectrum of frequency bands, 5G includes low-band (below 1 GHz) for wide coverage and good penetration, mid-band (1–6 GHz) for a balance between speed and range, and high-band (millimeter wave, above 24 GHz) for ultra-fast data transmission over shorter distances (Ali et al., 2020). However, meeting the diverse performance demands of 5G use cases requires highly efficient spectrum utilization. This paper emphasizes the importance of accommodating various spectrum access models licensed and unlicensed to ensure broad 5G adoption.

1.1.3. Cognitive Radio

The electromagnetic spectrum has been under intense pressure because to the rapid expansion of wireless communication technologies in the past few years. There is a real risk of congestion and shortage of the accessible spectrum due to the increased use of this naturally occurring scarce resource. A communication system's spectrum is the range of frequencies that can be used for transmission. An authoritative body within a specific area is responsible for allocating spectrum. Users must pay the authority in order to obtain a license that allows them to utilize a specific frequency range. The user is granted exclusive access to a specific spectrum band in return. A fixed entity, the wireless spectrum is divided into multiple frequency blocks (Khattab, 2013). Primary users (PU) refers to licensed individuals. They are the only ones authorized to enter these blocks. Every once in a while, the main user isn't making the most of their licensed spectrum. At various periods, less than 5% of the licensed spectrum is really being used (Cisco, 2020). Since the introduction of mobile communication technologies, the demand for spectrum has skyrocketed, making this limited resource increasingly congested and difficult to obtain. The average band occupancy in certain locations is above 25%, while in others it is as low as 0.2%, according to research that aims to examine spectrum occupancy within the 3.45 GHz to 3.65 GHz range (Cotton, 2020). It is reasonable to assume that the main licensed user does not constantly use the spectrum, even though they have heritage rights to it, which causes spectrum holes to form. Unused licensed frequency bands in a particular region at a particular moment are called spectrum holes or white voids (Singh, 2017). In an effort to solve the issue of spectrum scarcity, Mitola and Maguire initially proposed cognitive radio (Mitola & Maguire 1999). According to CR, some communication device may adjust its transmission parameters based on an assessment of the condition of the target frequency channel in a bid to satisfy some performance specifications. It is described that the CR is a device that can dynamically adapt the transmission properties and functionality when it senses its operational spectral environment (Singh, 2017). Primary users (PUs) are those individuals or entities that are the purchasers of the spectrum, which is licensed, and they are the only ones who have the right to use the spectrum. On the other hand, the secondary users (SUs) refer to those persons who have not paid to occupy the spectrum, and have no rights of exclusive use. As it is described in this concept, the unlicensed user, also referred to as a SU, can only use the unused portion of the spectrum when it is abandoned by the primary user. CR technology allows unlicensed SUs to opportunistically use the licensed spectrum bands by dynamically sensing and exploiting the unused or under-utilized frequencies. CR networks are adaptable to changing communication environment and enable coexistence of primary and SUs without exposing undesirable interference.

The primary function of a CR is to analyze the spectrum to detect a spectrum hole and permit a SU to utilize it without causing interference. Spectrum holes denote unoccupied frequency ranges within the radio spectrum. These are the segments of the spectrum that principal users do not utilize at a specific time and location, thereby allowing SUs to access the spectrum without inducing interference. If a spectrum hole exists, the CR must be equipped to transmit inside the vacant frequency range until the PU resumes its communication. The CR must consequently detect the spectrum and modify its broadcast settings accordingly (Yucek, 2009). A CR Network is the network within which a CR functions. CR is essential in 5G networks. 5G technology is defined by rapid transmission speeds, minimal latency, and elevated throughput. CR facilitates the use of unoccupied and underutilized airwaves, resulting in enhanced throughput in 5G networks.

1.1.4. Spectrum Sensing in Cognitive Radio Networks

The key role of CR in a CRN is to do spectrum sensing. The largest requirement of CRs within a CR network is spectrum sensing. Before starting a transmission, one must examine which spectrum occupancy exists and the presence of any other gaps in the spectrum within a given area. This is a vital role in the efficacy of the CR network. CR networks require spectrum sensing, which enables devices to locate and recognize any unused spectrum bands to use opportunistically. It is the determination of the existence of primary or licensed users within a frequency band and whether the band is free to be used by secondary or unlicensed users. CR employs numerous spectrum sensing methods among them being, energy detection, matching filter detection, and cyclostationary feature identification. There has also been the implementation of machine learning (ML) methodologies. The most commonly used technique is that of energy detection that is used to determine the presence of PUs using the energy level of the signal received relative to a set predetermined value. Matched filter detection Matched filter detection finds known PU signals across the spectrum using them as templates. Cyc-stat feature detection makes use of the cyclic nature of the significant user signals to determine their presence. Spectrum sensing is difficult due to the factors capable of compromising the accuracy of detection such as fading, interference, and noise. The spectrum can be sensed in order to use dynamic spectrum access, meaning that CR devices perceive which frequency bands are not occupied such that they can opportunistically use the spare band to improve network efficiency and capacity.

1.1.5. Machine Learning in Spectrum Sensing

The goal of ML is to discover new insights by analyzing previously collected data. A ML algorithm's main goal is to find a formula that, when fed data, produces answers to real-world problems. An evaluation of the system's output in relation to inputs called training data yields the mathematical formula. After learning the ropes during training, the algorithm decides what to do with new input based on its predictions. ML algorithms take features, which are discrete data properties, as input. Various features can be included in each data element, and each feature can be used to identify a certain quality or attribute of the data.

ML has three major subfields, which include supervised, unsupervised, and reinforcement learning. Under the supervised learning paradigm of ML, the algorithm finds the relationship between the inputs and outputs of a dataset by using a set of known input-output pairs. The use of data is labelled data since both inputs and their respective outputs are known (Burkov, 2019). Subcategories are further classification and regression. Regression to a continuous output may be used with any input. To address this issue, the data is aggregated and a model created out of this data. This model can then be utilized to give a prediction of the output should there be any input value (Burkov, 2019). In the pattern recognition problem of classification, each sample in a dataset has to be put into some of the possible a priori classes. Initially, the method is trained with a tagged dataset that contains the category of each element of data. Based on this information we can also train a model to correctly predict the type of any unlabelled input (Burkov, 2019). A ML algorithm is able to learn unsupervised without human intervention and guidance, provided the algorithm gathers data on its own and discovers interesting patterns and insights in the data set. The dataset has not been labelled by anybody. The two predominant types are clustering and dimensionality reduction.

To cluster data, an unsupervised learning algorithm creates a model that divides the unlabelled dataset into groups defined by shared characteristics. A feature vector with fewer features than the input vector is the outcome of a dimensionality reduction technique. Algorithms used in reinforcement learning take in data about their environment and adjust their behavior accordingly. There is a unique "reward" for every deed. Using states as inputs, the algorithm aims to acquire a policy that determines the optimum action or outcome (Burkov, 2019).

1.2: Statement of the Problem

The rapid growth of wireless communication technologies and the emergence of data-intensive applications in the 5G era have led to a substantial increase in spectrum demand. However, most of the spectrum is statically allocated, resulting in inefficient utilization where some frequency bands are used while others remain underused, and others are not used at all. Cognitive Radio technology has emerged as a promising solution to address this challenge by enabling dynamic spectrum access through intelligent spectrum sensing. The effectiveness of CR, however, heavily depends on the accuracy, speed, and reliability of its spectrum sensing mechanisms.

The conventional spectrum sensing methods including energy detection, cyclostationary feature detection and matched filtering are severely limited. These are inefficient operation when the signal to noise ratio is low as well as susceptible to uncertainty of noise. Besides, the techniques are frequently not as flexible as required to meet the highly dynamic and heterogeneous characteristics of 5G environments, where users, devices, and services vie against a limited spectral resource. Moreover, with the high concentration of 5G networks and simultaneous licensed and unlicensed users, the management of the interference will be more critical and complicated.

Recent breakthroughs in machine learning offer novel possibilities in improving spectrum sensing with data-driven schemes that can learn the environment and predict spectrum availability as well as improve accuracy of detection. Nevertheless, applying

ML to spectrum sensing is associated with its challenges such as selection of suitable learning algorithms, feature extraction of noisy data, real-time decision making and scalability in large networks. Moreover, lots of ML-based methods involve independent spectrum detection without taking the reduction of the interference as a fundamental achievement, which is essential in guaranteeing the quality of service in the 5G applications.

Thus, an urgent need is to have a powerful, intelligent and interference-conscious spectrum sensing architecture that exploits machine learning to enhance spectrum-use and minimize interference in 5G cognitive radio networks. This research aims to fill this gap by designing and testing a hybrid machine-learning system, that is, the application of Particle Swarm Optimization (PSO) and K-Means clustering, to improve the performance of spectrum sensing and add to the more efficient and reliable 5G communication schemes.

1.3: Research Objectives

1.3.1: Main Objective

To develop a machine learning based algorithm to optimize spectrum utilization by dynamically selecting the best available spectrum bands subject to real-time spectrum occupancy information.

1.3.2: Specific Objectives

i. To optimize spectrum utilization by reducing inter-user interference.
ii. To develop a machine learning algorithm for user classification.
iii. To formulate a machine learning algorithm for spectrum sensing.
iv. To validate the study by comparing it with other machine learning techniques.

1.4: Significance of the Study

The researchers found the research interesting due to the following reasons. To begin with, it resulted in better use of spectrums. The proposed ML-based solution allowed enhancing the accuracy and performance of spectrum sensing, which consequently led to the enhanced utilization of the spectrum in 5G CR networks, which also translated into the enhanced performance of the network and its capacity. Secondly, the study was supposed to assist in avoiding interference. Proper spectrum sensing in time and precision was considered essential in the prevention of deleterious interference with PUs, and the spectrum sensing based on the ML was estimated to improve the interference avoidance behaviors that ensured a smooth coexistence of the primary and SUs in 5G CR networks. Third, the study was focused on scalability and flexibility whereby ML algorithms can evolve with the changing communication conditions and be used to address the dynamism of the 5G CR networks. The study sought to examine scalability and flexibility of the spectrum sensing methods based on ML. Finally, the study was deemed to be relevant to the industry because spectrum sensing was brought out as a field of research that receives significant emphasis both in academia and among network operators. It was observed that spectrum sensing directly influences the proper use of scarce spectrum resources, and the study results were likely to be useful to network operators, policy-makers, and regulatory authorities to deploy the 5G network.

1.5. Justification of the Study

Spectrum sensing is also an important facet of CR networks due to the ability to detect and harness unused spectrum resource. It was said that through proper sensing and sensing available spectrum bands, CR devices could dynamically adjust their transmission parameters and frequency band in order to prevent the interference and enhance the efficiency of an entire network.

The rationale of the study was on the potential advantages it would offer to CR networks. The study was expected to enhance the overall spectrum detection accuracy, minimize interference with PUs and maximize the spectrum utilization efficiency by applying PSO and K-means algorithms in spectrum sensing. This should have brought about considerable performance and reliability enhancement of CR networks, which will eventually allow wider spectrum access to be dynamically and efficiently.

More so, the proposed study sought to address the current literature in the area of CR networks and ML applications that are significant in 6G applications. Through an extensive research and study on the efficiency of PSO and K-means in terms of spectrum sensing, the objective was to present worthy information and advice on the future research and development in the field.

The remaining structure of the thesis is as follows, chapter two gives the review of the literature, chapter three is the research methodology, chapter four reports results and discussion and chapter five gives the conclusion and suggestions to be given in future work.

CHAPTER 2: LITERATURE REVIEW

2.1: Introduction

The introduction of 5G wireless communication systems marks a significant phase in the evolution of mobile communication networks, characterized by significantly enhanced data rates, ultra-low latency, and massive device connectivity. However, this growth in wireless communication users has led to scarcity of available radio spectrum. Despite the scarcity, several studies have revealed that large portions of the licensed spectrum are underutilized. This observation has sparked increasing interest in dynamic spectrum access techniques, with CR as a key technology for improving spectrum utilization efficiency in modern wireless networks.

Cognitive Radio technology enables the opportunistic access of the spectrum by secondary (unlicensed) users without harmful interference to the primary (licensed) users. Spectrum sensing is one of the most important roles of CR and it provides the ability to see the unused frequency bands which are also known as spectrum holes or white spaces. Good spectrum sensing means that second-user can reliably sense and access channels available and reduce interference with the primary users. Nevertheless, classic spectrum sensing methods, including energy detection, matched filtering and cyclostationary feature detection, are limited by factors like low signal-to-noise ratio (SNR) performance and are not always well-adapted to highly-dynamic and heterogeneous 5G networks.

Introduction of machine learning in the process of spectrum sensing has received significant interest as a solution to these drawbacks. The machine learning processes provide evidence-based solutions that can learn based on past and real-time signal patterns and make smart decisions about spectrum occupancy. There are the ML techniques which include supervised learning, unsupervised learning and reinforcement learning. Hybrid models that integrate optimisation algorithms with clustering/classification methods, including Particle Swarm Optimisation and the K-Means clustering, have demonstrated enormous potential in improving accuracy and flexibility of spectrum sensing. Intelligent spectrum sensing will become an interference reduction method in the 5G network where a significant issue is the reduction in network density as a result of the high network density.

This literature review discusses and examines literature available in the field of spectrum sensing, cognitive radio, integration of machine learning, and interference management in 5G networks. The review discusses development of spectrum sensing methods first and an overview of traditional and ML-based methods is given. It has further gone to talk about how optimization techniques, especially PSO, and clustering techniques, such as K-Means have been used to improve sensing performance. Lastly, it determines research gaps and challenges that this study will seek to overcome by formulating a PSO-K Means based spectrum sensing framework that has been specific to interference reduction in 5G cognitive radio networks.

2.2: Related Work

An energy detection technique for spectral sensing was suggested in the work of Gupta (2019). By measuring the received signal's power, the CR was able to determine whether a signal was present or not using this method. By comparing the output of the energy detector with a specified threshold, spectral gaps were identified. The PU's background knowledge was not required. However, the energy detector was only able to detect PU signals whose energy levels were higher than the predetermined threshold. It was unable to detect situations in which the PU signal's energy level was too low, potentially falling below the predetermined threshold. Choosing the right threshold was difficult since it required knowledge of noise power to achieve a desired detection probability while keeping the false alarm probability below a set level. Precise determination of the noise power was not possible. On top of that, one needed to know the noise variance before setting the threshold. However, noise power varied over time, which made noise estimation even more inaccurate.

Suresh et al. (2018) suggested matched filtering detection. In this method, a matching filter was used in order to identify the PU when the transmitting signal was recognized. Where an unidentified signal matched an identified signal, the inference was that a PU existed in the spectrum. It offered a high signal-to-noise ratio of the specified input. In achieving this, the CR had to have a thorough knowledge of the assured features of the PU such as operating frequency and bandwidth.

George (2018) suggested the use of cyclostationary features to discover features. Signal statistics such as autocorrelation and mean were of periodic nature and gave birth to cyclostationary property. Intentionally choosing periodicity was to simplify spectrum detection. Their proposal utilized the cyclic correlation function in order to identify the presence of a PU at a particular spectrum. This method separated the signal of the PU and rejected the background. The rift between relayed signals and noise was utilized since the noise signals lacked periodic characteristics. This method was used to differentiate different modalities of communication and key consumers. The presence of main users was verified by analysing the periodicity of the received signal with this technique. Although computationally expensive and hard to implement, the cyclostationary feature detection method outperformed the energy detection-based spectrum sensing method at low SNR conditions.

Encouraging results had been demonstrated using ML methods to enrich the spectrum sensing of CR networks. Researchers studied different ML algorithms. Ali (2017) proposed an ML-based spectrum sensing method. A Naive Bayes classifier was his research of a classifier built using the received signal power and the cyclic prefix correlation. Gupta and (2019) utilized the correlation produced by the cyclic prefix in order to overcome SNR wall phenomenon in energy detector spectrum sensing.

In cooperative spectrum sensing, Saber (2020) employed a CR network whose PU signal was just one signal, and support vectors machines and kernel neural networks were used. Energy was one of the features that have been used to classify. The energy of the signal received by each CR in a cooperative CR network was summed together to constitute an energy vector. A k-means clustering algorithm was used to label classes. Supervised ML models such as K-nearest neighbours (KNN), Gaussian mixture models (GMM), and SVMs were used to make decisions in terms of the spectrum state (Bodong and January 2020). Support vector machines approach was better in terms of overall performance compared to all other ML techniques.

Devaraj et al. (2021) authors proposed an adaptive cluster-based heuristic method with cooperative spectrum sensing in 5G cognitive radio networks, which are ineffective in traditional spectrum access, the secondary users sense and report separately. Their approach has been to create dynamic groups of secondary users using a locally sensed spectrum information and use a heuristic algorithm that dynamically chooses cluster heads to organize sensing and reporting to minimize overhead and collision. The method dominates the more conventional algorithms in terms of higher detection accuracy, reducing false alarms, reduced sensing latency, and maximizing the use of slots. It is, however, based on idealized cooperation of the SU without taking into account malicious or non-cooperative behaviors.

It has been proposed to conduct a study in which a method that combines a mutated version of the Modified Whale Optimization Algorithm with a Spiking Neural Network in an orchestrating relay system can be used to detect spectrum holes in cognitive radio networks (Eappen et al., 2021). The methodology is to optimize training process of the Spiking Neural Network with the help of the genetically enhanced modified whale optimization algorithm, which allows greater accuracy in discovering the idle channels in various signal conditions. The authors compare the results of the performance metrics, which are detection probability, false alarm rate and convergence speed with the traditional sensing models through simulations. Its results show that the suggested strategy addresses detection accuracy and false alarms much better than traditional neural models without any optimization, which are characterized by better convergence and resilience. It is also, however, sensitive to parameter tuning to trade off network complexity and optimization efficacy.

Ivanov et al. (2022) conducted a research that examined the relevance of energy detection-based spectrum sensing in ultra-dense network deployed systems. The approach entailed the use of energy detection at cognitive access points and extensive simulations that studied detection probability and spectrum occupancy under a range of distributions of primary-user SNR. The results showed that SNR variations were strongly detected by energy and then the method became unreliable at low SNR.

Gupta et al. (2023) introduced a new energy-detecting algorithm in the context of multi-carrier waveforms in the cognitive radio system including a dynamic threshold mechanism to improve its spectrum sensing. The methodology used by them was to analytically derive a variable energy threshold which was specific to multi-carrier signals and compared its performance to that of conventional fixed-threshold energy detection and matched-filter methods using simulations. The results revealed that the dynamic-threshold algorithm was consistently superior to the conventional detectors in that it yields a high detection probability and better throughput and a low computational complexity.

The paper, Kansal et al. (2022), introduced a new Long Boosted Memory Algorithm that multiplies several Long Short-Term Memory (LSTM) predictors, using an AdaBoost framework in order to boost spectrum sensing in 5G and beyond (6G potential) systems, in under-utilized sub-terahertz bands (0.1 -1 THz). Their approach used time-series as a base to develop an ensemble of weak LSTM models with their outputs weighted to enhance the robustness of prediction. They optimized important parameters such as accuracy, sensitivity and specificity, probability of detection, training and sensing time, as well as, the computational complexity using extensive simulations that explore a variety of SNR scenarios (0–20 dB). The results indicated that the model was much more effective as compared to the models of deep learning and LSTMs.

Meena and Rajendran (2019) suggested a cooperative spectrum sensing-based integrated OFDM-based cognitive radio network to minimize interference and maximize the data throughput in dense networks. Their approach adopted a multi-antenna fusion center that served massive users and coordinated SUs into even K-means cluster to collaborate in sensing. Two-stage multi-slot channel assignment was applied as well as dynamic slot allocation to avoid sensing transmission overlaps. The results showed throughput and capacity improvement.

In Banumathi et al. (2022), the article included an in-depth survey and discussion of cooperative spectrum sensing methods that would be suitable in 5G cognitive radio in the real world. They discussed spectrum sensing methods and considered new CSS designs like the relay-assisted sensing, space-time-frequency coding and user cooperation models in next-generation networks. They found that strong cooperative spectrum sensing enhanced the accuracy of detector and spectral efficiency in high-density 5G networks, over single or individual node sensing. Nonetheless, empirical validation by simulations was missing.

A hybrid deep learning structure that combined long short term memory (LSTM) and extreme learning machines (ELM) networks to perform spectrum sensing in 5G cognitive radio networks was introduced by Mohanakurup et al. (2022). Their approach included the LSTM component to extract and learn temporal features based on spectrum data and the ELM component could cluster sensed signals in idle or occupied bands. The findings showed that the hybrid model was highly accurate and reliable in identifying spectrum holes at different SNR levels as compared to independent traditional statistical models. They, however, did not compare the results with other hybrid models.

Wang (2018) introduced a spectrum sensing technique using signal eigenvalues and a clustering technique. This method trained a signal eigenvalue-based classifier with a K-means-based clustering method; it then applied the classifier to check the presence of the primary user signal.

Arjoune and Kaabouch (2019) suggested a methodology that implied the classification of a dataset with the help of the K-means clustering, which distinguished between two classes of the received PU signal present and absent. The number of ML algorithms used was several such as KNN, KNN, logistic regression (LR), decision tree (DT) and random forest (RF) to classify the received signals in one of the two classes after splitting them. The research only dealt with the AWGN signal and did not indicate the nature of the modulation of the PU signal.

An evaluation of the literature by Janu (2021) analyzed how various clustering methods can be used in spectrum sensing and how effective these methods are. ML-based spectrum sensing methods were discovered to be more adaptable and limited less background information about the sensing environment than the less flexible, more traditional methods. Nonetheless, complex and ambiguous interference in the signal compromised the performance of the clustering algorithms thereby causing low performance of clustering at low SNR as well as the low overall sensing performance in the spectrum sensing system.

Arivudainambi et al. (2021) in their article provided a hybrid spectrum sensing and prediction model of cognitive radio networks that combined clustering algorithms, eigenvalue-based detection, and Bayesian inference to use the spectrum better and reduce the interference with the primary users. Its methodology was a three-stage algorithm; spectrum coordinator: selection with clustering algorithms (K-means, K-medoids and mean-shift), spectrum sensing: maximum to minimum eigenvalue (MME) detecting, and spectrum prediction: a Bayesian approach to prediction of idle time slots based on past sensing information. Simulation findings revealed that the model proposed dramatically improved detection accuracy and the sensing time. MME outperformed the traditional energy detection, particularly when the SNR was low and Bayesian inference was also successful in forecasting availability of spectrum in the future. But, it was based on centralized architecture, which could be scaled down and resulted in latency.

Parimala and Devarajan (2022) introduced a two-phase clustering algorithm, hybridization of Modified Fuzzy C-Means and K-Means algorithm, to enhance spectrum sensing in cooperative cognitive radio networks. The methodology began with fuzzy C - Means clustering that was applied in the analysis and labeling of the training data according to signal characteristics and thereafter the K-Means as an unsupervised classifier that determined the presence or absence of primary users. The Fuzzy C-Means and K-Means offered were found to boost performance indices such as classification accuracy, spectrum utilization and detection probability with low false alarm rate, sensing time and misclassification. The hybrid approach when compared to benchmark algorithms like standalone K-Means, FCM and SVM produced better results with less error convergence, and ability to handle class imbalances. However, the research indicated lower performance under low number of secondary users, lack of real time hardware validation, and limited applicability in mobile or highly dynamic environment.

A research study described in the paper by Kockaya and Developeli (2020) determined how to improve the performance of spectrum sensing in cognitive radio networks by optimizing the detection threshold based on a machine learning algorithm. The methodology took into consideration a thorough analysis of the energy detection and matched filter detection algorithms and the presentation of a new approach to the determination of the threshold using online learning. It employed dynamic adjustment of the threshold with historical data on detection and K-means clustering to classify and analyze the input data to reduce the overall error probability. The results indicated that the suggested algorithm-based threshold selection approach using online learning greatly enhanced the detection performance over a variety of fading channels and in low SNR, compared to the conventional dynamic threshold approaches. The algorithm, however, generated complexity and computation overhead.

Tao et al. (2025) study suggested a hybrid method that included Pearson correlation and memorial K-means clustering. The three major steps in the methodology were to extract features of sensor signals, convert the clustering model into a supervised learning framework via training, and use the memorial K-means clustering in order to save previous cluster centers and to enhance the accuracy of classification. Simulation of realistic network conditions was done in the model under a Rayleigh channel. The strategy minimized noise uncertainty and random initialisation of conventional clustering. The results revealed that the method was better at detecting and lowering the false alarm rate than traditional methods like energy detection or simple K-means. This method however was based on channel conditions that were not moving.

In Fernando and Lăzăroiu (2023), the methodology adhered to the PRISMA framework by integrating the quantitative analysis of the literature as a combination of major scholarly databases. The results have indicated that sophisticated spectrum-sensitive clustering algorithms and smart energy-harvesting systems were central to enhancing spectral efficiency, energy sustainability and network performance in dense CR settings.

Zhang and Xiao (2021) proposed an improved PSO algorithm-based cooperative spectrum sensing scheme to CR. It was based on a methodology that attempts to optimize the fusion weights of the soft fusion in order to combine signal detection information of multiple secondary users (SUs) in a more efficient way. These authors put forward a hybrid PSO strategy where they added immune algorithms, chaotic sequences and adaptive inertia weights to overcome premature convergence as well as achieve greater global search capabilities of the PSO. The results pointed to the fact that the proposed method, despite the adverse conditions, reduced the errors and missed detection probabilities significantly, providing the dependable detection outcomes with reduced SUs relative to other schemes.

Eappen and T (2020) paper discussed the trade-off between energy efficiency and spectrum sensing efficiency of 5G heterogeneous cognitive radio networks (CRNs). The methodology presented a new hybridization Particle Swarm Optimization (PSO) and the Gravitational Search Algorithm (GSA) called Hybrid PSO-GSA to address the optimization problem of energy efficiency in spectrum sensing. In this mixed strategy, mutation and crossover factors were considered to create a middle ground trade-off between exploration and exploitation, to identify spectrum holes with better energy use by optimizing the transmission power, sensing bandwidth and power spectral density. The results showed that the designed Hybrid PSO-GSA algorithm was much better in terms of optimizing the energy efficiency at different parameters as compared to the available PSO methods. The weakness was however, the fact that the proposed PSO-GSA is more complex in terms of time than the traditional PSO.

Shami et al. (2022) conducted a comprehensive survey aimed at giving a detailed review of the Particle Swarm Optimization, which is a well-known swarm-based optimization algorithm. The approach to this paper was a critical literature review, where PSO variants were classified according to the adjustments made to the controlling parameters, hybridisation with other meta-heuristic algorithms like genetic algorithms and differential evolution, and multi-swarm/cooperative strategies. It also explored binary PSO, neighborhood topologies, and use of PSO on feature selection problems. The results have indicated that though the original PSO had a good performance of optimization, it had the problem of premature convergence.

In the article by Gul et al. (2021), the significant interest was given to the improvement of the resilience of cooperative spectrum sensing (CSS) in cognitive radio networks (CRNs) to several categories of malicious users. The methodology suggested a scheme which incorporated the PSO algorithm at the Fusion Center(FC) to detect and counteract the effect of false sensing reports by malicious users. The FC received the local sensing data of secondary users (SUs). The main results showed that the given scheme based on PSO is effective to enhance the performance of spectrum sensing in the context of malicious attacks. Results of simulations, reported as ROC curves, always indicated that the PSO-enhanced fusion techniques had greater detection probability (Pd) and lower false alarm probability (Pf) than other conventional techniques. The paper has indicated that the inclusion of PSO rendered the decision of the FC less prone to misleading reports.

2.3: Conceptual Framework

This conceptual framework focuses on the role of variables in influencing spectrum sensing in 5G cognitive radio network with the aim of reducing inter-user interference. The independent variables are signal to noise ratio and threshold. They are designed in a way to optimize the clustering of signals for an accurate spectrum sensing. The dependent variables depend on SNR and threshold, which together determine how effective the algorithm is. Channel conditions and noise level are the control variables.

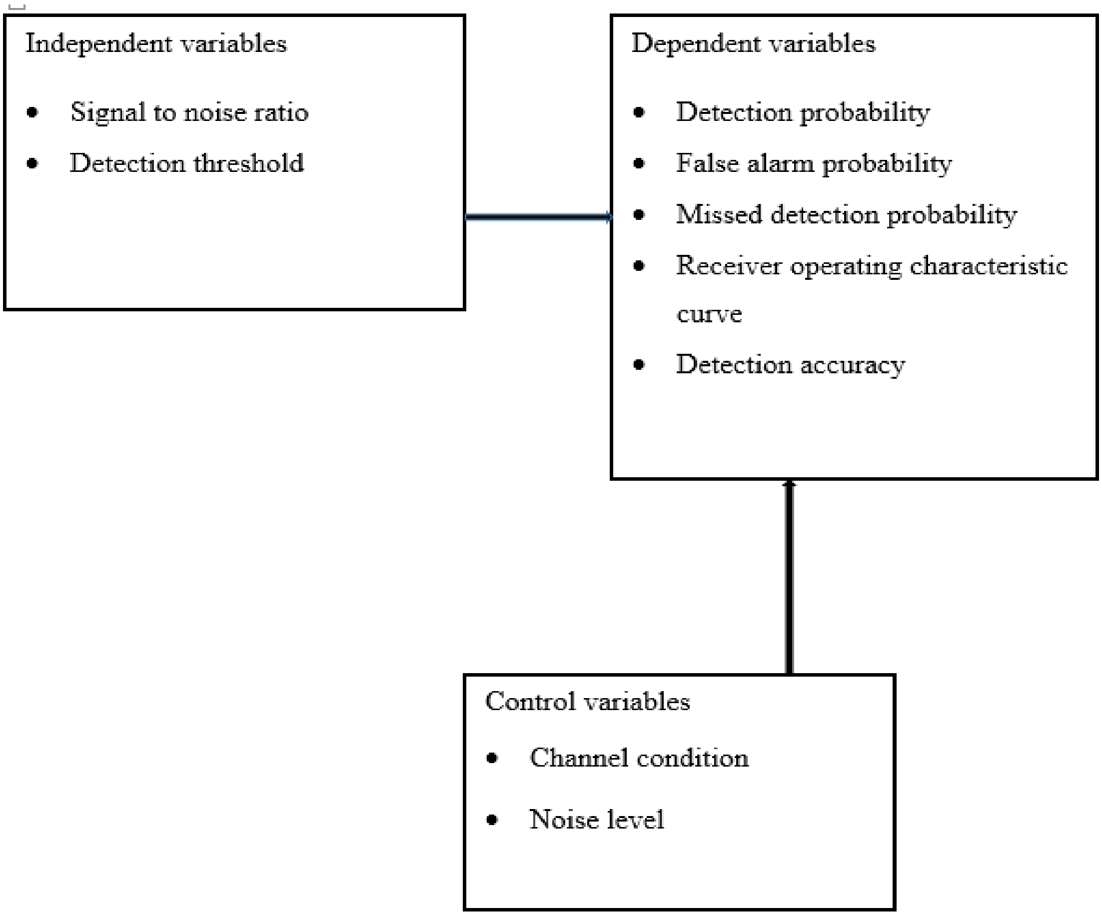

Figure 2. 1: Conceptual framework

2.4: Research Gap

While numerous studies have explored the application of traditional spectrum sensing techniques such as energy detection, matched filtering, and cyclostationary feature detection, these methods often suffer performance degradation in low signal-to-noise ratio (SNR) conditions, require prior knowledge of primary user (PU) signals, or exhibit high computational complexity. Recent research has increasingly focused on integrating ML approaches; SVM, KNN, Gaussian Mixture Models, LSTM, and clustering algorithms like K-means and fuzzy C-means, to enhance sensing accuracy and reduce false alarm rates. Although these ML-based methods show promise, several gaps still remain. First, many existing models rely on centralized architectures that may introduce latency and scalability challenges in dynamic 5G environments. Second, while some studies combine ML with optimization algorithms they often overlook interference reduction as a specific objective, instead focusing primarily on detection accuracy or energy efficiency. Additionally, limited research has been conducted on hybrid clustering and optimization techniques that directly targets both interference mitigation and real-time adaptability. Therefore, there exists a critical need for a scalable, adaptive, and interference-aware spectrum sensing framework that leverages hybrid machine learning techniques to improve spectrum utilization and reduce interference in 5G cognitive radio environments.

In the next chapter, the research methodology was discussed.

CHAPTER 3: METHODOLOGY

3.1: Introduction

This chapter presents the implementation and evaluation of the particle swarm optimization-K-means algorithm for spectrum sensing in 5G CR networks. The PSO-K algorithm combines the global search capability of particle swarm optimization with the local optimization efficiency of K-means clustering. This hybrid approach aims to enhance the accuracy and efficiency of spectrum sensing. The chapter includes the methodology for data generation, the details of the PSO-K algorithm, and the results of the performance evaluation.

CR enables unlicensed SUs to access the licensed spectrum bands opportunistically when they are under-utilized or unused by the PUs.

System Model

The hybrid algorithm presented is PSO-K Means algorithm that combines the PSO with k- means clustering to improve spectrum sensing in a 5G CR network. Figure 3.1 presents a system model of the CR network..

Particle Swarm Optimization-based K-means (PSO-K) method was used in this work to determine the presence or absence of the PU signal in spectrum sensing. This methodology consists of two major components, an optimization algorithm and a clustering algorithm. The detection procedure was enhanced by using energy based feature extraction. The flow chart shown in Figure 3.2, breaks down the working of the PSO-K Means algorithm and provides examples to each scenario.

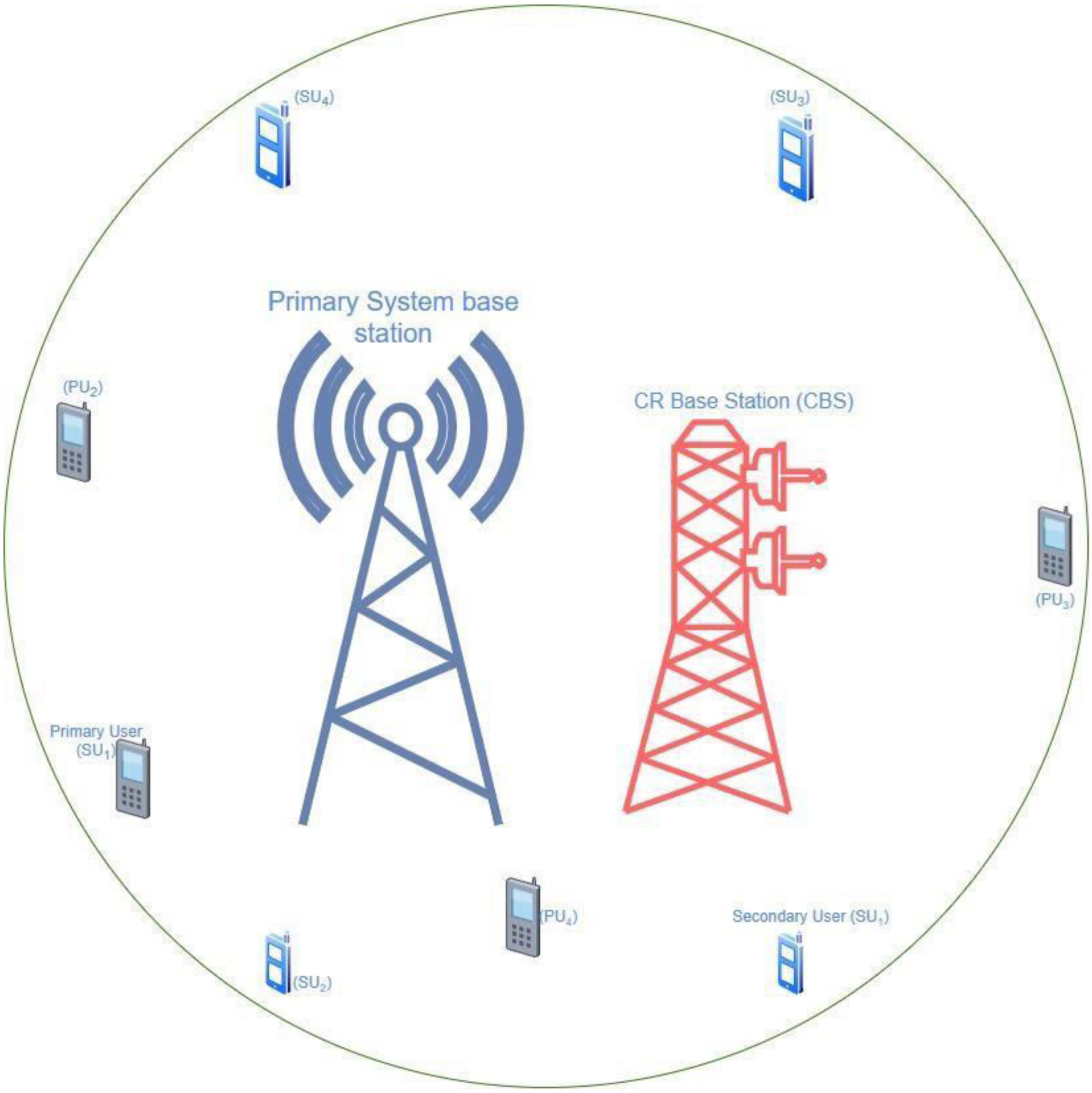

Figure 3. 1: Cognitive radio network system model

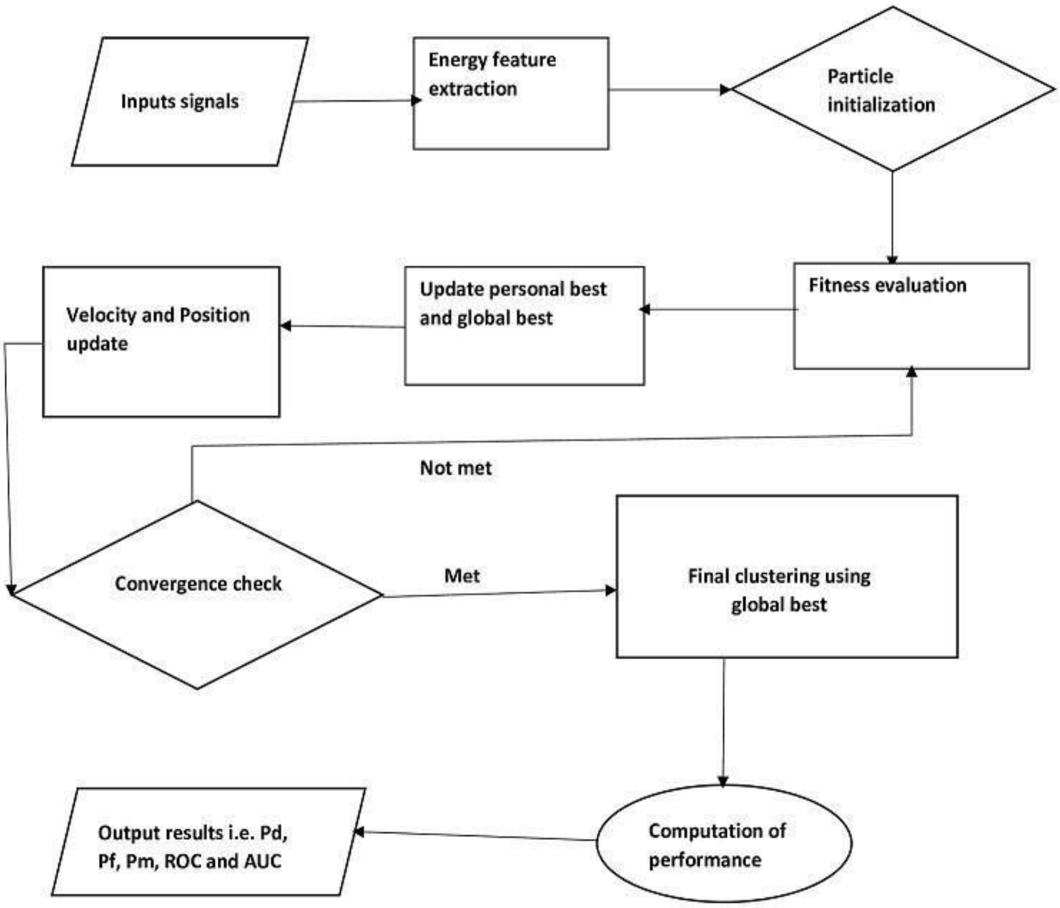

Figure 3. 2: Methodology flow chart

3.2: Simulation Parameters

In this part, the parameters that were employed in simulating the proposed spectrum sensing algorithm in a 5G CR network are given. They were selected in a cautious manner to represent the actual network situations, and at the same time to check the effectiveness of the suggested approach at large scale. The parameters are classified into different groups such as general settings, channel models, algorithm-specific settings and performance metrics. They are correctly chosen in order to simulate realistic and challenging conditions that are usually faced in 5G CR networks.

Detection threshold is manipulated to examine its effect on the key performance indicators, that is, the probability of detection, false alarm rate and probability of missed detection. The trade-offs of spectral sensing will be uncovered by shrewdly adjusting these parameters, and this will give the study a clue on how the proposed algorithm can be tailored within the idiosyncratic network conditions.

3.2.1: Algorithm Specific Parameters

The parameters that were used to simulate the proposed spectrum sensing algorithm in a 5G CR network are presented in this part. To a great extent they were chosen carefully to model the real network conditions, and simultaneously to test the efficiency of the proposed procedure on a large scale. The parameters fall into various categories that include general settings, channel models, algorithm-specific settings and performance metrics. They are properly selected according to the purpose of imitating real and problematic conditions that are typically encountered in 5G CR networks.

Detection threshold is varied in a manner that it investigates its impact on the key performance indicators, i.e., probability of detection, false alarm rate and probability of missed detection. It is the trade-offs of spectral sensing that shrewdly changing these parameters will reveal, and this will provide the study with an indication of how the proposed algorithm can be adapted to the idiosyncratic network conditions.

3.2.2: Channel Model Parameters

The range of SNR 0 dB-20 dB corresponds to different signal conditions and this would enable the proposed PSO-K means algorithm to be experimented with across.

b wide range of noise levels. The AWGN channel model is used because it represents the common random noise in the environment where 5G networks are supposed to be used.

During simulation of the algorithm, the following key variables were used.

Table 3. 1: Simulation parameters

S/N	Parameters	Description	Values
1	Algorithm	Algorithm type	PSO-K means Algorithm
2	Swarm size	Number of particles in PSO	1000
3	Clusters	Number of clusters in k-means	2
4	Iterations	Maximum number of PSO iterations	300
5	Distance Metrics	Type of distance metric	Euclidean
6	Channel type	Type of channel model	AWGN
7	Signal to Noise ratio range	Signal to noise ratio	0-20dB
8	Detection threshold	Variable	01-09
9	Frequency band	Frequency range	0.41 GHz to 7.125 GHz
10	Probability of detection	Ideal value of detection probability	1
11	Probability of false alarm	Ideal value of false alarm probability	0
12	Probability of missed detection	Ideal value of missed detection probability	0
13	AUC	Area under the curve for ideal system	1

3.3: Technical Approach

The means algorithm used was the PSO-K.

PSO-K Means technique is a composite approach whereby PSO is combined with K-Means clustering to enhance clustering effectiveness. The hybrid approach used in this work takes advantage of the global optimization capabilities of PSO to find useful initial cluster centroid which in turn is optimized using local search capabilities of K-Means algorithm. Particle Swarm Optimization is an algorithm of optimization based on the social behaviour of birds in flocks or fish in schools. Individuals of a flock or school are known as particles such as birds and fish respectively. Each member of the swarm is an indication of a potential solution to the optimization problem. Particles move through the solution space in a manner that follows basic principles based on their experiences as well as those of other particles. The movements of each particle change accordingly to these two positions. The most recognized position of a single individual, the personal best (PBest), and the most recognized position of the entire swarm, the global best (GBest).

K-Means is a common clustering algorithm, which splits a data set into K groups. The data are assigned to the cluster which the closest to the mean (centroid) is and the centroids are gradually refined until convergence is reached. K-Means aims to reduce the amount of squared distances between each datum and its cluster centroid which is known as within-cluster sum of squares (WCSS).

The K-means algorithm was optimized with PSO to determine the clustering. It was utilized to determine the neighborhood of the optimum solution by global search. The PSO output was taken as the initial point of the K-Means method which was subsequently applied to refine and generate the ultimate result. PSO algorithm plays a key role in the success of PSO-K Means hybrid algorithm, which has a number of benefits that provide an improvement in the overall process of the clustering. To begin with, PSO is efficient in global exploration that enables it to explore the full solution space to find the best cluster centroids. This worldwide search advantage can circumvent a significant weakness of the classic K- Means algorithm, which is likely to become stuck in local minima because it uses arbitrary starting localized centers. Having realized promising areas where the centroids could be found, PSO significantly enhances the probability of detecting a more precise and stable solution. The other important benefit of PSO is that it has the power to increase the speed with which the clustering process is brought to completion. PSO intelligently directs the swarm of particles to the most appropriate solutions, and thus reduces the search space giving the K-Means algorithm almost optimal initial centroids. This lowers the number of steps taken by K-Means to converge and the hybrid method is more efficient especially when one has large or complex data. Also, the flexibility of PSO enables it to support different datasets with different characteristics. The algorithm is able to change its search strategy dynamically depending on the changing fitness landscape and, hence, it is applicable to clustering tasks in low-noise or high-dimensional data setting. Moreover, PSO and K-Means have been combined so that the hybrid approach will have both the advantages of global optimization (PSO) and local refinement (K-Means) thus makes the clustering of the data even more robust and effective.

K- Means algorithm brings a number of key benefits to the PSO-K Means hybrid algorithm which makes it more effective in the clustering of tasks. The major advantage of K-Means is that it is effective in local optimization. As opposed to the global exploration of the solution space, the PSO algorithm, K-Means focuses on optimization of the clusters after pre-supplying the initial centroids. This repeated process of allocating the data points to the closest centroids and reassigning the centroids makes sure that the clusters are clear that are in close relation to the underlying data structure. Consequently, K-Means effectively optimizes the quality of the final clusters because the variance is reduced in each cluster hence the final results of the clustering become more accurate and meaningful. Besides, K-Means also adds strength to the hybrid approach since it can process large data sets. The simplicity and the scalability of the algorithm allow it to be used to handle large amounts of data without having to incur unnecessary computational costs. The latter is especially useful along with PSO that can provide K-Means with optimal initializations. The hybrid technique can be used to obtain high-quality clustering results also in challenging situations by utilizing the local refinement with K-Means. Also, K-Means improves the efficiency of the process of clustering as a whole. The iterative nature of the algorithm can result in a solution being found quickly once centroid values are initialised and the entire computation time is less than methods that do not use global search techniques at all. This efficiency besides the capacity of PSO to identify good initial centroids assures the effectiveness and efficiency of the hybrid method.

PSO-K Means algorithm relaxes part of the shortcomings of the classic K-Means algorithm, including sensitivity to initial centroid location and local minima traps. The hybrid algorithm steps were as follows, initialization, particle swarm optimization K-means clustering, final clustering, performance measurement.

3.3.1: Data Generation

The initial process is the process of initialization and involves such activities as data generation and collection, feature extraction and particle initialization, the definition of fitness, velocity initiation.

The process of generating or simulating data to be utilized in analysis, experimentation or hypothesis validation is known as the data generation. This is necessary in determining the performance of PSO- K means algorithm. In this research the data generation was conducted to reflect the spectrum environment and PU activity. To begin with, the process began with the definition of the spectrum environment. Spectrum sensing frequency band was defined. In this research the frequency chosen was a 5G mid band with frequency between 3.8 GHz 7 GHz. Determination of Signal Characteristics was carried out. The level of PU transmit power was assumed to be 20dBm. As modulation technique, orthogonal frequency division multiplexing (OFDM) was utilized and Additive White Gaussian Noise was assumed in channels. Secondly, PU Activities were modelled. This has been achieved through development of models of spectrum utilization by PUs. The model entailed regular patterns of the spectrum use. The patterns provided information on the active periods and the non-active periods (spectrum holes). PUs were then synthesized with synthetic signals based on the modulation scheme, power levels and signal characteristics. The signals were generated in Python since the language is rich in libraries and packages, which combine with signal processing. Finally, SU data were produced. The effects of the introduction of the presence of SUs to possibly transmit simultaneously with PUs were simulated.

3.3.2: Data Collection

Data collection can be defined as a systematic procedure of gathering information in order to back research, analysis or decision making. Data concerning the use of spectrum, the characteristics of the signal and network performance were obtained in this work. Such parameters as signal strength, GHz frequency bands and modulation types were taken into account. The spectrum usage scenario was modeled in python as simulation. Data collection parameters were then configured and python configured to execute a series of runs with varying signal- to-noise ratios (SNR) and threshold level. Simulations were operated according to the design and subsequently data gathered and tabulated. Theoretical values were then compared with simulated results with the aim of determining the validity of the simulation. A sample of 5G frequency occupancy state is provided in Table 3.1.

3.3.3: Feature Extraction

Data processing and ML require feature extraction, which is the crucial step to transform raw data into a sequence of features or attributes that can be used to model and classify it. This was a significant step in this study to obtain data in order to identify the presence or absence of PUs.

ML-based feature extraction was used. Initial data was recorded and it comprised of transmission of PU and any noise in the spectrum. It was preprocessed, meaning operations like removing noise and the normalization of the signal amplitude were performed.

Once reprocessing is completed, the pertinent features was then located and mined within data. The features derived on different characteristics of each frequency ranges were the energy levels. These characteristics were refined. (Ratanavilisagul, 2020)

Mean energy level extracted was then calculated. Calculations of these features were carried out as indicated in Equation 3.1.

$$E_n = \frac{1}{k} \sum_{k=1}^{k} \left(z_n(k)\right)^2 \qquad (3.1)$$

Where $(Z_n (k))^2$ is the energy of k^{th} sample.

3.3.4: Initialization

Initialization of particles is a very vital procedure that prepares the optimization process. The swarm size was found first. There was a sample size of 1000 particles. This was followed by the determination of the number of clusters (k) whereby two clusters were used in this study. The former cluster to signify PU present, and the latter cluster to signify PU absent. The number of repetitions were also established and 300 was used. Iterations indicate the highest amount of times that have been carried out before reaching the optimum solution.

Another step that was undertaken is Position Initialization. All the particles of the swarm had a starting random location in the search space. This location was a possible collection of centroid points to the K-means clustering algorithm. The locations were set randomly according to the range of the energy levels of signal.

Velocity initiation was performed. Velocity will dictate the manner in which the particle will explore the search space in subsequent iterations. Besides positions, the initial velocity of each particle was given. Velocities were also randomly started, though within a range so that the particles could test the search space well without going beyond the limits.

The equations 3.2 and 3.3 were used to identify the position and velocity of the particle respectively.

$$V_{id}^{t+1} = w\ x\ v_{id}^{t} + c_1 r_1 \left(p_{id}^{t} - x_{id}^{t}\right) + c_2 r_2 \left(p_{gd}^{t} - x_{1d}^{t}\right) \quad (3.2)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (3.3)$$

where w is the inertia weight, $x_{id}^{t}$and v_{id}^{t} represents the position and velocity of the particle i in the d dimensional space, at time t. c_1 is the self-learning factor, while c_2 is the group learning factor, r_1 and r_2 are random numbers ranging from 0 to 1. They introduce randomness in the movement of particles to promote exploration and prevent particles from being stuck in local minima. p_{id}^{t} is the particle's best value, or local best, and p_{gd}^{t} is the swarm's best value or global best.

3.3.5: Fitness Evaluation

The second step was swarm optimization. And procedure involved fitness evaluation, updating personal and global best positions, position and velocity update, and iteration.

A fitness function is key in optimization algorithms. It is used to evaluate how well a particular solution solves the problem. The fitness of each particle was evaluated based on how well the centroids were represented in the cluster of the dataset. The fitness function is often the within-cluster sum of squares (WCSS). It is the sum of the squared differences between each point in a cluster and the centroid of that cluster. It measures the compactness of the clusters formed by the algorithm. The goal of K- means clustering was to minimize the WCSS, leading to tightly packed clusters with low variance.

The WCSS for each particle in the clusters were calculated following the equation.

$$wcss = \sum_{k=1}^{k} \sum_{x_1 \in k} \left[\frac{(x_1 - \mu_k)^2}{1} \right] \tag{3.4}$$

where k is the number of clusters, x1 represents the data points and μk is the centroid of the cluster ck, x1-μk, is the euclidean distance between data point x_1 and the centroid μk

A lower WCSS indicates that the data points within a cluster are close to each other and the centroid, implying better-defined clusters. It was aimed at minimizing WCSS across all clusters.

Euclidean distance is a distance metric that is used to measure the "straight line" distance between two points in a multi-dimensional space. It is commonly used distance metric in the K-means algorithm to calculate the distance between data points and centroids. The Euclidean distance between two points x=(x1, x2,….,xn) and y= (y1,y2,…,yn) in an n-dimensional space is given by:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_1 - y_1)^2} \qquad (3.5)$$

where x1 and y1 represent the coordinates of the two points in the *i*-th dimension, $(x1-y1)^2$ is the squared difference between the coordinates of the two points.

K-means algorithm will identify the best clustering strategy in minimizing Within-Cluster Sum of Squares (WCSS) of each data point. Reducing the weighted centre-to-point (WCSS) distance that is computed as the sum of the squared distances between points and their centroid points necessitates reducing the EC distances within a cluster. K-means algorithm repeatedly computes the distance in degrees between the data points and the cluster centros. Following that, every point is assigned to the cluster to which it is the nearest in terms of the Euclidean distance formula. The centroids are repeatedly updated in the algorithm on the new cluster assignments and the algorithm is repeated until it converges. At the convergence point of cluster assignments further decreases in Within-Cluster Sum of Squares (WCSS) cannot occur.

The algorithm then successively varies the centroids in such a way that the Euclidean distances between the points and the centroids become minimum, thereby reducing the WCSS. When the reduction of WCSS becomes small, the process is considered to have ended meaning that the clusters have stabilized.

Each particle was then judged in terms of its fitness by computing the energy levels of the respective frequency band (Ratanavilisagul, 2020). Equation 3.6 was used.

$$Fitness = \frac{max\ d_1 y_i}{min\ \ d_2 y_i} \tag{3.6}$$

max (d1 (yi) = max a=1............ Nc ($\sum \forall z \in cij\ \ d(zp, mj)$/cij)

min (d2 (yi)) = min a=1.... *Nc, i≠j (d (mi, mj))*

where max (d1 (yi)) = maximum of average values of distances within same classes in the classification plan exhibited by the particle yi, min (d2 (yi)) = minimum of distances between classes in the classification plan exhibited by the particle yi, mj = the class j and mi = the level.

For the clusters considering Euclidean distance and WCSS, the fitness was expressed as

$$fitness = \frac{1}{wcss}$$

$$= \frac{1}{\sum_{k=1}^{k} \sum_{x_1 \in k} \left[\frac{(x_1 - \mu_k)^2}{1}\right]} \tag{3.7}$$

where k is the number of clusters, x_1, the data point, and μk is the centroid. The algorithm maximizes the fitness value by minimizing the WCSS.

Particles with higher fitness values were desired. The algorithm adjusted the positions and velocities of particles to explore the search space and find the best centroids that minimize WCSS.

3.3.6: Initialization of Personal Best

Personal best (Pbest) refers to the best solution that an individual particle has found during the search process. Each particle represented a potential solution in the optimization problem, and it only knows the best position it had visited so far. Then each particle tracked its best-known position. At the beginning, the initial position of each particle was considered as its pbest. The personal best fitness for each particle was then set to the fitness value obtained from the initial fitness evaluation. The personal best represented the best set of cluster centroids that the particle had identified for classifying the data. It helped distinguish between signals and noise in the spectrum. This step is important because it helped the PSO component of the algorithm guide the K-means clustering toward achieving optimal solution. The personal best was used in the velocity update equation, which determined how each particle moved in the search space.

3.3.7: Initialization of Global Best

The global best (Gbest) is the best position found by any particle in the entire swarm during the optimization process. It represents the most optimal solution that has been discovered up to the current point in time, across all particles. Initially, the gbest was set to the position of the particle with the best fitness value among the initial pBest values. The fitness of the gbest was then evaluated using the objective function, and it was updated as the algorithm was iterated. The global best served as a reference point for all particles in the swarm. Each particle was influenced by both its personal best and the global best when updating its position.

The effect of personal best and global best helped particles converge towards the optimal regions of the solution space. The global best corresponded to the set of cluster centroids that had produced the best clustering results at that time, meaning the most accurate separation of occupied and unoccupied spectrum bands, were achieved, leading to optimal spectrum sensing performance.

3.3.8: Position and Velocity Update

Velocity update mechanism governed how particles in the swarm moved through the solution space in search of optimal cluster centroids for spectrum sensing. Each particle represented a potential solution. The particle's velocity determined how it moved or its position from one iteration to the next. The velocity was influenced by both the particle's personal best position and the global best position found by the swarm. The velocity of a particle was updated using the following relation.

$$v_i^{t+1} = w \, x \, v_i^t + r_1 \, x \, c_1 \, x \, (p_{best} - x_i^t) + c_2 \, x \, r_2 \, x \, (g_{best} - x_i^t) \qquad (3.8)$$

where v_i^{t+1} is the velocity of the particle *i* at iteration t + 1. v_i^t is the velocity of the particle *i* at iteration t. w is the inertia weight, that controls the effect of previous velocity on the current velocity, $c1$ and $c2$ are the acceleration coefficients, controlling the influence of personal best and global best elements. $r1$ and $r2$ are random numbers that are distributed uniformly between 0 and 1. They introduce randomness to the particle's movement. $p_i$best refers to the personal best position of the particle *i.* gbest is the global best position of the particle found by the algorithm. x_i^t is the current position of the particle *i* at iteration *t*.

Once velocity of particles were updated, the position of the particles were then updated.

The position of each particle was updated by adding the updated velocity to the current position. The following relation was used.

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{3.9}$$

where X_i^{t+} is the new position of particle *i*.

The location of each particle would define a possible set of cluster centres that determined the boundary between any two spectrum states, occupied or unoccupied. The velocity update scheme enabled the particles to search through different configurations of these centroids and was directed by both personal bests and global bests of the swarm and resulted in the best achievable detection of spectrum holes or occupied bands.

The update process of velocity in PSO-K algorithm is very important towards directing swarm of particles towards the best solution. It is in between the new solutions and the good solutions known; that the particles converge to an optimal set of cluster centroids to enhance the spectrum sensing. This is essential to a better spectrum detection accuracy and efficiency which translates to improved spectrum utilization..

3.3.9: Iteration

The third was convergence and spectrum decision. It consisted of iteration, followed by clustering, when convergence was reached. One full step through the process of the algorithm in which the positions and velocities of all the particles in the swarm were changed is called iteration. The PSO-K algorithmic steps were repeated until convergence had occurred. In each step, the particles will move around the solution space according to their existing velocity and this velocity was informed by both personal best positions of the particles and the swarm best position. Once all the particles were initialized with with positions, velocities, pbest, and gbest, the algorithm started the iterative process. The counter of the iteration was initially set to zero. The velocity of each particle is updated on the basis of the current velocity, the difference between the current position and current best and difference between current position and global best of the particle. This step guides the particles toward better solutions by balancing exploration and exploitation in the search space. The iterative process involved updating the particles' positions and velocities, adjusting their pbest, and potentially updating the gbest, all aimed at converging on an optimal convergence point. Once the velocities are updated, the particles' positions (which correspond to the cluster centers) are also updated. This new position represents a new potential clustering of the data.

The algorithm was iterated for 250 iterations then a convergence region was achieved. During each iteration, the particles' positions and velocities were updated, their fitness evaluated, and the best solution found by the swarm was also updated.

3.3.10: Convergence

Convergence refers to the process by which the algorithm reaches a stable solution over successive iterations. In other words, convergence is the point at which the algorithm's particles stop making significant improvements or changes in their positions, indicating that the optimal. Convergence in the PSO-K Means algorithm occurred when the particles in the swarm had found an optimal solution, and further iterations could not significantly improve the solution. Several criteria were used to determine convergence.

Fitness Stabilization means that there is improvement in fitness across successive iterations has becomes negligible. It indicated that the particles had converted to a solution and were no longer making significant progress in finding a better one.

Position Convergence is another indicator of convergence. It is when the particles' positions, the centroids, become stable. The centroids were no longer changing significantly across iterations, suggesting that the optimal clustering had been found.

As the particles were converging towards the global best, their velocities tend to decrease. Reduction in velocity across the swarm, indicating that the particles were converging to a common solution and that the search space exploration was nearing completion.

After convergence, the final positions of the centroids after iterations were used to classify the data points into clusters. Clustering was then performed on the set of frequency bands to group them into clusters based on their similarities in energy levels. The K-means algorithm clustered the particles into 2 groups based on their positions in the search space. Equation 3.10 was used.

$$d(z_p m_j) = \sqrt{\sum_{k=1}^{n_d}}(zpk - mjk)^2 \tag{3.10}$$

where Nd is the number of attributes, $_{k=1}$Zp is the object, Zpk is the attribute k in the object p, mj is the class j, mjk is the attribute k in the class j and d (Zp, mj) is the distance between the object p and class j.

Each class was calculated according to equation 3.11

$$m_j = \frac{1}{n_j}\sum_{zp\in cj} Zp \tag{3.11}$$

3.3.11: Spectrum Decision and Performance Measurement

The final clustering result was then evaluated to make a decision to determine clusters that represent occupied frequency bands and those that represent unoccupied frequency bands. After the convergence of the PSO-K Means algorithm and clustering, the next critical step was to measure the performance metrics. In standard measures, these indicators are commonly considered as the probability of detection (Pd), probability of false alarm (Pf), and probability of missed detection (Pm).

Cluster labelling was carried out in such a way that one cluster denoted occupied state and the other one vacant state. This algorithm identifies such clusters in terms of centroid locations, and in one such cluster one centroid is generally the presence of a PU (occupied spectrum) and the other centroid is generally the absence (vacant spectrum).

To evaluate the performance, the results from the algorithm were compared with a ground truth dataset, which contained the actual states of the spectrum bands (whether they are occupied or vacant). This ground truth was essential for calculating the performance metrics.

True Positives (TP) refer to the number of correctly detected occupied spectrum bands. False Positives (FP) refer to the number of incorrectly detected occupied spectrum bands (actually vacant). True Negatives (TN) are those correctly potential vacant spectrum bands.

False Negatives (FN) can be defined as the amount of missed detections where the spectrum was occupied, but the algorithm reported vacuity. These metrics were subsequently applied in assessing the performance of the algorithm in relation to probability of detection, probability of false alarm, probability of missed detection and detection accuracy. Probability of detection measures how correctly the technique can identify presence of PU when it is actually there. Pd is calculated mathematically following the relation in equation 3. 12.

$$Pd = \frac{TP}{TP + FN} \tag{3.12}$$

Where TP is True positive and refers to the number of instances where the sensor correctly detects the presence of a signal. FN is nonexistent, and can be defined as the count of incorrect sensor failures to recognize a signal which is actually present.

Probability of false alarm quantifies the probability of the technique to conclude the spectrum is occupied when not occupied. Mathematically is shown by the formula 3.13.

$$Pf = \frac{FP}{FP + TN} \tag{3.13}$$

where FP is false positive and TN is true negative. The equation represents the ratio of falsely detected occupied bands over all truly vacant bands.

Probability of missed detection represent a case where the technique fails to detect that the spectrum is occupied. The missed is calculated by the equation 3.14.

$$Pm = \frac{FN}{TP + FN} \tag{3.14}$$

FN is false negative and TP is true positive. Probability of missed detection can also be given as Pm = 1-Pd.

3.4: Performance Metrics of a Cognitive Radio Network (CRN)

The subsequent performance measures evaluate the efficacy of a CR network. Pd: the likelihood that the CR accurately identifies the spectrum as occupied.

The CR may mistakenly believe that the spectrum is empty when it is really in use; this is represented by the Pm.

The CR could inaccurately detect an occupied spectrum as an empty one, which is known as the Pf.

Over a range of signal-to-noise ratios (SNRs), receiver operating characteristic curves (ROCs) show the correlation between the Pf and the Pd.

3.5: Performance Evaluation

The states, H0 and H1, are used as sensed states for the absence and presence of primary signals, respectively. The four possible cases for the detected signal are:

Declaring H0 when H0 is true (H0/H0) Declaring H1 when H1 is true (H1/H1) Declaring

H0 when H1 is true (H0/H1)

Declaring H1 when H0 is true (H1/H0)

Accurately identifying a PU signal and a spectral void are the first two cases, respectively. Missed detection and false alarm are seen in the last two scenarios, respectively. The link between the false alarm probability (Pf) and the detection probability (Pd) over different signal-to-noise ratio (SNR) values is shown by receiver operating characteristic curves (ROC). With respect to our proposed spectrum sensing paradigm, the detection probability (Pd) and the false alarm probability (Pf) are the success metrics.

Pd is the probability of correctly recognizing the main user when it is actually present (the band is occupied H1). Because interference with the PU occurs at low Pd values (failed detection), it was desirable to keep the Pd value high. Equation 3.15 was used to calculate the Pd.

$$pd = \left(decision\frac{H_1}{H_1}\right)$$

$$= \frac{N_c}{N} \times 100 \qquad (3.15)$$

where Nc is the number of times in which the signal is detected and N is the number of the captured signals.

Pf is the probability of correctly identifying the PU when they are not actually present (when the band is not occupied, H0) compared to when they are present (when the band is occupied, H1). A high Pf value (failed detection) reduces spectrum use efficiency, so it is crucial to keep the false alarm rate in spectrum sensing to a minimum. Equation 3.16 was used to calculate the Pf.

$$pf = \left(decision\ \frac{H_1}{H_0}\right)$$

$$= \frac{N_e}{N} \times 100 \quad (3.16)$$

where, Ne is the number of times in which the signal is detected, and *N* is the number of captured signals.

The next chapter provides the results of the algorithm and discussions of findings in relation to existing literature.

CHAPTER 4: RESULTS AND DISCUSSION

This section describes the performance analysis and evaluation of the simulation results. The primary objective of this study was to develop an ML-based algorithm to optimize spectrum utilization. To evaluate the performance of spectrum sensing techniques, results were analyzed based on the performance metrics discussed below.

4.1 Optimization of Spectrum Utilization

Clustering was performed, and the performance metric was measured to analyze the performance of the algorithm. Probability of detection and the ROC curve were used to measure the performance of PSO-K means in order to minimize inter-user interference.

4.1.1 Receiver Operating Characteristic Curve

The Detection probability, also known as true positive rate shows the likelihood of CR to correctly detect the presence of a PU in the spectrum (Sarjit, 2024). It is represented by equation 3.12.

The PSO-K means algorithm achieved a high detection probability. In figure 4.1, the Receiver operating characteristic curves shows the true positive rate on the x- axis, against false positive rate on the y-axis. It is clearly seen that the Pd has a value of 0.93 or 93%, which shows that the proposed technique detects the presence of PU by 93% and indicates a robust spectrum sensing capability. The ROC curve generated for the proposed PSO-K Means spectrum sensing system resulted in a high AUC. This high value demonstrates the system's strong ability to distinguish between unoccupied and occupied channel. It validates the usefulness of the hybrid algorithm in maximizing the clustering boundary to achieve proper classification. The proposed method performs better as compared to the conventional energy detection methods, which normally give lower AUC values under the same circumstance. This is better than the conventional energy detection which usually results in a Pd value of approximately 0.7 (Burkov, 2019). The optimum ROC curve with spectrum sensing has an AUC value of 1 or 100%. Usual Pd values in the literature are 0.7 to 0.9 (Burkov, 2019). This study has improved to 0.93, an indication that the suggested approach is more efficient in the PU signal detection.

ROC curve representation assesses the effectiveness of detection system. It shows the trade-off between the system's sensitivity (true positive rate) and its specificity (false positive rate) with varying detection threshold.

True Positive Rate (TPR) or Probability of Detection (Pd) is also called sensitivity, is the proportion of actual positives (presence of a signal) correctly identified by the system. Normally plotted on the Y-axis of the ROC curve.

False positive rate (FPR) or probability of false alarm, also known as 1 - Specificity, is the proportion of actual negatives (absence of a signal) incorrectly identified as positives by the system. It is plotted on the X-axis of the ROC curve.

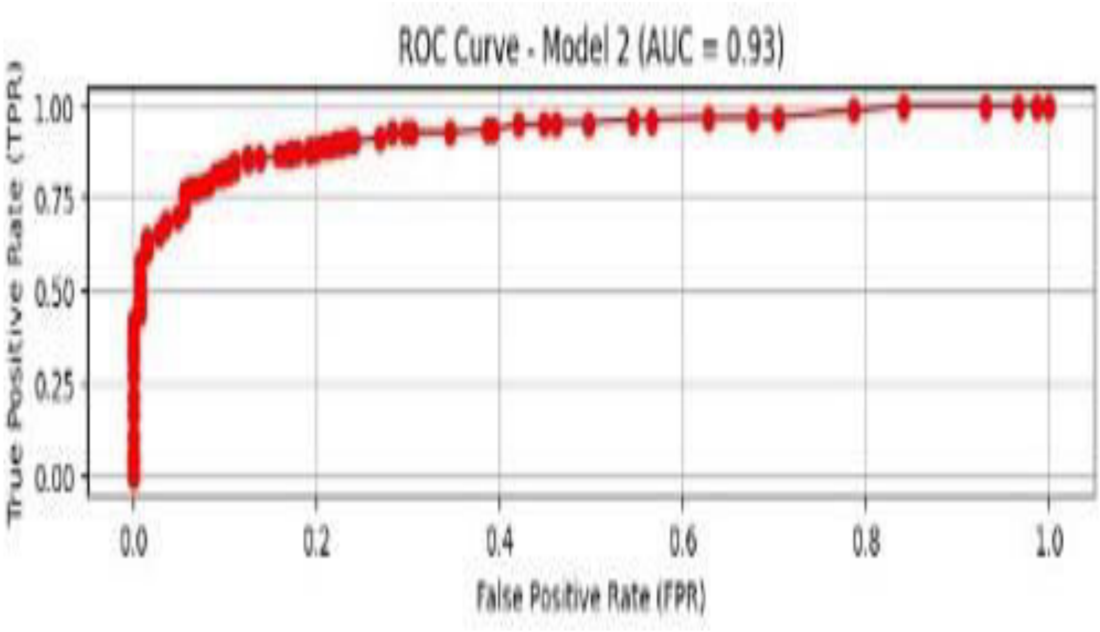

Figure 4. 1Receiver operating characteristics curve for PSO-K

4.1.2 Probability of Detection

Figure 4.2 presents a graph of Pd against detection threshold. It is seen that increasing detection threshold lowers Pd, and vice versa is also true. A lower detection threshold makes the system more sensitive, meaning it is more likely to detect even weak signals from the PU. A higher threshold makes the system less sensitive, making it possible to miss weaker signals from the PU, leading to lower detection probability. It is usually very important to keep the detection threshold to a value that gives a higher Pd .In this study, the detection threshold has been at a value that gives a high Pd, in order to allow accurate sensing while minimizing interference between users.

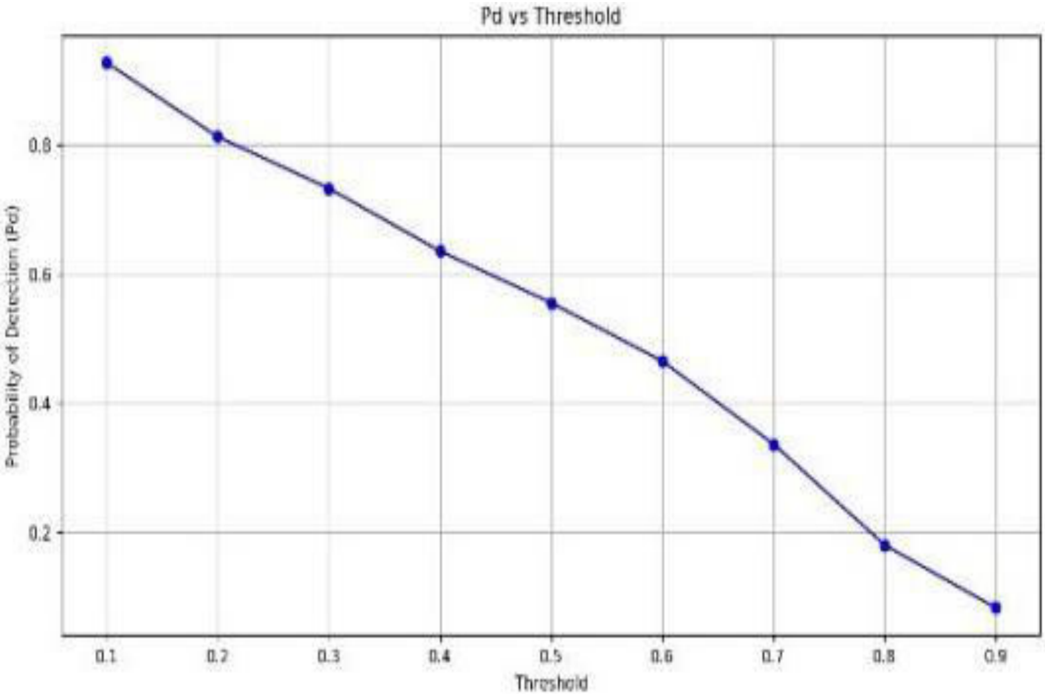

Figure 4. 2: probability of detection vs threshold

4.1.3 Signal-to-Noise Ratio

Noisy environments affect spectrum sensing. SNR is a vital metric for communication networks. An evaluation of the channel's quality is given. When properly implemented, spectrum sensing can reliably provide accurate detection results. Noise, as well as the effects of fading and shadowing on the channel in question, make this practically impossible. A considerable detection probability was achieved across various SNR levels by the PSO-K means approach. Figure 4.3 shows how the signal-to-noise ratio is correlated with the detection chance. Gains in signal-to-noise ratio are directly proportional to increases in detection probability. From 0dB to 5dB, the detection probability (shown in figure 4.3) is very low; nevertheless, it grows exponentially from 5dB to 20dB. This trend demonstrates the PSO-K means algorithm's sensitivity to signal quality, exhibiting best performance in settings with elevated SNR. Suboptimal detection performance in a low SNR channel indicates that this approach is adversely influenced by the channel conditions, preventing the SUs from capitalizing on all available transmission chances.

This high Pd minimizes the risk of interference between the secondary and the PU. In 5G CR network the SU will seldom interfere with the primary leading to better performance. This suggests that the algorithm can effectively identify spectrum holes, allowing for optimal use of the available spectrum

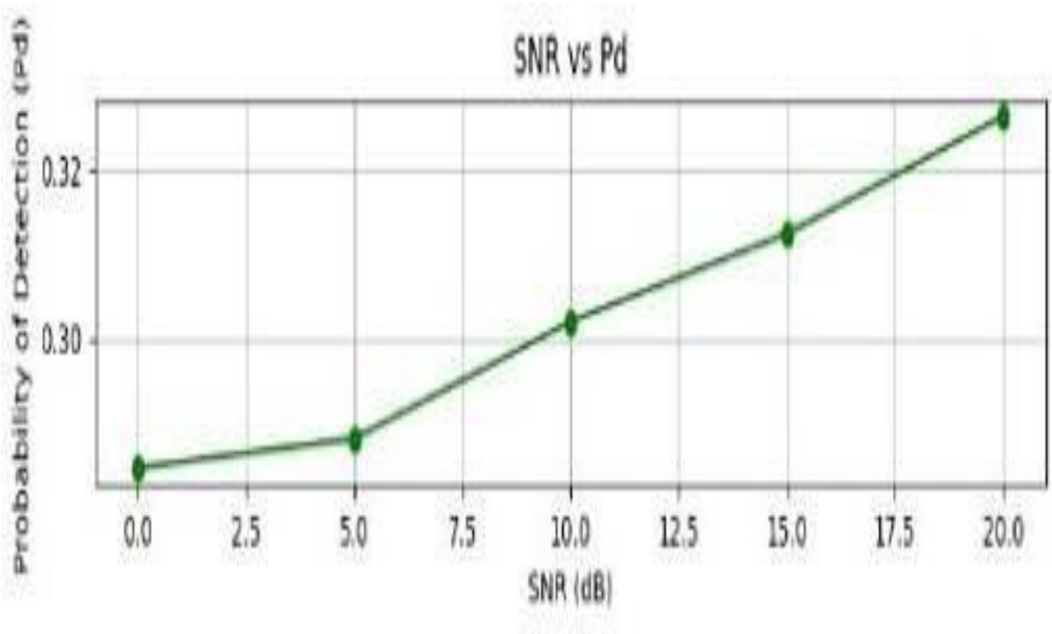

Figure 4. 3: Graph of detection probability vs SNR

Figure 4.4 illustrates that the chance of false alarm is elevated at low SNR values and diminished at higher SNR values. At a high signal-to-noise ratio, the PSO-K means accurately identifies the presence of a principal user.

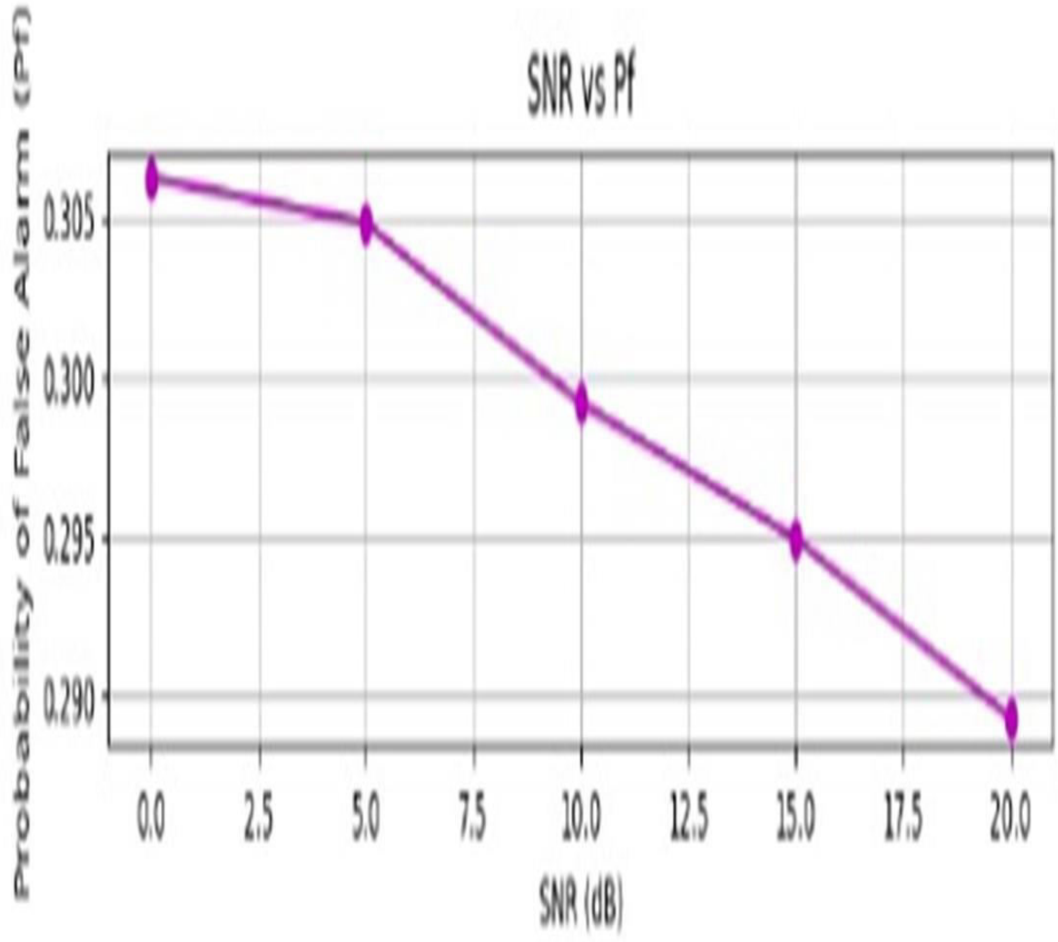

Figure 4. 4: Graph of False alarms probability vs SNR

The performance of the spectrum sensing algorithm was evaluated at different SNR levels. The ML-based approach maintained a high Pd across a wide range of SNRs, demonstrating better resilience in noisy environments. Below 5dB, the Pf is higher, meaning that low SNR affects spectrum sensing. The Pf then decreases compared to SNR. In the literature, traditional methods often struggle at low SNRs, with significant drops in Pd. The improved performance at low SNRs in this study indicates that the ML approach is more effective in challenging environments, making it more suitable for real-world applications in 5G networks.

Missed detection is a barrier for spectrum sensing, as it allows the SU to interfere with the primary signal. Figure 4.5 illustrates that the chance of missed detection is elevated at low SNR values; however, as the SNR value increases, the Pm diminishes. Generally, it is seen that the PM decreases as the SNR value increases.

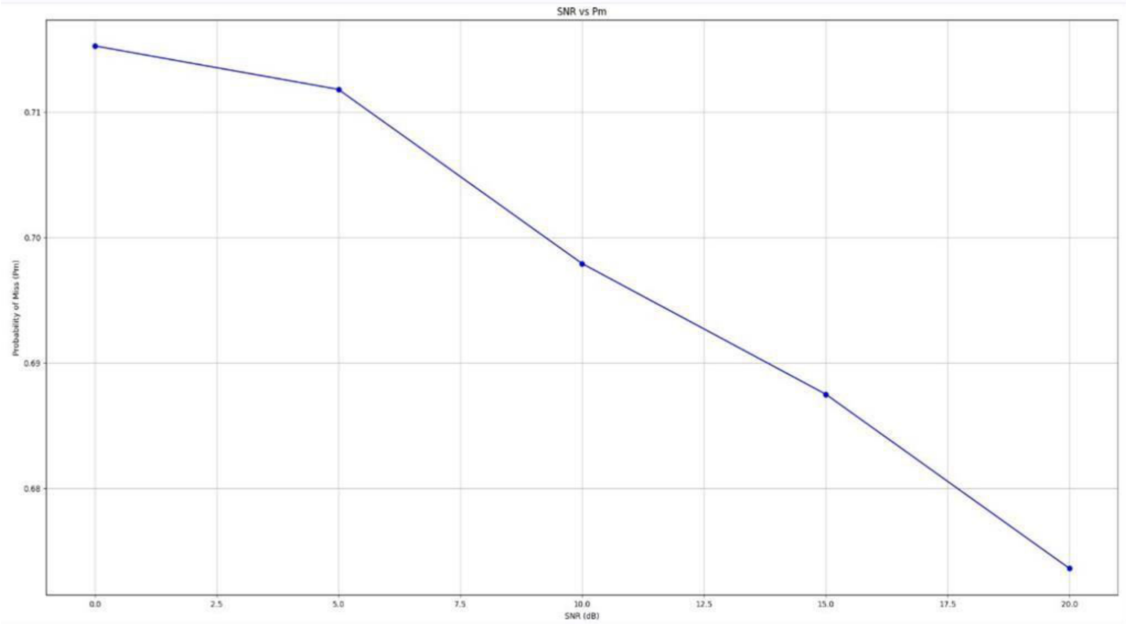

Figure 4. 5: Probability of Missed detection vs SNR

4.2 User Classification and Spectrum Sensing

User classification is the ability to differentiate between primary and SUs. It determines how the spectrum is accessed and managed. Spectrum sensing is the process where SUs detects the presence or absence of PUs in a particular spectrum band. Effective classification helps in making decisions about spectrum allocation, avoiding interference, and optimizing network usage. Probability of false alarm and probability of missed detection were used to analyze user classification and spectrum sensing.

Algorithm for user classification

```
Input:
    X ← user feature matrix (shape: n_users × n_features)
    K ← 2  # PU and SU

PSO Parameters:
    swarm_size ← 1000
    max_iter ← 300
    w_max ← 0.9
    w_min ← 0.4
    c1 = c2 = 2

Initialize:
    For each particle i in range(swarm_size):
        position[i] ← Random centroids (2 × n_features)
        velocity[i] ← zeros_like(position[i])
        p_best[i] ← position[i]

    g_best ← best performing p_best

Function assign_clusters(X, centroids):
    For each user x in X:
        Calculate Euclidean distance to each centroid:
            d = √Σ(x_j - μ_kj)²
        Assign to closest centroid (label 0 or 1)
    Return labels

Function compute_wcss(X, labels, centroids):
    wcss = 0
    For each cluster k in {0,1}:
        cluster_points = X[labels == k]
        wcss += Σ(||x - μ_k||²)
    Return wcss

Function fitness(wcss):
    return 1 / (wcss + 1e-8)  # Higher fitness = better clustering

For t in range(max_iter):
    w = w_max - ((w_max - w_min) * t / max_iter)  # Gradually reduce inertia

    For i in range(swarm_size):
        centroids = reshape(position[i], shape=(2, n_features))
        labels = assign_clusters(X, centroids)
        wcss = compute_wcss(X, labels, centroids)
        current_fitness = fitness(wcss)

        If current_fitness > fitness(p_best[i]):
            p_best[i] = position[i]

        If current_fitness > fitness(g_best):
            g_best = position[i]

        # Update velocity and position
        r1, r2 = random numbers in [0,1]
        velocity[i] = w * velocity[i] \
                    + c1 * r1 * (p_best[i] - position[i]) \
                    + c2 * r2 * (g_best - position[i])
        position[i] = position[i] + velocity[i]

Final_centroids = reshape(g_best, shape=(2, n_features))
Final_labels = assign_clusters(X, Final_centroids)

# Output:
Return Final_labels  # where 0 = Primary User (PU), 1 = Secondary User (SU)
```

4.2.1 Probability of False Alarm

Table 4.1 shows trade-off between the probabilities of detection versus the probability of false alarm. There is a relation among the probability of detection, probability of false alarm and threshold. When the detection threshold was adjusted from 0.2 to 0.5, there was a noticeable decrease in the false alarm rate from 0.79 to 0.5, at the expense of a slight reduction in detection probability. This trade-off highlights the importance of carefully tuning the detection threshold to balance detection accuracy and false alarms. Lowering the threshold lowers the probability of detection as well as the probability of false alarm, while increasing the threshold, increases the probability of detection and the probability of false alarm. A lower detection threshold makes the detector to be very sensitive meaning even small signals are detected. But this increases the likelihood of false alarms when noise is being mistaken for a signal. This leads to a higher probability of false alarm. A higher threshold on the other hand reduces the sensitivity of the detector, which implies that only stronger signals are detected. This decreases the false alarm rate but increases the risk of missing actual signals, resulting in a lower probability of detection.

The proposed method recorded a probability of false alarm of 0.08, which is lower than the 0.12 reported in energy detection methods (Burkov, 2019). A probability of false alarm value of 0.08 demonstrates a significant improvement. This reduction in the probability of false alarm is critical as it minimizes unnecessary spectrum sensing interference and improves the overall efficiency of spectrum utilization.

Table 4. 1: Table of Probability of detection, false alarm rate and detection of threshold

Threshold	Pd	Pf
0.1	0.93	0.88
0.2	0.81	0.79
0.3	0.73	0.69
0.4	0.64	0.61
0.5	0.56	0.50
0.6	0.47	0.42
0.7	0.34	0.29
0.8	0.18	0.21
0.9	0.08	0.10

4.2.2 Probability of Missed Detection

In figure 4.6 a plot of the probability of missed detection versus threshold is presented. It is seen raising detection threshold increases the probability of missed detection, while lowering it decreases the probability of missed detection. Lowering detection threshold increases the system's sensitivity to detecting weak signals and leads to a reduction in probability of missed detection because the system is more likely to detect signals that are weak or close to the noise level, in turn reducing the probability of missed detection.

Conversely, higher detection threshold decreases detector sensitivity meaning that only stronger signals are detected. This increases probability of missed detection as the system is less likely to detect weaker signals. In this work the detection probability is high and ensures that the probability of missed detection is kept low. The proposed technique achieved a probability of missed detection of 0.08, indicating an improvement over traditional methods where probability of missed detection gives typically value of 0.2. This lower probability of missed detection indicates that the ML-based approach is more reliable in identifying spectrum occupancy, which is crucial for avoiding harmful interference with PUs in 5G CR network.

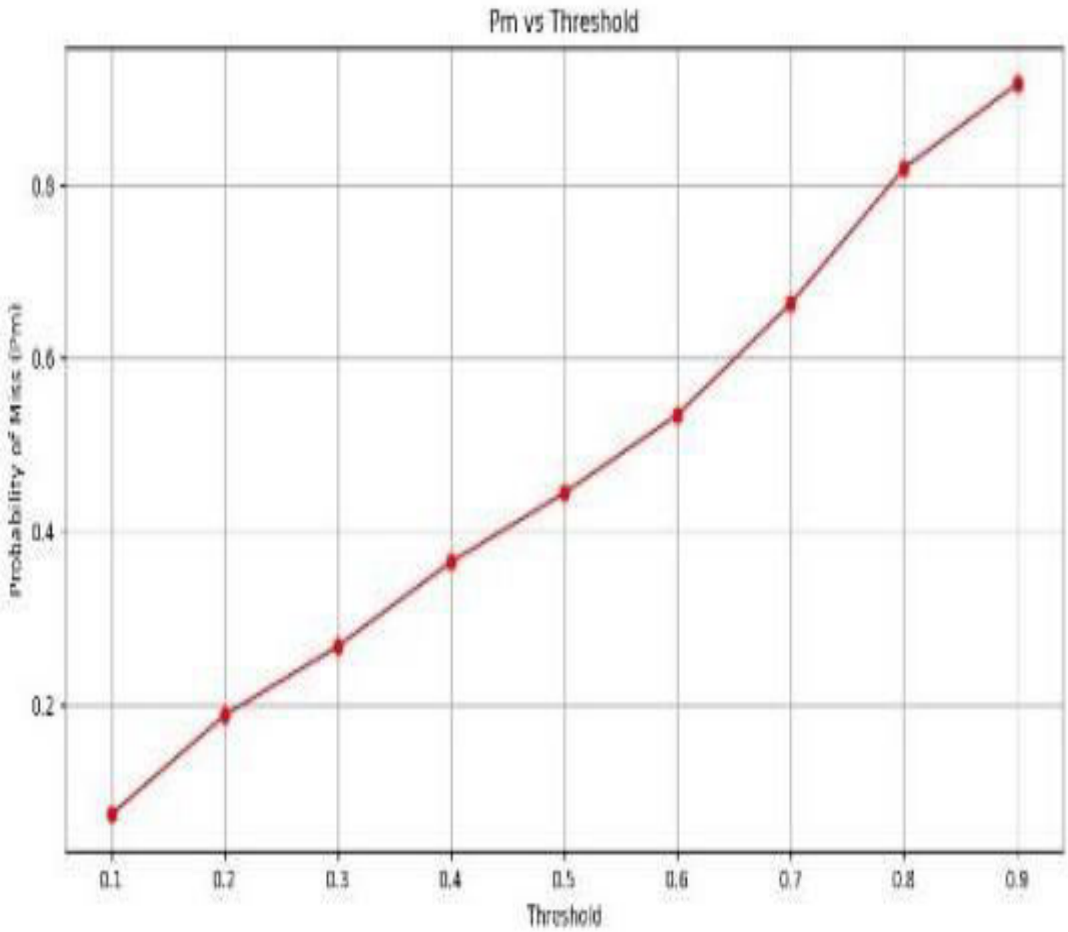

Figure 4. 6: Probability of missed detection vs threshold

In literature, traditional energy detection methods have a higher probability of missed detection especially in low SNRs. The improved performance at low SNRs in this study indicates that the ML approach is more effective in challenging environments, making it more suitable for real-world applications in 5G networks. In 5G CR networks, the proposed algorithm will ensure that interference with PUs is highly minimized.

4.3 Validation of Results

Sensing accuracy in CR networks is the capability of CR system to identify correctly the presence or absence of a PU in a spectrum band. Precise spectrum sensing is highly significant in CRNs and it has a direct impact on both the efficiency of the network, the minimization of interference among the PU and SU, and overall use of the spectrum at hand. To determine the comparison of performance of the proposed detectors, we measured the detection accuracies of the various detectors under the 5G CR network environment.

The fact that sensing accuracy is crucial in the following holds. Firstly, it protects the PU. A primary objective of CR network is to make sure that SUs will not interfere with PUs. High sensing accuracy helps to protect PUs by reliably detecting their presence and vacating the spectrum when PUs want to transmit. Secondly, it increases efficiency in spectrum utilization. Accurate spectrum sensing allows for better utilization of the spectrum by SUs. When spectrum vacancies are detected accurately, SUs can opportunistically access them without causing interference on the PU, thereby improving overall spectrum efficiency. Lastly, accuracy improves network performance**.** In a 5G CR environment, where the demand for spectrum is high and the available spectrum is limited, accurate sensing is crucial in maintaining network performance. It ensures that SUs can access the spectrum only when it is available.

Important elements of sensing accuracy include performance metrics of spectrum sensing i.e. probability of detection, probability of false alarm, probability of missed detection and receiver operating characteristic curve: the roc curve graphically shows the sensing accuracy. The area under the roc curve also is a summary of the total sensing accuracy as AUC of 1 signifies ideal sensing accuracy.

Sandeep, (2018) analyzed spectrum sensing using energy detection method. Also, in (Atapattu, 2019), the author analyzed the area under the receiver operating curve. From their analysis, we see that the accuracy value for energy detection was 0.5, or 50%. In literature, the authors performed a comprehensive analysis of the performance of logic regression algorithm in spectrum sensing based on the detection accuracy. It was observed that the logic regression gave a detection accuracy of 0.85 or 85% when analysis was done on 5G CR networks. In the paper by (Gao, 2021) author proposed a spectrum sensing algorithm based on the random forest. The performance of the algorithm was analyzed under the same environment, where energy levels of the received signal were used as features. In their study, the detection was found to be 0.86 which is 86%.

In this study the PSO-K algorithm has produced high detection accuracy values. Detection accuracy when measure considering the true negatives and false positives gives the PSO-K means algorithm an accuracy of 0.94, or 94%. These results show an improvement in detection accuracy by 9.3%. It implies that application of PSO-K means algorithm in spectrum sensing in 5G CR networks will lead to reduced interference on primary and SU in the CR network.it will also enhance network performance and improve spectrum utilization in CR network. Table 4.3 shows comparison between PSO-K means and other methods of spectrum sensing. The PSO-K means algorithm demonstrated higher accuracy and better performance as compared to other spectrum detection methods.

Table 4. 2: Accuracy comparison table

Algorithm	Detection accuracy
Logistic Regression	0.85
Random Forest	0.86
Energy Detection	0.55
PSO-K means	0.94

4.4: Implications for 5G Cognitive Radio Networks

The results obtained from the proposed ML-based spectrum sensing algorithm show clear improvements over traditional methods in several key performance metrics, including detection probability, false alarm rate, and missed detection probability. The ROC curve and AUC also prove the effectiveness of the chosen approach in the accurate PU signal detection at minimal interference with 5G CR networks. The improvements can be very critical in improving spectrum efficiency as well as reducing interference and hence the suggested approach will be very appropriate in being deployed in the 5G network. The improved spectrum sensing method offered by the PSO-K means algorithm lead to an improved use of the spectrum and less interference in 5G CR networks. That would result in the more stable communication and the improvement of the service quality in areas where the spectrum resources are overly contested.

The following chapter gives the conclusion and work recommendation.

CHAPTER 5: CONCLUSION AND RECOMMENDATION

5.1: Conclusion

The detection accuracy analysis of the proposed PSO-K Means has shown an improvement of 9.3 per cent to the conventional energy detection methods. This improvement highlights the efficiency of integrating machine learning (ML) tools that is the hybrid Particle Swarm Optimization (PSO) and K-Means clustering, in addressing spectrum sensing issues in 5G CR networks. The method enhances the spectrum utilization of the 5G CR networks, which result in the better performance of networks and less interference.

The results indicate that PSO-K Means algorithm improves detection probability, and reduces false alarm and missed detection probabilities significantly. Such gains translate to better user classification so that primary and secondary users are well separated which is minimizes inter-user interference and leads to better overall network performance.

Besides better accuracy, PSO-K Means method has scalability, and thus, it is applicable to real-time spectrum sensing in dynamically 5G surroundings. It can serve as an acceptable candidate to implement in practical CR systems owing to its capability of maintaining high sensing accuracy in different conditions.

5.2: Recommendations

The present study validates the power of the PSO-K Means algorithm when used in an ideal simulated condition, Future research efforts should be directed at investigating how the algorithm works in relation to other wireless fading channel models including Rayleigh, Rician and Nakagami fading. This would bring additional information on its strength and flexibility in real-world implementation of 5G.

REFERENCES

Abdelmonseh Ali, W. H. (2017). Advances on Spectrum Sensing for Cognitive Radio Networks: Theory and Applications. *IEEE Communications Surveys and Tutorials, 19*(2), 1277 - 1304.

Ali, F., & Yigang, H. (2019). Spectrum for next generation technologies. *ICIE*, 188–191. https://doi.org/10.1145/3328833.3328884

Ali, F., Yigang, H., Shi, G., Sui, Y., & Yuang, H. (2020). Future generation spectrum standardization for 5G and internet of things. *Journal of Communications*, 276–282. https://doi.org/10.12720/jcm.15.3.276-282

Aneesh Sarjit, S. M. (2024). Efficient improvement of energy detection technique in cognitive radio networks using K-nearest neighbour (KNN) algorithm. *EURASIP Journal on Wireless Communications and Networking*.

Arivudainambi, D., Mangairkarasi, S., & Kumar, K. a. V. (2021). Spectrum prediction in cognitive radio network using machine learning techniques. *Intelligent Automation & Soft Computing*, *32*(3), 1525–1540. https://doi.org/10.32604/iasc.2022.020463

Arjoune, Y., & Kaabouch, N. (2019). Spectrum sensing, a machine learning method for cognitive radio systems.

Banumathi, J., Sangeetha, S., & Dhaya, R. (2022). Robust Cooperative Spectrum Sensing Techniques for a Practical Framework Employing Cognitive Radios in 5G Networks. *Artificial Intelligent Techniques for Wireless Communication and Networking*, 121–138. https://doi.org/10.1002/9781119821809.ch9

Bodong Shang, L. L. (JanuaryY 2020). Machine Learning Meets Point Process: Spatial Spectrum. *IEEE WIRELESS COMMUNICATIONS LETTERS, , 9*(1).

Burkov, A. (2019). The Hundred page Machine Learning Book.

Cisco. (2020). *Cisco Annual Internet Report (2018–2023) White Paper.*

Devaraj, S. A., Aruna, T., Muthukumaran, N., & Roobert, A. A. (2021). Adaptive cluster-based heuristic approach in cognitive radio networks for5Gapplications. *Transactions on Emerging Telecommunications Technologies*, *33*(1). https://doi.org/10.1002/ett.4383

Dimpal Janu, K. S. (2021). Machine learning for cooperative spectrum sensing and sharing: A survey. *Transaction on Engineering Telecommunications Technologies, 33*(1).

Eappen, G., T, S., & Nilavalan, R. (2021). Cooperative relay spectrum sensing for cognitive radio network: Mutated MWOA-SNN approach. *Applied Soft Computing*, *114*, 108072. https://doi.org/10.1016/j.asoc.2021.108072

Eappen, G., & T, S. (2020). Hybrid PSO-GSA for energy efficient spectrum sensing in cognitive radio network. *Physical Communication*, *40*, 101091. https://doi.org/10.1016/j.phycom.2020.101091

Eldar, Y. C. (2012). *Compressed sensing: Theory and applications.*

Fernando, X., & Lăzăroiu, G. (2023). Spectrum sensing, clustering algorithms, and Energy-Harvesting technology for Cognitive-Radio-Based Internet-of-Things networks. *Sensors*, *23*(18), 7792. https://doi.org/10.3390/s23187792

Gao, Z. W. (2021). Spectrum Sensing Algorithm Based on Random Forest in Dynamic Fading Channel. *Advances in Intelligent Systems and Computing*, 1-5.

Gemi Rachel George, S. C. (2018). Cyclostationary Feature Detection Based Blind Approach.

Gul, N., Ahmed, S., Kim, S. M., & Kim, J. (2021). Robust spectrum sensing against malicious users using particle swarm optimization. *ICT Express*, *9*(1), 106–111. https://doi.org/10.1016/j.icte.2021.12.008

Gupta, M. S. (2019). Progression on spectrum sensing for cognitive radio networks A survey, classification,challenges and future research issues. *Journal of Network and Computer Applications*, 47–76.

Hassan, Y. E.-T. (2012). Learning-based spectrum sensing for cognitive radio systems.

Gupta, J., Gupta, A., Chauhan, S., & Srivastava, R. (2023). A dynamic Threshold-Based Cognitive radio energy detection algorithm for multi-carrier waveforms. In *Algorithms for intelligent systems* (pp. 193–200). https://doi.org/10.1007/978-981-99-5792-7_15

Hwang, C.-H. L.-L.-C. (2010). Spectrum sensingin wideband OFDM cognitive radios. *IEEE Transactions on Signal Processing*, (pp. 709–719.).

Ivanov, A., Stoynov, V., Mihaylova, D., & Poulkov, V. (2022). Applicability assessment of energy detection spectrum sensing in cognitive radio based ultra-dense networks. *AIP Conference Proceedings*. https://doi.org/10.1063/5.0099506

Jan, S. U. (2018). Performance analysisof support vector machine-based classifier for spectrum sensing in cognitive radio networks. *international conference on cyber- enabled distributed computing and knowledge discovery.*

Jondral, F. K. (2005). Software-Defined Radio—Basics and Evolution to Cognitive Radio.*Journal of Wireless Communication Network.*

Kaabouch, Y. A. (2019). On Spectrum Sensing, a Machine Learning Method for Cognitive Radio Systems. *IEEE International Conference on Electro Information Technology (EIT,*. Brookings, SD, USA, 2019, pp. 333-338,.

Kansal, P., Gangadharappa, M., & Kumar, A. (2022). Long boosted memory algorithm for intelligent spectrum sensing in 5G and beyond systems. *Journal of Network and Systems Management*, *30*(3). https://doi.org/10.1007/s10922-022-09652-w

Khattab, A. P. (2013). Cognitive radio networks: From theory to practice.

Kockaya, K., & Develi, I. (2020). Spectrum sensing in cognitive radio networks: threshold optimization and analysis. *EURASIP Journal on Wireless Communications and Networking*, *2020*(1). https://doi.org/10.1186/s13638-020-01870-7

Mani Shekhar Gupta, K. K. (2019). Progression on spectrum sensing for cognitive radio networks: A survey, classification, challenges and future research issues. *Journal of Network and Computer Applications, 147*, 47-76.

Meena, M., & Rajendran, V. (2019). Spectrum Sensing and Resource Allocation for Proficient Transmission in Cognitive Radio with 5G. *IETE Journal of Research*, *68*(3), 1772–1788. https://doi.org/10.1080/03772063.2019.1672585

Michael G. Cotton, L. P. (2020). *" Technical Report TR-20-548, on 3.45–3.65 GHz Spectrum Occupancy from Long-Term Measurements in 2018 and 2019 at Four Coastal Sites,*

U.S. Department of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences.

Mitola, j. (1999). *Cognitive Radio Making Software Radios More Personal.* EEE Personal Communications.

Mohanakurup, V., Baghela, V. S., Kumar, S., Srivastava, P. K., Doohan, N. V., Soni, M., & Awal, H. (2022). 5G cognitive radio networks using reliable hybrid deep learning based on spectrum sensing. *Wireless Communications and Mobile Computing*, *2022*, 1–17. https://doi.org/10.1155/2022/1830497

Muzaffar, M. U. (2012). Learning based spectrum sensing in OFDM cognitive radios.

Nouri, M. B. (2021). A wideband 5G cyclostationary spectrum sensing method by kernel least mean square algorithm for cognitive radio networks. *IEEE Transactions on Circuits and Systems II: Express Briefs.*

Osseiran, A. M. (2016). *5G mobile and wireless communications technology.* Cambridge University Press.

Parimala, V., & Devarajan, K. (2022). Modified fuzzy C-means and K-means clustering based spectrum sensing using cooperative spectrum for cognitive radio networks applications. *Journal of Intelligent & Fuzzy Systems*, *43*(3), 3727–3740. https://doi.org/10.3233/jifs-212863

Ratanavilisagul, C. (2020). Modification fitness function of particle swarm optimization. *International Conference on Engineering and Industrial Technology (ICEIT2020).* Bangkok, Thailand.

Saber M, A. E. (2020). Spectrum Sensing for Smart Embedded Devices in Cognitive Networks using Machine Learning Algorithms. *24th International Conference on Knowledge-Based and Intelligent Information &*, *176*, pp. 1-10.

Saman Atapattu, C. T. (2019). Analysis of area under the ROC curve of energy detection. *IEEE transaction on wireless communication*.

Sandeep., K. (2018). Performance of ED Based Spectrum Sensing Over α–η–μ Fading Channel. *Wireless personal communication*.

Shah, H. A. (2018). Reliable machine learning basedspectrum sensing in cognitive radio networks. *Wireless Communications and Mobile Computing*.

Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, *10*, 10031–10061. https://doi.org/10.1109/access.2022.3142859

Shihao, J. Y. (2008). Bayesian compressive sensing. *IEEE Transactions on Signal Processing,*, 2346–2356.

Singh, S. (2017). COGNITIVE RADIO . *Technology Connect, 3,* (31).

Suresh Dannana, B. P. (2018). Spectrum Sensing Using Matched Filter Detection. *International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)*, *22*, pp. 1-14.

Tao, B., Wu, J., Dou, X., Wang, J., & Xu, Y. (2025). Memorial K-means clustering for cooperative spectrum sensing in cognitive wireless sensor networks at low SNR regimes. *Sensor Review*. https://doi.org/10.1108/sr-10-2024-0883

Yang, J. J. (2018). Low complexity sub-nyquist wideband spectrum sensing for cognitive radio. *IEEE Acces, 2018*(6), 45166–45176.

Yonghua Wang, Y. Z. (2018). Yonghua Wang. *Wireless Communication and Mobile Computing, 2018*, 1-11.

Yucek, T. A. (2009). A survey of spectrum sensing algorithms for cognitive radio applications. *IEEE Communications Surveys & Tutorials, 1*(11), 116–130.

Zhang, J., & Xiao, X. (2021). Soft fusion-based cooperative spectrum sensing using particle swarm optimization for cognitive radio networks in cyber-physical systems. *Concurrency and Computation Practice and Experience*, *35*(13). https://doi.org/10.1002/cpe.6295

APPENDICES

Appendix A: Python Code for PSO-K Means Spectrum Sensing

```
Input:
    - Signal dataset D = {x₁, x₂, ..., xₙ}, each x has features (e.g., energy, frequency, time)
    - Number of clusters k = 2 (Channel Idle, Channel Busy)
    - Swarm size = 1000 particles
    - Max iterations = 300
    - Inertia weight: w_initial = 0.9, w_final = 0.4
    - PSO parameters: c₁ = cognitive factor, c₂ = social factor

Output:
    - Cluster labels (Idle or Busy) for all x ∈ D
    - Final centroid positions (μ₁, μ₂)

Begin

1. Normalize dataset D using Min-Max or Z-score normalization.

2. Initialize swarm:
   For each particle i = 1 to 1000:
       - Randomly initialize centroids: Ci = {μ₁, μ₂}
       - Initialize velocities Vi = {v₁, v₂}
       - Set personal best: pBest_i = Ci
       - Compute initial fitness_i using:
           WCSS = Σ_(k=1)^2 Σ_(x ∈ C_k) [(x - μ_k)²]
           Fitness_i = 1 / WCSS

3. Set global best:
   gBest = pBest_i with the best (highest) fitness

4. For each iteration t = 1 to 300:

    a. Update inertia weight:
        w = w_initial - ((w_initial - w_final) * (t / max_iter))
        // Linear decrease from 0.9 to 0.4

    For each particle i in 1 to 1000:

        i. **K-Means Assignment**:
            For each data point x ∈ D:
                - Compute distance to centroids:
                  d(x, μ) = √Σ_(j=1)^n (x_j - μ_j)²
                - Assign x to the closest centroid

        ii. **Update centroids**:
            For each cluster j in {1, 2}:
                μ_j = (1 / n_j) * Σ_(x ∈ C_j) x
                // Cluster mean
```

```
        iii. **Compute fitness**:
            - Intra-cluster distance (compactness):
                WCSS_i = Σ_(k=1)^2 Σ_(x ∈ C_k) [(x - μ_k)²]
            - Inter-cluster separation:
                d1 = max_(a=1..Nc) (1 / |C_a|) * Σ_(z ∈ C_a) d(z, μ_a)
                d2 = min_(i ≠ j) d(μ_i, μ_j)
                Fitness_i = d1 / d2
            OR use compactness-based fitness:
                Fitness_i = 1 / WCSS_i

        iv. **Update personal best**:
            If Fitness_i > Fitness(pBest_i):
                pBest_i = Ci

    b. **Update global best**:
        If any Fitness(pBest_i) > Fitness(gBest):
            gBest = pBest_i

    c. **PSO Update - Position and Velocity**:
        For each particle i:
            For each centroid dimension d:
                v_id^(t+1) = w * v_id^t
                            + c₁ * r₁ * (pBest_id^t - x_id^t)
                            + c₂ * r₂ * (gBest_d^t - x_id^t)

                x_id^(t+1) = x_id^t + v_id^(t+1)

5. **Final Clustering** using gBest centroids:
    For each data point x ∈ D:
        Assign x to cluster with closest gBest centroid using:
        d(x, μ_j) = √Σ_(i=1)^n (x_i - μ_ji)²

6. **Label clusters**:
    For each cluster j ∈ {1, 2}:
        avg_energy_j = (1 / |C_j|) * Σ_(x ∈ C_j) x.energy

    - Assign label:
        If avg_energy is low → Channel Idle
        If avg_energy is high → Channel Busy

7. Return:
    - Cluster labels for all x ∈ D
    - Final centroids μ₁ and μ₂
    - Channel classification (Idle or Busy)

End
```

Appendix B: Simulation Parameters

S/N	Parameters	Description	Values
1	Algorithm	Algorithm type	PSO-K means Algorithm
2	Swarm size	Number of particles in PSO	1000
3	Clusters	Number of clusters in k-means	2
4	Iterations	Maximum number of PSO iterations	300
5	Distance Metrics	Type of distance metric	Euclidean
6	Channel type	Type of channel model	AWGN
7	Signal to Noise ratio range	Signal to noise ratio	0-20dB
8	Detection threshold	Variable	01-09
9	Frequency band	Frequency range	0.41 GHz to 7.125 GHz
10	Probability of detection	Ideal value of detection probability	1
11	Probability of false alarm	Ideal value of false alarm probability	0
12	Probability of missed detection	Ideal value of missed detection probability	0
13	AUC	Area under the curve for ideal system	1