

EFFECT OF INPUT CHANNEL REDUCTION ON EEG SEIZURE DETECTION

by

Md Shaikh Abrar Kabir

20200255002

MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONIC ENGINEERING



Ahsanullah University of Science and Technology
Dhaka, Bangladesh

July, 2023

This thesis titled, “**Effect of Input Channel Reduction on EEG Seizure Detection**”, submitted by Md Shaikh Abrar Kabir, Roll No.: 20200255002, Session: Fall 2022, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of MASTER OF SCIENCE in Electrical and Electronic Engineering on 30th July, 2023.

BOARD OF EXAMINERS

Mr. Md. Jakaria Rahimi Associate Professor Department of Electrical and Electronic Engineering Ahsanullah University of Science and Technology	Chairman (Supervisor)
Dr. Tareq Aziz Professor & Head Department of Electrical and Electronic Engineering Ahsanullah University of Science and Technology	Member (Ex-Officio)
Mr. Monjur Morshed Associate Professor Department of Electrical and Electronic Engineering Ahsanullah University of Science and Technology	Member
Dr. Md. Taslim Reza Professor Department of Electrical and Electronic Engineering Islamic University of Technology	Member (External)

Candidate's Declaration

This is to certify that the work presented in this thesis entitled, "Effect of Input Channel Reduction on EEG Seizure Detection", is the outcome of the research carried out by Md Shaikh Abrar Kabir under the supervision of Md. Jakaria Rahimi, Associate Professor, Electrical and Electronic Engineering (EEE), Ahsanullah University of Science and Technology (AUST), Dhaka-1208, Bangladesh.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma, or other qualifications.

Signature of the Candidate

Md Shaikh Abrar Kabir
20200255002

Dedication

This research focuses on people who are impacted by epilepsy and the families they belong to. We are motivated to work for improvements in seizure detection and enhance the quality of life for persons with epilepsy by their tenacity, resiliency, and unyielding spirit in the face of difficulties. The epilepsy community, which consists of sufferers and advocacy organizations, is the focus of our work. We are continually inspired to seek out better solutions by audience opinions, and stories of experiences which serve as constant reminders of the value of our research.

Additionally, we dedicate this work to the scientists and researchers who have devoted their whole professional lives to researching epilepsy and creating ground-breaking new technology. Their innovative work opened the door for improvements in seizure detection, providing hope and better outcomes for people with epilepsy. Also, the work is dedicated to the scientists and researchers of the future who will expand on what we have discovered. We believe this study will provide a springboard for increased comprehension, creativity, and advancement in the area of EEG-based seizure detection.

Finally, we commit this study to the goal of creating a world in which people with epilepsy are able to enjoy their lives to their greatest ability without the interruption of seizures. We hope our combined efforts advance the development of more reliable, usable, and efficient seizure detection techniques, thereby enhancing the standard of individuals' life with epilepsy.

Contents

Certification	ii
Candidate's Declaration	iii
Dedication	iv
List of Figures	viii
List of Tables	xii
Acknowledgement	xiii
Abstract	xiv
1 Introduction	1
1.1 Background and present state of the problem	1
1.2 Importance of input channel reduction for practical applications	2
1.3 Research Objectives	3
1.4 Thesis Outline	4
2 Literature Review	5
2.1 Epilepsy: An Overview	5
2.2 Overview of EEG seizure detection techniques	6
2.3 Previous studies on input channel reduction in EEG analysis	8
2.4 Existing methods and algorithms for feature extraction and classification	9
2.5 Gaps and limitations in current research	10
3 State of the art Classifiers	12
3.1 Machine Learning Models	12
3.1.1 Logistic Regression (LR):	12
3.1.2 Gaussian Naive Bayes (GNB):	13
3.1.3 Decision Tree (DT):	14
3.1.4 Random Forest Tree (RFT):	15

3.1.5	K-Nearest Neighbour (KNN):	16
3.1.6	Support Vector Machine (SVM):	17
3.2	Deep Learning Models	18
3.2.1	Vanilla Neural Network (VNN):	18
3.2.2	Convolutional Neural Network (CNN):	19
3.2.3	Gated Recurrent Neural Network (RNN):	21
3.2.4	Convolutional Gated Recurrent Neural Network(C-RNN):	22
3.2.5	Inception Convolutional Gated Recurrent Neural Network (IC-RNN):	22
3.2.6	Convolutional Densely Connected Gated Recurrent Neural Network (C-DRNN):	25
3.2.7	ChronoNet:	26
3.2.8	Convolutional Neural Network - Long Short Term Memory (CNN-LSTM):	27
4	Methodology	29
4.1	Overview	29
4.2	Dataset description and selection criteria	30
4.3	Proposed Classifiers	31
4.3.1	Proposed Model 1 (ChronoNet-M):	31
4.3.2	Proposed Model 2 (CNN-LSTM-M):	32
4.4	Experimental setup and evaluation metrics	36
5	Results and Analysis	38
5.1	Performance Comparison with All Channels	38
5.2	Effect of Single Channel Exclusion	39
5.2.1	Logistic Regression (LR):	39
5.2.2	Gaussian Naive Bayes (GNB):	43
5.2.3	Decision Tree (DT):	46
5.2.4	Random Forest Tree (RFT):	49
5.2.5	K-Nearest Neighbour (KNN):	52
5.2.6	Support Vector Machine (SVM):	55
5.2.7	Vanilla Neural Network (VNN):	58
5.2.8	Convolutional Neural Network (CNN):	61
5.2.9	Gated Recurrent Neural Network (RNN):	64
5.2.10	Convolutional Gated Recurrent Neural Network (C-RNN):	67
5.2.11	Inception Convolutional Gated Recurrent Neural Network (IC-RNN):	70

5.2.12	Convolutional Densely Connected Gated Recurrent Neural Network (C-DRNN):	73
5.2.13	ChronoNet:	76
5.2.14	Convolutional Neural Network - Long Short Term Memory (CNN-LSTM):	79
5.2.15	Proposed Model 1 (ChronoNet-M):	82
5.2.16	Proposed Model (CNN-LSTM-M):	85
5.3	Effect of Multiple Channel Exclusion	88
5.3.1	Performance with Dropping Random Channels	88
5.3.2	Performance with Selected 8 channels	95
6	Discussion	96
6.1	Performance Comparison with All Channels	96
6.2	Effect of Single Channel Exclusion	96
6.3	Effect of Multiple Channel Exclusion	97
6.4	Effect of Selected 8 Channels	97
6.5	Performance of ChronoNet & ChronoNet-M	98
6.6	Performance of CNN-LSTM & CNN-LSTM-M	100
7	Conclusions	103
7.1	Major contributions	104
7.2	Challenges	104
7.3	Future Scope	105
References		106

List of Figures

3.1	Architecture of VNN	19
3.2	Architecture of CNN	20
3.3	Architecture of RNN	22
3.4	Architecture of C-RNN	23
3.5	Architecture of IC-RNN	24
3.6	Architecture of C-DRNN	25
3.7	Architecture of ChronoNet	26
3.8	Architecture of CNN-LSTM	28
4.1	Block diagram of EEG seizure detection process	30
4.2	Electrode placement of the EMOTIV EpoC+ EEG headset (AF: Anterior-Frontal, F: Frontal, FC: Fronto-Central, T: Temporal, P: Parietal, O: Occipital), Black dots indicate the location of reference electrodes.	31
4.3	Architecture of ChronoNet-M	32
4.4	Architecture of CNN-LSTM-M	33
4.5	Flow chart of the machine learning models and VNN	34
4.6	Flow chart of the deep learning models	35
4.7	The electrodes used in the reduced montage [4].	36
4.8	The electrodes used in the reduced montage based on CSP.	37
5.1	Accuracies for Guinea-Bissau dataset with no channel exclusion	38
5.2	Accuracies for Nigeria dataset with no channel exclusion	39
5.3	Accuracies for Guinea-Bissau dataset using LR	40
5.4	Accuracies for Nigeria dataset using LR	40
5.5	Confusion Matrices for Guinea-Bissau dataset using LR	41
5.6	Confusion Matrices for Nigeria dataset using LR	41
5.7	ROC curves for Guinea-Bissau dataset using LR	42
5.8	ROC curves for Nigeria dataset using LR	42
5.9	Accuracies for Guinea-Bissau dataset using GNB	43
5.10	Accuracies for Nigeria dataset using GNB	43
5.11	Confusion Matrices for Guinea-Bissau dataset using GNB	44

5.12	Confusion Matrices for Nigeria dataset using GNB	44
5.13	ROC curves for Guinea-Bissau dataset using GNB	45
5.14	ROC curves for Nigeria dataset using GNB	45
5.15	Accuracies for Guinea-Bissau dataset using DT	46
5.16	Accuracies for Nigeria dataset using DT	46
5.17	Confusion Matrices for Guinea-Bissau dataset using DT	47
5.18	Confusion Matrices for Nigeria dataset using DT	47
5.19	ROC curves for Guinea-Bissau dataset using DT	48
5.20	ROC curves for Nigeria dataset using DT	48
5.21	Accuracies for Guinea-Bissau dataset using RFT	49
5.22	Accuracies for Nigeria dataset using RFT	49
5.23	Confusion Matrices for Guinea-Bissau dataset using RFT	50
5.24	Confusion Matrices for Nigeria dataset using RFT	50
5.25	ROC curves for Guinea-Bissau dataset using RFT	51
5.26	ROC curves for Nigeria dataset using RFT	51
5.27	Accuracies for Guinea-Bissau dataset using KNN	52
5.28	Accuracies for Nigeria dataset using KNN	52
5.29	Confusion Matrices for Guinea-Bissau dataset using KNN	53
5.30	Confusion Matrices for Nigeria dataset using KNN	53
5.31	ROC curves for Guinea-Bissau dataset using KNN	54
5.32	ROC curves for Nigeria dataset using KNN	54
5.33	Accuracies for Guinea-Bissau dataset using SVM	55
5.34	Accuracies for Nigeria dataset using SVM	55
5.35	Confusion Matrices for Guinea-Bissau dataset using SVM	56
5.36	Confusion Matrices for Nigeria dataset using SVM	56
5.37	ROC curves for Guinea-Bissau dataset using SVM	57
5.38	ROC curves for Nigeria dataset using SVM	57
5.39	Accuracies for Guinea-Bissau dataset using VNN	58
5.40	Accuracies for Nigeria dataset using VNN	58
5.41	Confusion Matrices for Guinea-Bissau dataset using VNN	59
5.42	Confusion Matrices for Nigeria dataset using VNN	59
5.43	ROC curves for Guinea-Bissau dataset using VNN	60
5.44	ROC curves for Nigeria dataset using VNN	60
5.45	Accuracies for Guinea-Bissau dataset using CNN	61
5.46	Accuracies for Nigeria dataset using CNN	61
5.47	Confusion Matrices for Guinea-Bissau dataset using CNN	62
5.48	Confusion Matrices for Nigeria dataset using CNN	62
5.49	ROC curves for Guinea-Bissau dataset using CNN	63

5.50	ROC curves for Nigeria dataset using CNN	63
5.51	Accuracies for Guinea-Bissau dataset using RNN	64
5.52	Accuracies for Nigeria dataset using RNN	64
5.53	Confusion Matrices for Guinea-Bissau dataset using RNN	65
5.54	Confusion Matrices for Nigeria dataset using RNN	65
5.55	ROC curves for Guinea-Bissau dataset using RNN	66
5.56	ROC curves for Nigeria dataset using RNN	66
5.57	Accuracies for Guinea-Bissau dataset using C-RNN	67
5.58	Accuracies for Nigeria dataset using C-RNN	67
5.59	Confusion Matrices for Guinea-Bissau dataset using C-RNN	68
5.60	Confusion Matrices for Nigeria dataset using C-RNN	68
5.61	ROC curves for Guinea-Bissau dataset using C-RNN	69
5.62	ROC curves for Nigeria dataset using C-RNN	69
5.63	Accuracies for Guinea-Bissau dataset using IC-RNN	70
5.64	Accuracies for Nigeria dataset using IC-RNN	70
5.65	Confusion Matrices for Guinea-Bissau dataset using IC-RNN	71
5.66	Confusion Matrices for Nigeria dataset using IC-RNN	71
5.67	ROC curves for Guinea-Bissau dataset using IC-RNN	72
5.68	ROC curves for Nigeria dataset using IC-RNN	72
5.69	Accuracies for Guinea-Bissau dataset using C-DRNN	73
5.70	Accuracies for Nigeria dataset using C-DRNN	73
5.71	Confusion Matrices for Guinea-Bissau dataset using C-DRNN	74
5.72	Confusion Matrices for Nigeria dataset using C-DRNN	74
5.73	ROC curves for Guinea-Bissau dataset using C-DRNN	75
5.74	ROC curves for Nigeria dataset using C-DRNN	75
5.75	Accuracies for Guinea-Bissau dataset using ChronoNet	76
5.76	Accuracies for Nigeria dataset using ChronoNet	76
5.77	Confusion Matrices for Guinea-Bissau dataset using ChronoNet	77
5.78	Confusion Matrices for Nigeria dataset using ChronoNet	77
5.79	ROC curves for Guinea-Bissau dataset using ChronoNet	78
5.80	ROC curves for Nigeria dataset using ChronoNet	78
5.81	Accuracies for Guinea-Bissau dataset using CNN-LSTM	79
5.82	Accuracies for Nigeria dataset using CNN-LSTM	79
5.83	Confusion Matrices for Guinea-Bissau dataset using CNN-LSTM	80
5.84	Confusion Matrices for Nigeria dataset using CNN-LSTM	80
5.85	ROC curves for Guinea-Bissau dataset using CNN-LSTM	81
5.86	ROC curves for Nigeria dataset using CNN-LSTM	81
5.87	Accuracies for Guinea-Bissau dataset using ChronoNet-M	82

5.88	Accuracies for Nigeria dataset using ChronoNet-M	82
5.89	Confusion Matrices for Guinea-Bissau dataset using ChronoNet-M . . .	83
5.90	Confusion Matrices for Nigeria dataset using ChronoNet-M	83
5.91	ROC curves for Guinea-Bissau dataset using ChronoNet-M	84
5.92	ROC curves for Nigeria dataset using ChronoNet-M	84
5.93	Accuracies for Guinea-Bissau dataset using mod CNN-LSTM-M . . .	85
5.94	Accuracies for Nigeria dataset using CNN-LSTM-M	85
5.95	Confusion Matrices for Guinea-Bissau dataset using CNN-LSTM-M . .	86
5.96	Confusion Matrices for Nigeria dataset using CNN-LSTM-M	86
5.97	ROC curves for Guinea-Bissau dataset using CNN-LSTM-M	87
5.98	ROC curves for Nigeria dataset using CNN-LSTM-M	87
5.99	Accuracies for Guinea-Bissau dataset using ChronoNet for 12 channels	88
5.100	Accuracies for Nigeria dataset using ChronoNet for 12 channels . . .	88
5.101	Accuracies for Guinea-Bissau dataset using ChronoNet for 10 channels	89
5.102	Accuracies for Nigeria dataset using ChronoNet for 10 channels . . .	89
5.103	Accuracies for Guinea-Bissau dataset using ChronoNet for 8 channels .	90
5.104	Accuracies for Nigeria dataset using ChronoNet for 8 channels	91
5.105	Accuracies for Guinea-Bissau dataset using ChronoNet for 6 channels .	92
5.106	Accuracies for Nigeria dataset using ChronoNet for 6 channels	93
5.107	Accuracies for Guinea-Bissau dataset using ChronoNet for 4 channels .	94
5.108	Accuracies for Nigeria dataset using ChronoNet for 4 channels	94
5.109	Accuracies for Guinea-Bissau dataset using ChronoNet for 2 channels .	95
5.110	Accuracies for Nigeria dataset using ChronoNet for 2 channels	95
6.1	AUC of Proposed Model1: ChronoNet-M	99
6.2	Comparison of models evaluation times	99
6.3	Accuracy comparison of CNN-LSTM-M with other models for 14 channels	101
6.4	AUC of CNN-LSTM-M	101

List of Tables

5.1	Accuracies for selected models with AF3, AF4, FC5, FC6, T7, T8, O1, and O2 channels	95
6.1	Worst accuracies with best performing models	97
6.2	Minimum accuracies for ChronoNet model	98
6.3	Maximum accuracies for ChronoNet model	98
6.4	Comparison of Trainable Parameters	100
6.5	Accuracies for CNN-LSTM and CNN-LSTM-M with 8 channels	100

Acknowledgement

We would like to extend our heartfelt appreciation to everyone who helped this research study on the effect of input channel reduction on EEG seizure detection be completed successfully. Firstly, we would like to express profound appreciation to the epileptic patients who kindly agreed to take part in the study and provided us with their EEG data. The knowledge of seizure detection techniques and their prospective uses has advanced significantly as a result of their contribution. We appreciate the help from the medical personnel and specialists who helped gather the data. They played a critical role in guaranteeing the precision and caliber of the EEG recordings with their knowledge and assistance. We owe a huge debt of gratitude to the creators and producers of the Emotiv EPOC+ device for giving us access to the tools required to collect the EEG data. Our research and the advancement of seizure detection have both benefited greatly from their revolutionary technology.

For the direction, inspiration, and competence throughout the project, we are grateful to our research advisers and mentors. Their insightful suggestions and input have greatly influenced our research strategy and analysis, improving the study's overall quality. We also want to express our gratitude to our coworkers and fellow students who helped us out and gave us advice and support during different parts of our research project. We significantly benefited from their insightful discussions, which deepened our understanding and assisted us in overcoming obstacles.

Finally, we are grateful to our parents and friends for their continuous support, patience, and faith in our capabilities which have been a continual source of inspiration throughout the whole journey. We gratefully thank each and every person who helped make this research endeavor a success for their individual and combined efforts.

Abstract

This research focuses on detecting seizures using electroencephalogram (EEG) signals in a cost-effective manner, especially in resource-constrained healthcare settings like remote areas in less developed countries such as Bangladesh. The study proposes reducing costs by minimizing the number of EEG channels used for diagnosing seizures. This approach can significantly impact overall expenses and make it more feasible to provide diagnostic facilities in remote regions. By using fewer EEG electrodes, the placement becomes simpler, requiring less expertise and time.

Two publicly available datasets of EEG signals from a low-cost consumer-grade EEG headset with 14 channels have been chosen to ensure lower costs during data acquisition and classifier design. Fourteen state-of-the-art classifiers from recent literature have initially been employed to distinguish between epileptic individuals and healthy individuals. The significance of each channel has been assessed by observing the decline in accuracy of all classifiers when excluding a single channel at a time. Furthermore, the effect of excluding multiple channels on the best-performing models has been investigated to identify the optimal electrode configuration. Alongside the data-driven experimental approach, prior knowledge-based methods from recent literature have been utilized to find the best electrode montage. As a result, an 8-channel EEG montage has been selected as the most effective configuration.

Additionally, this research has introduced two novel classifiers based on the two best classifiers, ChronoNet and CNN-LSTM. They are termed ChronoNet-M and CNN-LSTM-M. The impact of channel reduction on these newly proposed classifiers has been thoroughly investigated. The study demonstrates that the novel approach for classifier selection, along with the reduced EEG channels, improves cost-effectiveness and maintains high detection accuracy. 94.2% accuracy has been achieved using ChronoNet with only 8 channel EEG monatge. Among the newly developed classifiers CNN-LSTM-M has achieved higher accuracy (90.55%) than CNN-LSTM (86.13%) with 8 channel EEG montage. The trainable parameters and the evaluation time for the CNN-LSTM-M has been reduced about 78% and 63% respectively.Overall, this research contributes to the advancement of EEG seizure detection systems by providing an efficient and economical solution without compromising accuracy.

Chapter 1

Introduction

1.1 Background and present state of the problem

Epilepsy is one of several neurological diseases that may be monitored and diagnosed using the EEG (electroencephalogram) technology. Recurrent seizures, which are rapid and unexpected blasts of electrical impulses in the brain, are a hallmark of the neurological condition epilepsy, which is a chronic condition. Since it aids in comprehending the features of seizures and keeps track of their frequency and intensity, EEG seizure detection is essential for identifying and treating epilepsy patients. However, because brain activity is so complicated and variable, interpreting EEG data for seizure identification is a difficult undertaking. The identification of seizure occurrences is challenging due to the frequent low amplitude and considerable noise interference of EEG data. Furthermore, significant inter- and intra-individual variability in EEG data need the development of solid and trustworthy algorithms for precise seizure identification [1] [2] [3]. The goal of current research is to solve the difficulties in EEG seizure detection by simplifying the system. In order to extract important characteristics from EEG signals while reducing noise and unimportant data, complexity reduction techniques are used. These methods can improve seizure detection algorithms' effectiveness and accuracy, leading to better epilepsy diagnosis and care. The following two techniques have drawn the attention of modern researchers among the many approaches offered to simplify EEG seizure detection.

- i. Reduction of the number of input data channels: A variety of frequency-domain, time-domain, and time-frequency analysis techniques are used to draw out significant characteristics from a small number of EEG signal channels. Pre-processing of signals to remove noise and artifacts from EEG data, pre-processing techniques in-

cluding artifact removal, filtering, and baseline correction are used. This improves the signal's quality and raises the signal-to-noise ratio, making seizure detection more precise [4] [5] [6].

- ii. Complexity reduction of classifiers: Advanced machine learning algorithms, including support vector machines (SVM) [1] [7], random forests [1] [8], and deep learning models [1] [9], are employed for automated seizure detection. These algorithms categorize EEG data into seizure and non-seizure categories using the derived characteristics, enabling real-time seizure detection and event prediction.

Although complexity reduction approaches for EEG seizure detection have made tremendous progress, there are still problems to be solved. EEG signals have a significant degree of variability and are subject-specific, which makes it challenging to get high accuracy and dependability in practical situations. Future studies will focus on creating reliable algorithms that can accommodate individual variances, adapt to shifting brain dynamics, and enhance seizure detection systems' overall performance.

1.2 Importance of input channel reduction for practical applications

A useful method for identifying and tracking neurological problems, including seizure detection, is electroencephalography (EEG). EEG devices have historically used a large number of channels to precisely record brain activity. However, there is an increasing demand for workable and affordable alternatives that speed up the EEG procedure without sacrificing accuracy. This article examines the significance of number of input channel reduction in real-world EEG applications, emphasizing its advantages for faster detection, reduced expertise requirements, and cost-effectiveness.

- Faster Detection: In time-sensitive situations when rapid diagnosis and action are essential, seizure detection is a crucial component of EEG analysis. The processing and interpretation of EEG data may be greatly sped up by lowering the number of input channels. It is possible to detect seizures in real-time or very close to real-time by efficiently allocating computer resources when there are fewer channels to process. Faster reaction times enabled by this faster detection enable healthcare workers to deliver timely interventions and enhance patient outcomes.

- Reduced Expertise Requirements: Neurologists or skilled technicians often possess the particular knowledge and ability needed to interpret and analyze EEG data. However, the lack of access to these professionals could cause delays in treatment and diagnosis. Reducing the number of input channels in EEG systems can free up expert resources. The complexity of interpreting data is decreased by employing fewer channels, making it more approachable for a larger spectrum of healthcare practitioners. As a result, early assessments can be carried out by doctors with less specialist knowledge, possibly increasing the efficacy and availability of EEG-based seizure identification.
- Cost-effectiveness: Large-channel traditional EEG systems can be expensive to maintain and buy new equipment for. EEG technology is more affordable and available to medical facilities with limited resources due to input channel reduction. The price of equipment and related parts, including amplifiers and electrodes, can be greatly decreased by using fewer channels. Additionally, the cost of training and maintenance is decreased due to the streamlined setup and easier data processing. Because of its affordability, EEG technology is being used more widely in different healthcare settings, which increases the number of people who may use it.

1.3 Research Objectives

The main objectives of our study are:

- To analyze the performance of various state-of-the-art classifiers used for disease diagnosis from EEG signals.
- To investigate the effect of input channel reduction on the performance of selected state-of-the-art classifiers.
- To find the optimal number of input channels while decoding EEG signal with the least performance degradation of the different ML models.
- To optimize the performance of selected classifiers through necessary modifications.

This research may lead to the creation of an automated, low-cost EEG seizure detector for Bangladesh (and similar nations).

1.4 Thesis Outline

This document is divided into five chapters beyond the introduction. Chapter 2 presents a selected literature review on seizure prediction focusing on recent advances and challenges. Chapter 3 showcases the state of art classifiers for EEG-based seizure prediction. Chapter 4 describes the various methods employed throughout the experimental work, including signal processing, machine and deep learning approaches. It introduces 2 novel classifiers. Chapter 5 reports not only the results obtained from the machine learning algorithms but also their statistical validation. Chapter 6 showcases a thorough discussion of the obtained results and other aspects of this thesis. Chapter 7 presents a conclusion and addresses future perspectives in this field of study.

Chapter 2

Literature Review

In a general pediatric critical care unit (PICU), neurological conditions and symptoms account for approximately a fourth of admissions [10]. The cornerstone of the therapy of these individuals is ongoing neurological monitoring [11] [12]. This includes continuous EEG (cEEG) monitoring, mainly to assess the state of consciousness/encephalopathy and to identify seizures [13] [14]. Electroencephalography (EEG) first recorded in the year of 1924 [15], is a widely used noninvasive method applied to numerous domains from brain–computer interface [16] [17] [18] [19], to emotion [20] [21], to cognition [22], to brain diseases [23]. Epilepsy, a neurological issue, has been considered a worldwide problem and is one of the principal dangers to human lives. The seizure's symptoms can vary broadly [24] [25].

EEG is a well-liked diagnostic tool despite having a poorer spatial resolution than brain imaging methods like MRI and CT but superior temporal resolution, cost, and non-invasiveness. However, analyzing EEG data takes time and requires the assistance of qualified researchers. A rising number of people are interested in automated EEG interpretation using machine learning methods to solve this. It is possible to lessen the workload of neurologists by automating the first stage of EEG interpretation, which entails identifying aberrant or regular patterns of brain activity.

2.1 Epilepsy: An Overview

The International League Against Epilepsy (ILAE) described epilepsy as a brain illness in 2005 that is characterized by a recurrent propensity to have epileptic seizures and the associated cognitive, psychological, and social consequences. A brief episode of aberrant and excessive neuronal activity in the brain is known as an epileptic seizure. Due to

the limits of these definitions, the ILAE Task Force proposed a more useful operational definition in 2014. This description covers a number of situations, such as having many seizures spaced by 24 hours or longer, having a single seizure with a significant chance of recurrence, or being diagnosed with a particular epilepsy syndrome. This innovative method makes rapid treatment commencement possible while acknowledging the possibility of seizure recurrence following a single event. Importantly, epilepsy may not be a lifelong condition, as it can be considered "resolved" if a person remains seizure-free for a decade, with a minimum of five years off Anti-epileptic drugs (AEDs) or upon surpassing the age associated with age-dependent epilepsy syndromes. Nonetheless, even when resolved, there remains a possibility of its return.

2.2 Overview of EEG seizure detection techniques

EEG (electroencephalography) seizure detection techniques refer to the methods and algorithms used to identify and classify epileptic seizures from EEG signals. EEG is a noninvasive method that measures the electrical activity of the brain through electrodes placed on the scalp. Seizure detection is crucial in the diagnosis and treatment of epilepsy, as it helps in understanding the frequency, duration, and characteristics of seizures.

There are several approaches and techniques used for EEG seizure detection, ranging from basic signal processing methods to advanced machine learning algorithms. Here is an overview of some commonly employed techniques:

Visual Inspection: This is the simplest form of seizure detection, where EEG experts visually examine the recorded EEG signals to identify abnormal patterns indicative of seizures. However, this method is subjective and relies heavily on the expertise and experience of the interpreter.

Time-Domain Analysis: In EEG tests, electrodes are associated with the scalp using a paste-like substance or cap. The electrodes register the electrical activity of the brain [26] [27] [28]. So, time-domain data are accumulated from the electrodes. Various time-domain features are extracted from the EEG signal to capture seizure-related characteristics. These features include amplitude, duration, slope, and statistical measures like mean, variance, and skewness. Threshold-based techniques or pattern recognition methods can be used to classify seizures based on these features.

Frequency-Domain Analysis: EEG signals are transformed into the frequency domain using techniques such as the Fourier Transform or Wavelet Transform. Frequency-

domain features, such as power spectral density, spectral entropy, or coherence, are then extracted. These features can be used in combination with classification algorithms to detect seizures.

Statistical Analysis: Statistical methods like autoregressive modeling, hidden Markov models, or Gaussian mixture models can be used to analyze the statistical properties of EEG signals. By modeling the normal and seizure states, these techniques can help identify deviations from the baseline and detect seizures.

Machine Learning Approaches: Machine learning technology has benefited to diverse domains in our modern society [29] [30]. With the advancements in machine learning, various algorithms have been applied to EEG seizure detection. Supervised learning techniques like support vector machines (SVM), random forests, or artificial neural networks (ANN) can be trained on labeled datasets to classify EEG signals as seizure or non-seizure. Unsupervised learning techniques like clustering or anomaly detection can also be used for seizure detection without prior labeling.

Deep Learning: Deep learning, a subcategory of machine learning technology, has been showing excellent performance in pattern recognition [31], dramatically improving classification accuracy. The research outcomes of deep learning in speech recognition [32] and computer vision [33] have been successfully utilized to develop practical application systems, which are remarkably influencing our life and even changing our lifestyle. Although EEG domain is far behind compared to the domains, such as computer vision [34] and speech recognition [35] in terms of adopting deep learning, significant progress has been achieved in the last decade. Deep learning models, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in EEG seizure detection. These models can automatically learn hierarchical representations from EEG data, enabling more accurate seizure detection and classification.

It's worth noting that EEG seizure detection techniques are typically combined with pre-processing steps, such as noise removal, artifact removal, and feature normalization, to enhance the accuracy of detection. Additionally, the performance of these techniques depends on factors like the quality of the EEG recordings, the specific characteristics of the seizure types, and the availability of annotated datasets for training and evaluation.

Overall, EEG seizure detection techniques play a vital role in the diagnosis, monitoring, and treatment of epilepsy, assisting healthcare professionals in understanding and managing this neurological condition. Ongoing research aims to develop more robust and efficient algorithms to improve the accuracy and efficiency of EEG seizure detection.

2.3 Previous studies on input channel reduction in EEG analysis

Neurological diseases and symptoms account for a significant portion of admissions to a general pediatric intensive care unit (PICU). Continuous neurological monitoring, including continuous EEG (cEEG) monitoring, is crucial for managing these patients. cEEG is primarily used to assess the state of consciousness/encephalopathy and identify seizures, which can indicate potentially treatable conditions. Prompt diagnosis and treatment of seizures are important in critically ill children, but their clinical identification can be challenging, especially in cases of non-convulsive seizures or when neuromuscular blockade is used [4].

cEEG is often the only way to detect seizures in such situations. However, its use increases the workload of ICU staff, particularly in terms of electrode placement, which is time-consuming. To address this, a reduced EEG montage with fewer electrodes is sometimes used to save time and increase availability. However, using a reduced electrode set may compromise the diagnostic value of the method, as focal seizure patterns may not be captured adequately [4].

Studies have shown varying results regarding the effectiveness of montages with a reduced electrode set. Some research suggests that generalized phenomena can be recognized using two to four electrodes, while other studies indicate a decrease in the detection of seizures when the electrode count is reduced. Sensitivity levels ranging from 67% to a significant decrease in seizure detection have been reported in different studies using reduced electrode montages [4].

This study aims to assess the impact of a reduced montage with only eight electrodes on seizure detection in a pediatric population using data from an epilepsy monitoring unit (EMU). The selected montage, consisting of frontal, central, temporal, and occipital electrodes on each side, is commonly used in ICU settings and aligns with German guidelines for diagnosing brain death. The advantage of using EMU-derived data is that the EMU employs maximum resources for seizure detection, including continuous EEG and video monitoring, dedicated staff, and active participation of patients and caregivers in reporting events [4].

By analyzing seizure detection rates in the EMU-derived data, this study aims to provide insights into the effectiveness of the reduced electrode montage in a pediatric population. The high level of resources and expertise available in EMUs allows for an assessment of seizure detection rates at their highest achievable level [4].

2.4 Existing methods and algorithms for feature extraction and classification

The researchers focus on automating the first step of EEG interpretation using a recently released dataset called the TUH Abnormal EEG Corpus. They employ recurrent neural network (RNN) architectures that take the raw EEG time-series signal as input. Unlike previous studies that used traditional machine learning algorithms or deep learning techniques such as convolutional neural networks (CNNs), RNNs were specifically explored for this task. The combination of raw time series and RNNs eliminates the need for hand-crafted features and allows the classifier to automatically learn relevant patterns [36].

The researchers introduce a novel deep gated RNN called ChronoNet, which incorporates concepts from 1D convolution layers, gated recurrent units, inception modules, and densely connected networks. ChronoNet achieves superior performance compared to previous studies using hand-engineered features, surpassing their results by 3.51%. By further optimizing ChronoNet with inception layers and densely connected recurrent layers, the overall improvement reaches 7.77% over previous studies. Furthermore, ChronoNet sets a new benchmark for the TUH Abnormal EEG Corpus, outperforming a recently published study. The paper's primary contributions include the proposed network architecture, its application in EEG classification, and the systematic analysis of each component's impact on performance. The utility of ChronoNet extends beyond EEG interpretation to general time series analysis, as demonstrated by applying it to the Speech Commands Dataset [36].

Deep learning has gained remarkable development in various fields, particularly in image and natural language processing. Convolutional neural networks (CNNs) are well-known deep learning models that extract features through convolutional layers, pooling layers, normalization layers, and fully connected layers, enhancing performance in diverse tasks. Nonetheless, CNNs lack the ability to retain memory of previous time series patterns, making it challenging for them to learn the most crucial features from EEG signals, presented as time series. This hinders accurate construction of the relationship between raw EEG signals and the recognition of epileptic seizures [37].

To address this limitation, recurrent neural networks (RNNs) have been utilized in natural language processing and speech recognition. Long short-term memory (LSTM), a type of RNN architecture, is widely employed for processing time series data. LSTM effectively handles the gradient vanishing problem in basic RNNs and captures long-term dependencies, enabling more effective acquisition of temporal features in sequen-

tial data. To improve the modeling capabilities of deep neural networks and account for temporal sequences, researchers have combined the strengths of CNNs and RNNs, creating convolutional LSTM neural networks. This fusion has proven effective in tasks requiring large vocabulary, providing a relative improvement over standalone LSTMs. Numerous studies have focused on extracting temporal and spatial features by combining CNN and LSTM models, demonstrating the effectiveness and superiority of this approach. A CNN-LSTM model is proposed to enhance EEG signal classification by extracting features, enabling automatic recognition and detection of epileptic seizures in real-time or recorded EEG data [37].

2.5 Gaps and limitations in current research

Deep learning has had some success deciphering EEG, but there are still several obstacles in the way of its wider deployment. In addition to the challenges associated with decoding EEG signals' high dimensionality and poor signal-to-noise ratio, complex real-world application situations and the algorithm's inherent limits have complicated research and development.

- i. Many-sample labeled EEG information sets are still required since the impact of current deep learning methods is not completely represented. High-quality labeled data is essential for deep learning to work well. EEG data with complete and correct labeling is still hard to come by in current research, particularly in clinical research, and the sample size is limited. Future studies will need to apply new machine learning methods like transfer learning to compensate for the drawbacks of small sample sizes, in addition to gathering and sorting huge samples of EEG data.
- ii. The generalizability and repeatability of the current model have not been thoroughly tested in multi-center and longitudinal data. The major test collection, as well as various laboratories/hospitals, and the EEG data display varied features since the EEG data is significantly influenced by equipment and experimental people. EEG data also exhibits significant intra- and inter-individual variability. However, the majority of currently created models are based on data that was gathered from the same center and at the same time. To check for adequate generalization and repeatability, the model has to be validated on multi-center longitudinal data.
- iii. Real-time decoding is challenging due to the complexity of deep learning models, which is still high. Deep learning allows for continual model adjustment in re-

sponse to the application. Although the depth, complexity, and activation function of the model can improve classification model performance, they also have drawbacks that make real-time execution challenging and result in longer training times. Due to these issues, EEG signal decoding will need more resources and have fewer practical applications .

- iv. There is a need to improve deep learning's interpretability in EEG research. The most crucial objective in EEG-based psychological and medical research is not categorization accuracy. To understand more about psychological or medical conditions, one must use machine learning models. An essential objective of this research is to identify predictive EEG traits that might disclose brain pathways. Deep learning models must therefore become more interpretable in order to become an effective tool for researching brain systems.
- v. The use of unlabeled EEG data is not supported by the current deep learning model. Most of the EEG data sets utilized in current research are labeled data. Deep learning models are therefore mostly supervised learning. However, a lot of EEG data—particularly in medical research—remains unlabeled or with incorrect labeling. As a result, unsupervised or semi-supervised deep learning techniques also need to be continually improved in order to be used with EEG data that has incorrect or missing labels.

In conclusion, the network architecture of CNN, Deep Belief Networks (DBN), Autoencoders (AE), and RNN are mostly used in the present application of deep learning in EEG decoding. It is based on a number of established brain-computer interface (BCI) classification paradigms, the classification and prediction of cognitive states such as emotional weariness, and the detection of clinical seizures. Although there have been numerous successful uses of sleep categorization, there are still several issues with current research, including a lack of multi-center verification and significant complexity. Data gathering and sorting are necessary in order to get beyond deep learning's restrictions and issues with EEG decoding. The collaboration between brain science processes and deep learning algorithms enables us to go beyond the restrictions faced by deep learning in EEG analysis. To satisfy the demands of real-time online applications and be suited for multi-center, huge sample sizes, robust and efficient deep learning algorithms must be continually developed in future research. Furthermore, the requirements of the application could not be satisfied by a single kind of deep learning algorithm. As a result, employing integrated learning and reinforcement learning, numerous distinct models may also be merged in addition to the model's architecture being optimized. The goal is to maximize the benefits of several models to achieve better performance.

Chapter 3

State of the art Classifiers

3.1 Machine Learning Models

3.1.1 Logistic Regression (LR):

In binary classification problems, where the objective is to forecast the probability of an input belonging to one of two classes, a statistical analysis technique logistic regression is used there. Since its data points are not grouped in line rows, this method differs significantly from linear regression. The algorithm for optimization is used by the training classifier to identify the ideal regression coefficient in the regression formula [38] [39].

The sigmoidal or S-shaped curve that translates the input characteristics to the predicted probability is an assumption made by the logistic regression model. The linear combination of the input characteristics is converted to a value between 0 and 1 using the sigmoid function, also known as the logistic function. The sigmoid function is represented by the following formula:

$$p = \frac{1}{1 + e^{-z}} \quad (3.1)$$

Where p represents the probability of the event occurring, and z is the linear combination of the input features and their respective coefficients:

$$z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad (3.2)$$

In this equation, b0 is the intercept term, b1, b2, ..., bn are the coefficients correspond-

ing to the input features x_1, x_2, \dots, x_n . The coefficients $b_0, b_1, b_2, \dots, b_n$ are estimated using a process called maximum likelihood estimation, which seeks to maximize the likelihood of observing the given set of data. It takes a dataset containing labeled instances to train a logistic regression model. The examples include class labels (0 or 1) that correlate to the input characteristics. The model is then trained by minimizing the gap between the predicted probabilities and the actual class labels in the training data by optimizing the coefficients.

An optimization approach like gradient descent is frequently employed during the training phase to iteratively change the coefficients depending on the computed error or loss. The binary cross-entropy loss, which assesses the similarity between the predicted probabilities and the actual class labels, is the most widely used loss function in logistic regression.

Once trained, the logistic regression model may be used to predict outcomes from brand-new datasets. The model calculates the linear combination of the input characteristics and their coefficients, applies the sigmoid function, and generates the projected probability given a set of input features. To categorize the input into one of the two classes, a threshold can be applied to this probability.

A few benefits of logistic regression include its efficiency, interpretability, and simplicity. In comparison to more complicated models, it is less prone to overfitting and can handle input characteristics that are both numerical and categorical. However, logistic regression makes the assumption that the input attributes and the outcome's log chances are linearly related, which may restrict its effectiveness when the connection is not linear.

3.1.2 Gaussian Naive Bayes (GNB):

One of the top 10 data mining methods one is the naive Bayes classifier, which is an effective classifier. The Bayesian rule and probability theorems are the foundation of Bayesian classifiers. Bayes relies on two suppositions. The first is that attributes are conditionally independent given the class label, and the second is that no latent attribute influences the label prediction process. A variation of the Naive Bayes method called Gaussian Naive Bayes makes the assumption that the continuous input variables have a Gaussian (normal) distribution [40] [41] [42]. It is a well-known probabilistic classification technique that is utilized in machine learning, and it performs especially well when working with continuous or numerical data.

Based on Bayes' theorem, the Naive Bayes method makes the assumption that given the

class labels, the features (input variables) are conditionally independent of one another. Naive Bayes classifiers have been shown to function effectively in several real-world contexts in spite of this "naive" assumption.

Each class in the Gaussian Naive Bayes model is represented by a Gaussian distribution with mean and variance. The training data is used by the algorithm to determine the mean and variation of each feature for each class. These learnt parameters are used to determine the likelihood that a given feature value belongs to each class during the prediction phase.

3.1.3 Decision Tree (DT):

A popular machine learning approach for classification and regression applications is the decision tree. By learning straightforward decision rules derived from the input data, it develops a model that predicts the value of a target variable. The program creates a model of decisions and their outcomes that resembles a tree.

In order to produce homogenous subsets inside each partition, the decision tree method recursively divides the data depending on the values of the input characteristics. Based on a criterion that optimizes the separation of the target variable, it chooses the optimum feature to split the data at each phase. A stopping condition, such as reaching a maximum depth or having a minimum number of instances in a leaf node, must be satisfied before the splitting process can cease. The decision tree's leaf nodes represent class labels in the classification context, and the majority class inside each leaf node is used to determine the predicted class for cases that pass through that leaf node.

Interpretability is one of the benefits of decision trees since the decision rules are simple to comprehend and depict. They are reasonably resilient to outliers and can handle both numerical and categorical information. Additionally, non-linear correlations and interactions between features can be captured via decision trees.

Nevertheless, decision trees are prone to overfitting, particularly when they are too deep and complicated. Techniques like pruning, establishing a limit tree depth, or using ensemble methods like Random Forests are frequently used to reduce overfitting. Random Forests integrate many decision trees to provide predictions, which enhances the model's capacity for generalization.

3.1.4 Random Forest Tree (RFT):

An ensemble learning technique called Random Forest uses many decision trees to provide predictions. It is a well-liked and effective machine learning technique used for both regression and classification problems. The idea of decision trees is built upon by Random Forest, which also tackles some of its drawbacks. Instead of relying on a single decision tree, Random Forest creates an ensemble of decision trees, each trained on a random subset of the training data [43]. The randomness comes from two key aspects: random sampling of the training data and random feature selection. Here's an overview of the Random Forest algorithm:

- **Random Sampling:** Random Forest randomly chooses replacement subsets from the training data. This process is known as "bagging" or bootstrap aggregating. In the forest, a single decision tree is trained using a bootstrap sample from each subgroup.
- **Random Feature Selection:** Random Forest randomly chooses a subset of features to take into account for splitting at each node of the decision tree. This choice aids in introducing variation among the forest's decision trees.
- **Voting or Averaging:** The Random Forest ensemble of decision trees makes the final class prediction for classification problems. The anticipated class designation is given to the group receiving the most votes. When performing regression tasks, the ensemble calculates the final projected value by averaging the predictions from each decision tree.
- **Voting or Averaging:** The Random Forest ensemble of decision trees makes the final class prediction for classification problems. The anticipated class designation is given to the group receiving the most votes. When performing regression tasks, the ensemble calculates the final projected value by averaging the predictions from each decision tree.

Random Forest offers several advantages over individual decision trees:

- **Improved Accuracy:** In many cases, Random Forest outperforms a single decision tree in terms of accuracy, especially when working with complicated datasets or high-dimensional feature spaces. The group of trees makes it easier to identify a wider variety of links and patterns in the data.
- **Reduction of Overfitting:** Random Forest lessens the possibility of overfitting by building an ensemble of decision trees and only taking into account portions of

the data and characteristics. Combining several trees makes it easier to identify stronger and more universal patterns in the data.

- Feature Importance: By examining the performance drop when a certain feature is randomly permuted, Random Forest may calculate the relevance of a feature. Understanding the importance of characteristics in the prediction process is made easier with the use of this information.

But when dealing with a lot of trees and data, Random Forest might be computationally costly. Additionally, because Random Forest contains numerous trees and their aggregated predictions, it could be harder to comprehend than a single decision tree.

3.1.5 K-Nearest Neighbour (KNN):

A straightforward yet frequently successful non-parametric classification technique is the k-Nearest-Neighbor (kNN) algorithm. In order to classify a data record t , its k nearest neighbors are obtained, creating a neighborhood for t . The categorization for t is often determined by the neighborhood's data records voting in the majority, with or without taking into account distance-based weighting [44] [45].

Here's an overview of the KNN algorithm:

- Data Preparation: A set of labeled examples with input characteristics and accompanying class labels or target values serve as the training data.
- Choosing the Value of K : The nearest neighbors to take into account are represented by the value of K . Usually, cross-validation methods or experiments are used to find this value.
- Distance calculation: The method determines the distance between each test instance and every other instance in the training data for a given test instance. Manhattan distance, Euclidean distance, and cosine similarity are three frequently used distance measures.
- Finding Nearest Neighbors: Based on the measured distances, the KNN algorithm chooses the K training examples that are closest to the test instance.
- Classifying or Predicting: The approach selects the class label for classification tasks that is most common among the K closest neighbors. In order to forecast the target value of the test instance in regression tasks, the method computes the average or weighted average of the target values of the K nearest neighbors.

- Performance Evaluation: Depending on the nature of the issue, the performance of the KNN algorithm is assessed using a variety of evaluation measures, such as accuracy, precision, recall, or mean squared error.

Key factors and KNN features include the following:

- The selection of the distance measure and the value of K has an impact on KNN. Different distance measures and K values could provide various outcomes.
- KNN is a lazy learning algorithm, which means that it does not formally construct a model during training. To make predictions, it instead keeps the training data in memory.
- As the algorithm must calculate distances between every training instance during prediction, its computational cost increases with the amount of the training data.
- The "curse of dimensionality", where performance suffers as the number of input characteristics rises, can be a problem for KNN. Techniques for feature selection or dimensionality reduction can help to alleviate this.

3.1.6 Support Vector Machine (SVM):

SVM is a mathematical entity, an algorithm for maximizing a particular mathematical function with respect to a given collection of data [46]. It is particularly effective in handling complex datasets with a clear margin of separation between classes or with non-linear relationships. The goal of SVM is to identify the optimum hyperplane for separating data points from various classes. The margin, which is the separation between the hyperplane and the closest data points for each class, is maximized by the hyperplane, a linear decision boundary. The objective is to minimize classification mistakes while achieving the greatest possible separation between classes.

The capability of SVM to handle both linear and non-linear decision boundaries is one of its essential characteristics. SVM uses the kernel technique if the data cannot be linearly separated in the original feature space. The data may be transformed into a higher-dimensional space where it is linearly separable using SVM using the kernel technique. SVM can efficiently learn non-linear decision boundaries by mapping the data into this higher-dimensional feature space. In SVM, choosing the right kernel function is essential. The linear kernel, polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel are examples of common kernel functions. Each kernel function has unique characteristics and is appropriate for various data kinds and decision limits.

Finding the hyperplane or support vectors that characterize the decision boundary is required for SVM training. The data points nearest to the decision boundary, known as support vectors, are very important in defining the hyperplane. SVM seeks to optimize the margin between the support vectors of various classes, strengthening its resistance to outliers and enhancing generalization to untried data.

SVM provides a number of benefits, but there are a few things to take in mind as well. SVM may require a significant computing investment, especially when working with big datasets or intricate kernel functions. Cross-validation techniques are frequently used to find the optimum values for parameters like the regularization parameter and the kernel-specific parameters since proper parameter tuning is essential for achieving the greatest performance.

3.2 Deep Learning Models

3.2.1 Vanilla Neural Network (VNN):

A vanilla neural network is a simple deep learning architecture often referred to as a standard or basic neural network. An input layer, one or more hidden layers, and an output layer are the network's three primary building blocks. Artificial neurons, also known as nodes or units, are the building blocks of each layer and are connected by weighted connections. The strength of the connections between neurons is represented by the weights.

Each neuron in a layer of a standard neural network is linked to every neuron in the layer above it. Due to this connectedness, intricate patterns and linkages in the data may be discovered as well as information can spread. Each neuron applies an activation function to the weighted sum of its inputs in order to analyze the data. The sigmoid function, hyperbolic tangent (tanh) function, and rectified linear unit (ReLU) function are examples of common activation functions. It may identify non-linear patterns in the data because of these activation functions.

Using an optimization approach like gradient descent, the network modifies the weights of the connections during training. The objective is to minimize a loss function that measures the discrepancy between the outputs that are expected and those that are actually produced. The weights are updated as part of this iterative learning procedure, also known as backpropagation, to enhance the performance of the network. For a variety of applications, including as classification, regression, and pattern recognition,

vanilla neural networks can be employed. They could encounter issues like overfitting or trouble identifying complicated correlations in the data, though.

Figure 3.1 represents the VNN used in this study. The network consists of 3 densely connected layers; each containing 16 neurons and rectified linear unit (ReLU) as activation function. S-shaped rectified linear activation function (Sigmoid) function has been used for output layer. Learning rate = 0.001 and batch size = 512 have been used. The model has been fit for 100 epochs.

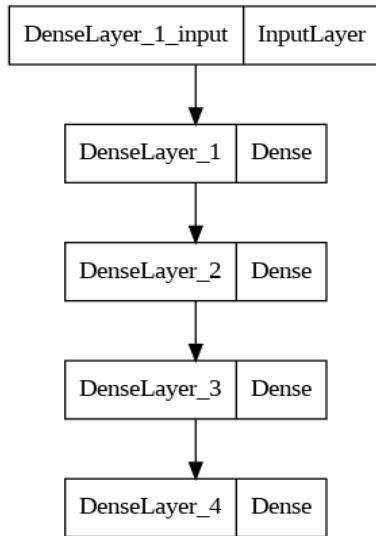


Figure 3.1: Architecture of VNN

3.2.2 Convolutional Neural Network (CNN):

Convolutional Neural Networks (CNNs) employ a parameter sharing scheme, which is used in convolutional layers to control and reduce the number of parameters. For controlling overfit or reducing the spatial size of data, number of trainable parameters and computation, a pooling layer is designed [47].

Here's an overview of the key components and operations in a CNN:

- **Convolutional Layers:** The basic elements of CNNs are these layers. They process the input data using a set of teachable filters, also referred to as kernels. Each filter creates a feature map that captures regional patterns or features via convolution over the input using element-wise multiplication. Different sorts of characteristics are extracted using a variety of filters.
- **Pooling Layers:** Pooling layers are frequently used to down sample the feature maps after convolution. The feature maps' spatial dimensions are reduced but

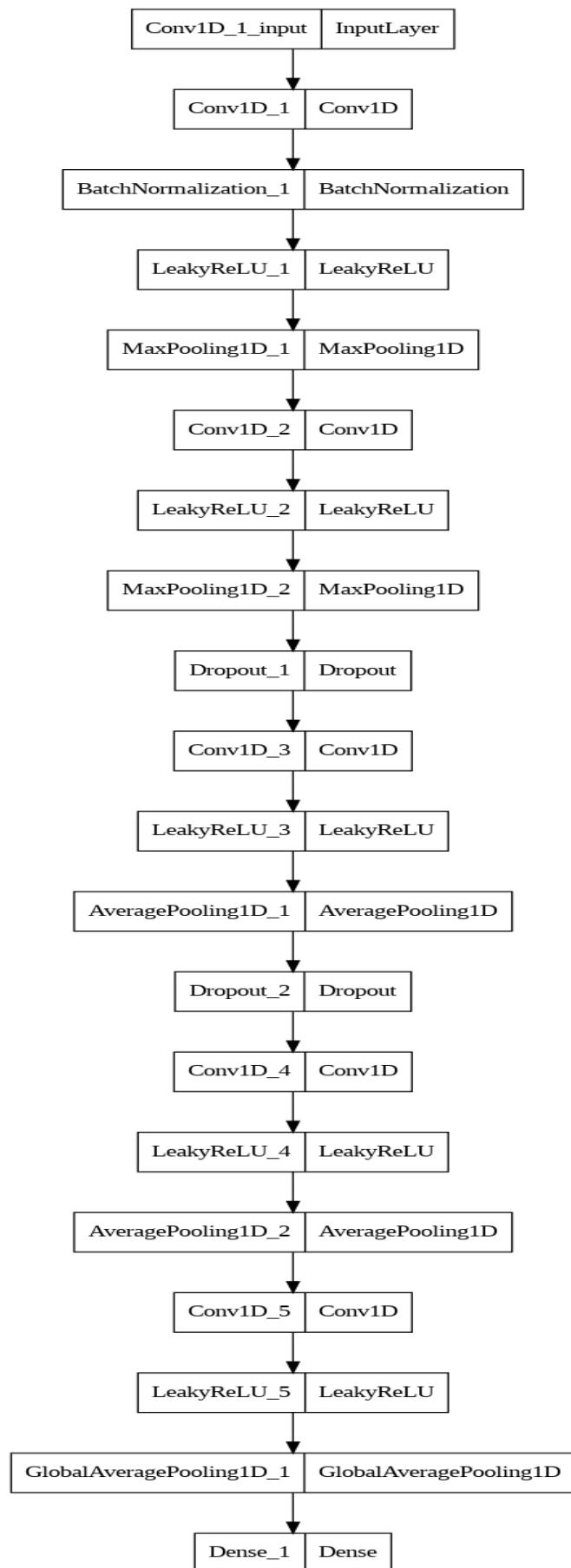


Figure 3.2: Architecture of CNN

vital data is preserved by pooling. Maximum or average values within a certain frame are chosen in pooling techniques known as max pooling and average pooling, respectively.

- Activation Functions: After each convolutional or pooling layer, non-linear activation functions like ReLU (Rectified Linear Unit) are applied element-by-element to the feature maps. By introducing non-linearities, activation functions allow the CNN to understand intricate correlations in the data.
- Fully Connected Layers: Predictions based on the learnt characteristics are made toward the network’s end using one or more fully linked layers. Similar to a conventional neural network, these layers link every neuron in the previous layer to every neuron in the layer after that.

Because of its shared weights and shared parameter scheme, CNNs are excellent at capturing spatial and local patterns. They are therefore very good at tasks like segmentation, object identification, and picture classification. Additionally, CNNs are more computationally efficient than fully connected networks because of the inclusion of convolutional and pooling layers, which minimize the number of parameters.

Figure 3.2 shows the the CNN model used in this research. Each one dimensional convolutional layer contains 5 filters, kernel size = 3, and strides = 1. For maxpooling and averagepooling layers, pool size = 2 and strides = 2. Dropout rate = 0.2 has been used. Sigmoid activation function has been used for output layer. Learning rate = 0.0006 and batch size = 32 have been used. The model has been fit for 100 epochs.

3.2.3 Gated Recurrent Neural Network (RNN):

A class of neural networks called recurrent neural networks (RNN) is used to analyze variable-length sequential input. The activation of an RNN’s recurrent hidden state changes with time and is influenced by the preceding time step. Using a set of weights and activation functions, an RNN’s hidden state is updated depending on the current input and the prior hidden state. The network can describe complicated interactions in the sequential data due to the activation function, which incorporates non-linearity. Both the rectified linear unit (ReLU) function and the hyperbolic tangent (tanh) function are frequently utilized activation functions in RNNs.

RNNs are susceptible to the vanishing and exploding gradient problem, which occurs when the gradients either become too small (vanishing) or too large (exploding) during the backpropagation process. Two popular models in use are the long short-term

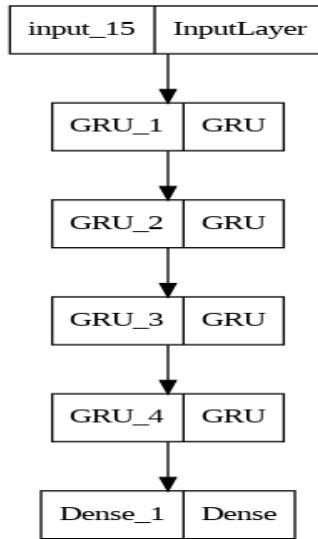


Figure 3.3: Architecture of RNN

memory (LSTM) [48] [49] [50] and the gated recurrent unit (GRU) [51].

Figure 3.3 shows the deep gated RNN model used in this study. Each of the four GRU layers contains 32 units with hyperbolic-tangent (tanh) activation. The first 3 layers return sequences as well. Sigmoid activation function has been used for output layer. Learning rate = 0.001 and batch size = 128 have been used. The model has been fit for 50 epochs.

3.2.4 Convolutional Gated Recurrent Neural Network(C-RNN):

C-RNN is nothing but concatenation of CNN and RNN layers. Instead of simple RNN, deep gated RNN layers have been used to get the advantages of GRUs.

Figure 3.4 shows the the C-RNN model used in this research. Same GRU layers of the RNN model has been used here. Each convolutional layer contains 32 filters, kernel size = 4 and strides = 2 with ReLU activation function. Learning rate = 0.0006 and batch size = 512 have been used. The model has been fit for 50 epochs.

3.2.5 Inception Convolutional Gated Recurrent Neural Network (IC-RNN):

Convolutional neural network (CNN) researchers at Google first announced the Inception architecture in 2014. Utilizing many concurrent convolutional layers with various filter sizes within the same network layer is the primary principle underlying the In-

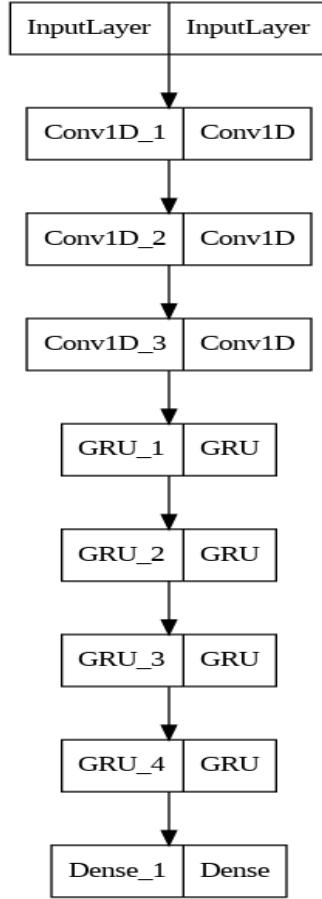


Figure 3.4: Architecture of C-RNN

ception design. The network can learn both fine-grained details and high-level global information concurrently since it can record characteristics at different geographical scales and resolutions.

The "Inception Module" is the "Inception" of the Inception architecture. This module consists of a set of parallel convolutional layers with 1×1 , 3×3 , and 5×5 filters, among other filter sizes. The 1×1 convolutional layers used for dimensionality reduction to lower the computing cost are added after these parallel convolutional layers.

In IC-RNN model, inception architecture has been used in CNN layers and concatenated with gated RNN layers. Figure 3.5 shows the the IC-RNN model. Same GRU layers of the RNN model has been used here. Before GRU layers, three similar blocks have been repeated. Each block contain three parallel convolutional layers followed by a concatenation layer. All three convolutional layers have 32 filters and strides = 2. The kernel sizes for the three layers are 2,4, and 8 respectively. Learning rate = 0.0006 and batch size = 512 have been used. The model has been fit for 40 epochs.

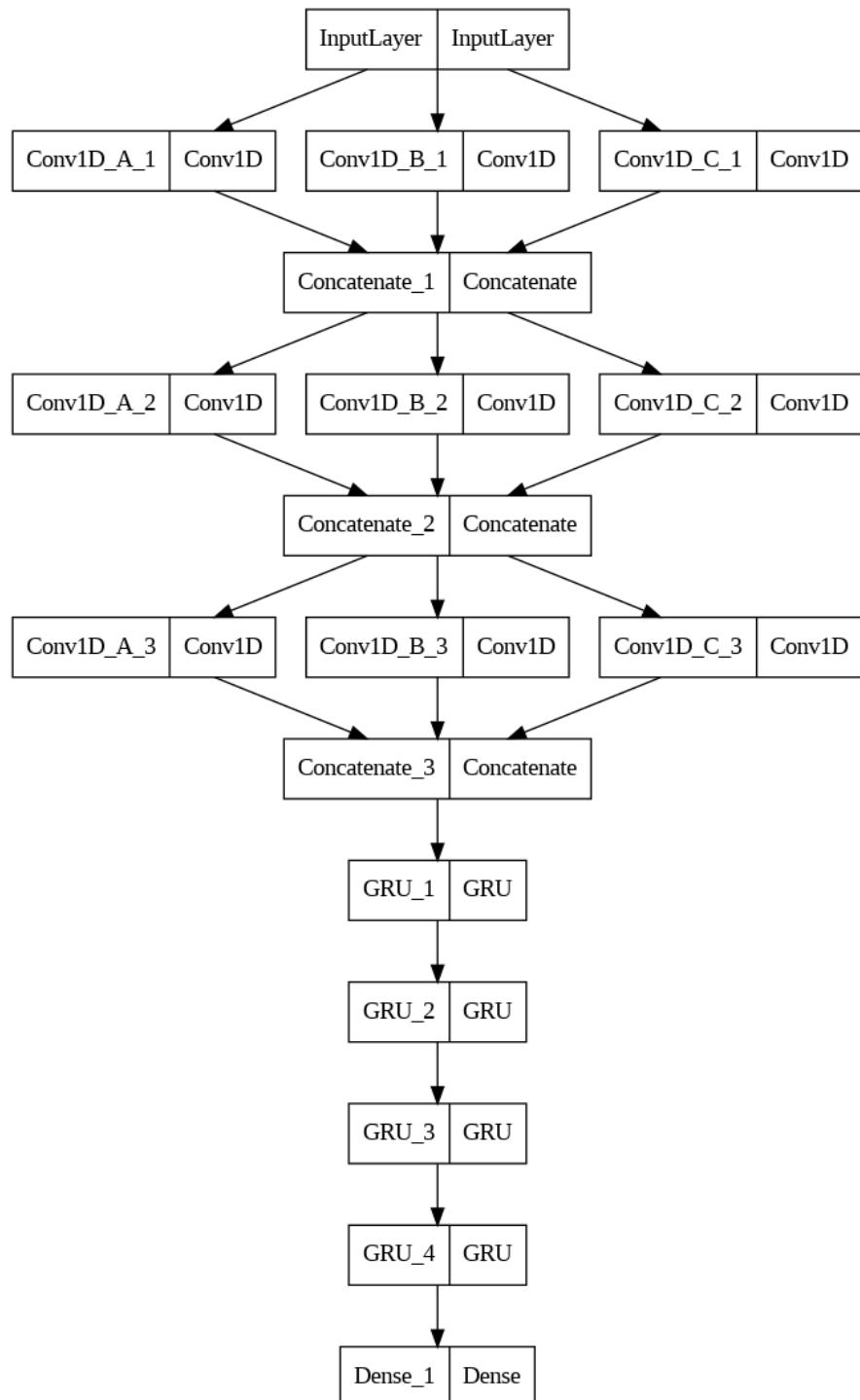


Figure 3.5: Architecture of IC-RNN

3.2.6 Convolutional Densely Connected Gated Recurrent Neural Network (C-DRNN):

C-DRNN is a modified version of C-RNN. 2-3 GRU units are concatenated multiple times which can be called densely connected gated RNN. As C-RNN, this network also have multiple convolutional blocks. Figure 3.6 shows the the C-DRNN model. Same as C-RNN. Only 2 concatenation layers have been introduced in between the GRU layers. Learning rate = 0.0006 and batch size = 512 have been used. The model has been fit for 50 epochs.

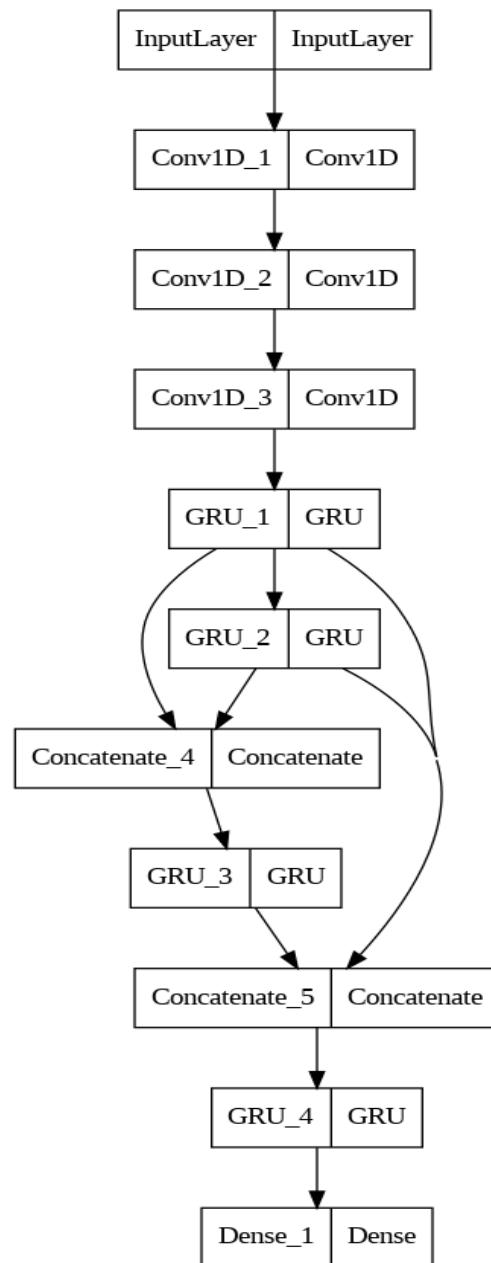


Figure 3.6: Architecture of C-DRNN

3.2.7 ChronoNet:

In ChronoNet, inception architecture in CNN layers and densely connected gated RNNs are combined. Figure 3.7 shows the ChronoNet model which is nothing but the combination of IC-RNN and C-DRNN. Learning rate = 0.0006 and batch size = 32 have been used. The model has been fit for 100 epochs.

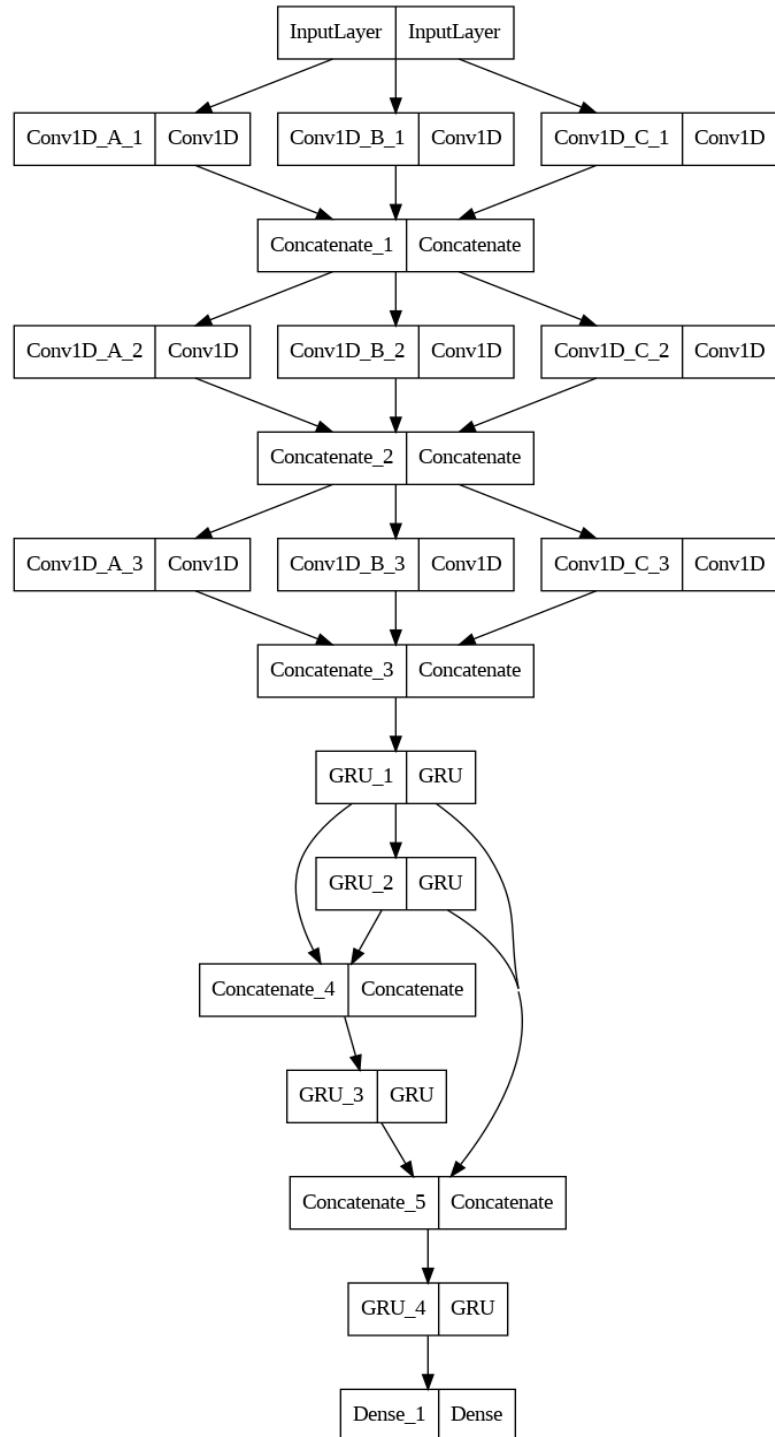


Figure 3.7: Architecture of ChronoNet

3.2.8 Convolutional Neural Network - Long Short Term Memory (CNN-LSTM):

Recurrent neural network (RNN) architectures such as Long Short-Term Memory (LSTM) are created with the goal of solving the vanishing and exploding gradient problem and accurately modeling long-term relationships in sequential data. The vanishing or exploding gradient problem, which affects traditional RNNs, prevents the network from capturing and learning long-term relationships because the gradients shrink or expand as they are backpropagated over time. By including a memory cell, LSTM gets over this restriction and enables the network to store and transmit data over lengthy periods of time.

The input gate, forget gate, and output gate are the three essential parts of the LSTM architecture. The network is able to manage what information to remember, forget, and output thanks to these gates, which govern the flow of information into, out of, and inside the memory cell.

Here's a high-level description of how an LSTM unit operates:

- Forget Gate: The forget gate chooses which data from the memory cell should be forgotten based on the current input and the prior concealed state. With values near to 0 suggesting information to be forgotten and values close to 1 indicating information to be maintained, it uses a sigmoid activation function.
- Input Gate: The input gate selects the fresh data that should be kept in the memory cell. It employs a tanh activation function to create fresh candidate values to be added to the memory cell and a sigmoid activation function to generate a candidate update from the current input and the prior hidden state.
- Update Memory: The memory cell is updated by the interaction of the input gate and forget gate. While the input gate updates the memory cell with fresh candidate values, the forget gate chooses which data should be removed from the previous memory cell.
- Output Gate: Which memory cell information should be output as the current concealed state is decided by the output gate. It creates the hidden state by applying a sigmoid activation function to the current input and the previous hidden state as input, multiplying that result by the tanh of the updated memory cell.

In this network CNN layers are concatenated with LSTM layers. Figure 3.8 shows CNN-LSTM model. All the convolutional layers have 1 stride and kernel of size 3.

The convolutional layers have 128, 512, 1024, and 256 filters respectively. The dropout rate is 0.5. The dense layers have ReLU activation function and 256, 256, 128, and 64 neurons respectively. Both LSTM layers contain 64 units. The first one returns sequences as well as state additionally. Learning rate = 0.0006 and batch size = 128 have been used. The model has been fit for 60 epochs.

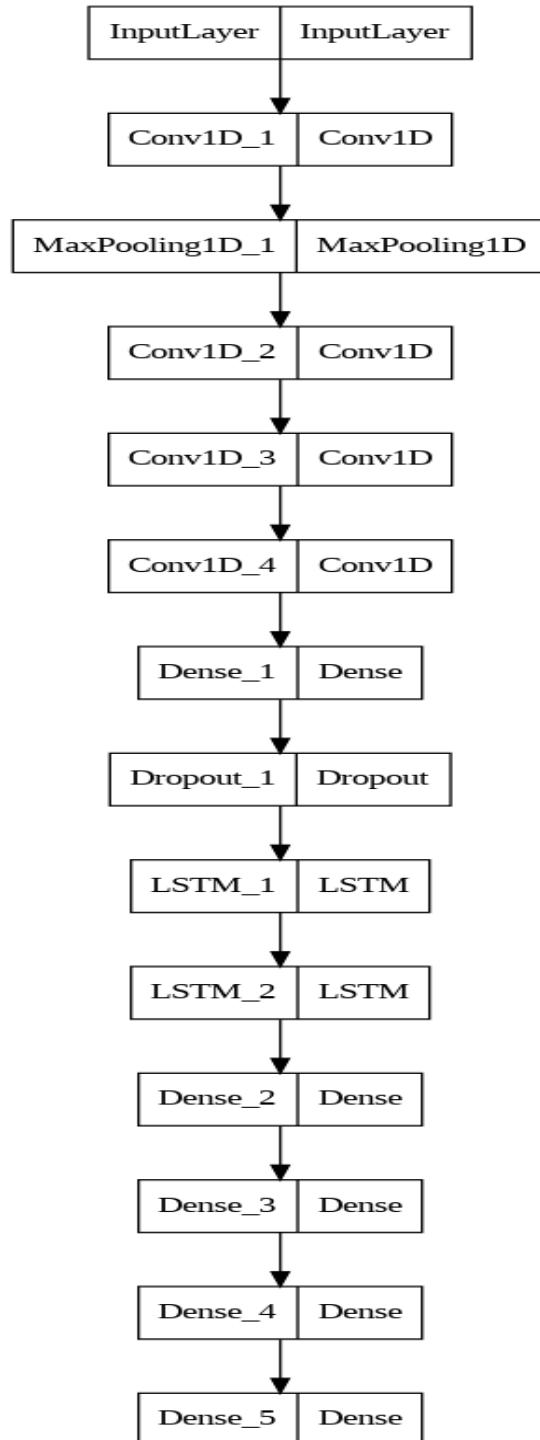


Figure 3.8: Architecture of CNN-LSTM

Chapter 4

Methodology

4.1 Overview

The outline of our proposed methodology and experimentation can be summarized in the following steps:

- i. First selected state-of-the-art classifiers have been reproduced and tested for publicly available standard datasets.
- ii. The effect of channel reduction have been investigated.
 - Standard publicly available databases used in recent literature have been set as reference databases.
 - Most informative and sensitive channels have been identified using the knowledge available in recent literature along with the trial and error method.
- iii. Finally, the architectures of selected classifiers have been modified for complexity reduction through
 - Analysis of existing selected state-of-the-art classifier and reduce the model complexity based on the knowledge available in recent literature.
 - Evaluate the performance of the modified classifier.
 - Validate and test the performance characteristics of the modified classifier.

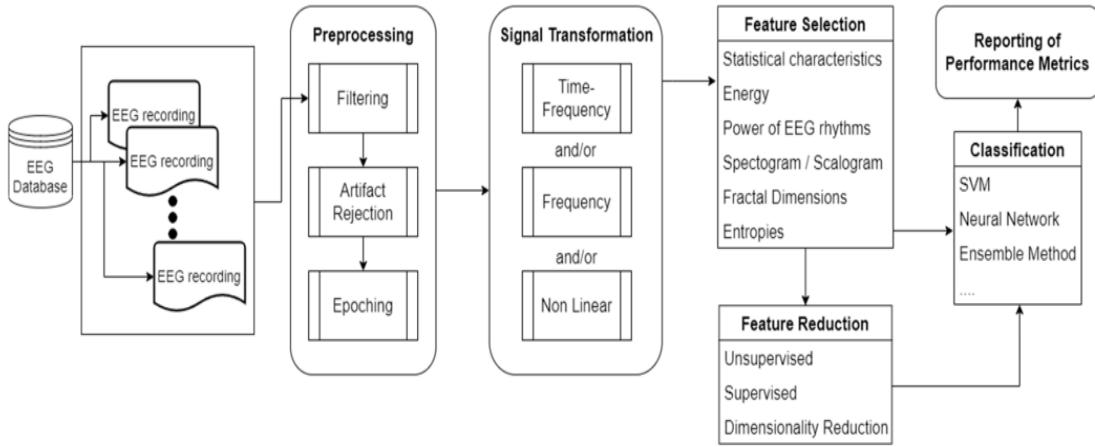


Figure 4.1: Block diagram of EEG seizure detection process

4.2 Dataset description and selection criteria

In this study, a total of 97 samples have been collected from both Nigeria and Guinea Bissau, consisting of 51 samples from individuals with epilepsy and 46 samples from individuals without epilepsy. The data have been recorded at a sampling rate of 128 Hz. To extract relevant features, a bandpass filter has been applied to focus on the frequency range of 1 Hz to 30 Hz, which includes delta (0.5-1 Hz), delta (1-2 Hz), delta (2-4 Hz), theta (4-8 Hz), alpha (8-16 Hz), and beta (16-32 Hz) bands, known to contain important information related to epilepsy [52].

To process the continuous data, these have been divided into fixed-length epochs using the *mne* module. Initially, all 14 EMOTIV EPOC+ channels corresponding to the 10-20 system have been extracted. Then, to reduce the complexity of the data, one channel at a time has been removed, resulting in a feature set of 13 channels. Additionally, a further reduction has been made, and a total of 8 channels have been thoughtfully selected from the available frontal, parietal, occipital, and temporal zones.

To accommodate machine learning models and vanilla neural networks that require 2-dimensional data, features have been further extracted by calculating the mean power spectral density within each frequency band. For other neural networks, the fixed-length epochs of the channels have been used directly as features. In total, 15 accuracy results have been obtained from each model, considering 14 & 13 channel configurations and feature extraction techniques. Then for the best models 8 channel configurations have been used. For the best performing model, 2, 4, 6, 10, and 12 channel configurations

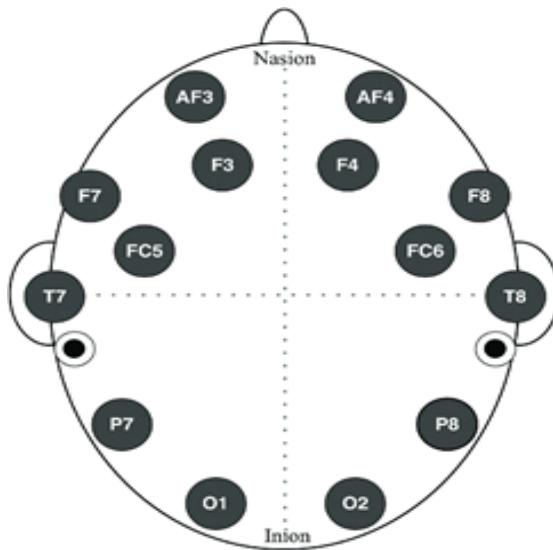


Figure 4.2: Electrode placement of the EMOTIV Epoch+ EEG headset (AF: Anterior-Frontal, F: Frontal, FC: Fronto-Central, T: Temporal, P: Parietal, O: Occipital), Black dots indicate the location of reference electrodes.

have also been employed.

In summary, the study involved careful selection and combination of 8 channels from the available 14 EMOTIV EPOC+ channels, covering frontal, parietal, occipital, and temporal zones. Different feature extraction methods have been applied based on the model requirements, including mean power spectral density calculation and direct use of fixed-length epochs.

4.3 Proposed Classifiers

4.3.1 Proposed Model 1 (ChronoNet-M):

Same GRU layers of the RNN model (Figure 3.3) have been used in this model. Before GRU layers, three similar blocks have been repeated. Each block contain two parallel convolutional layers followed by a concatenation layer. All two convolutional layers have 32 filters and strides = 2. The kernel sizes for the two layers are 2 and 4 respectively. Learning rate = 0.0006 and batch size = 512 have been used. The model has been fit for 100 epochs. Figure 4.3 shows the ChronoNet-M model.

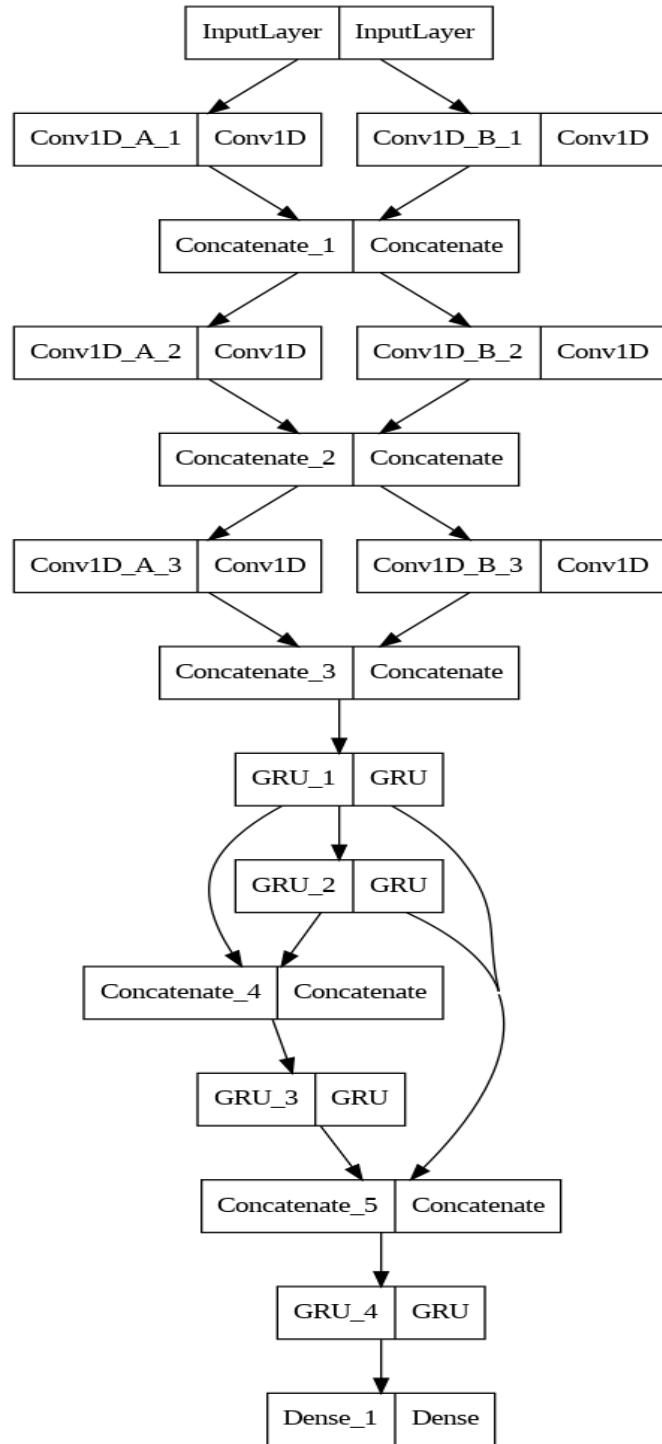


Figure 4.3: Architecture of ChronoNet-M

4.3.2 Proposed Model 2 (CNN-LSTM-M):

In the model, all the convolutional layers have 1 stride and a kernel of size 3. The convolutional layers have 128, 512, and 256 filters respectively. The dropout rate is

0.5. The dense layers have ReLU activation function and 256, 128, and 64 neurons respectively. Both LSTM layers contain 64 units. The first one returns sequences as well as states additionally. Learning rate = 0.0006 have been used. 512 and 128 batch sizes have been used for Guinea-Bissau and Nigeria dataset respectively. The model has been fit for 100 epochs. Figure 4.4 shows CNN-LSTM-M model.

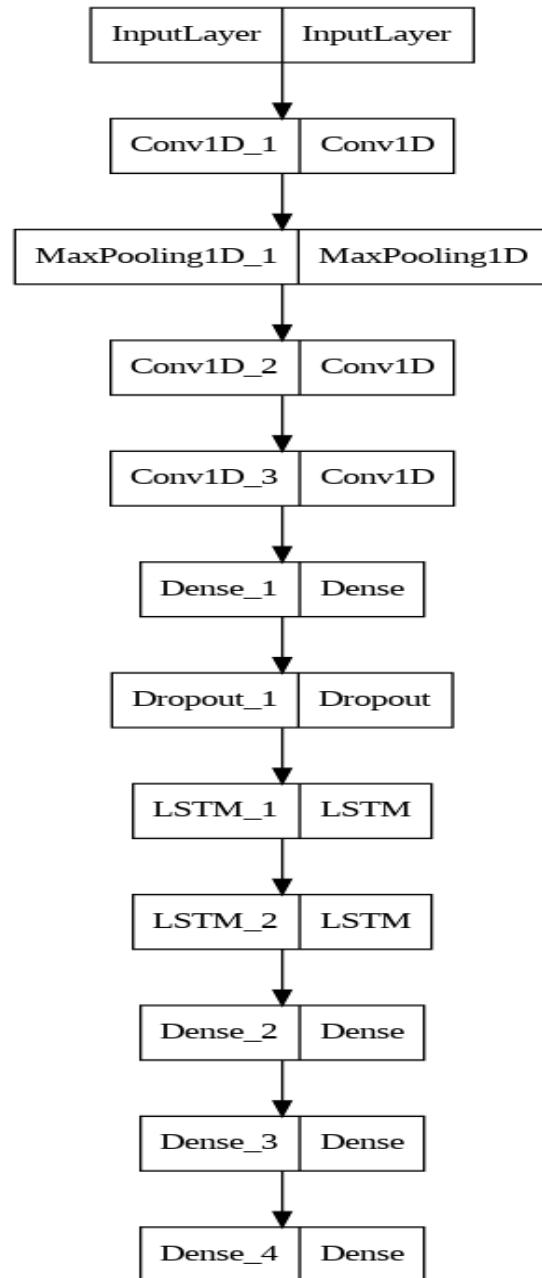


Figure 4.4: Architecture of CNN-LSTM-M

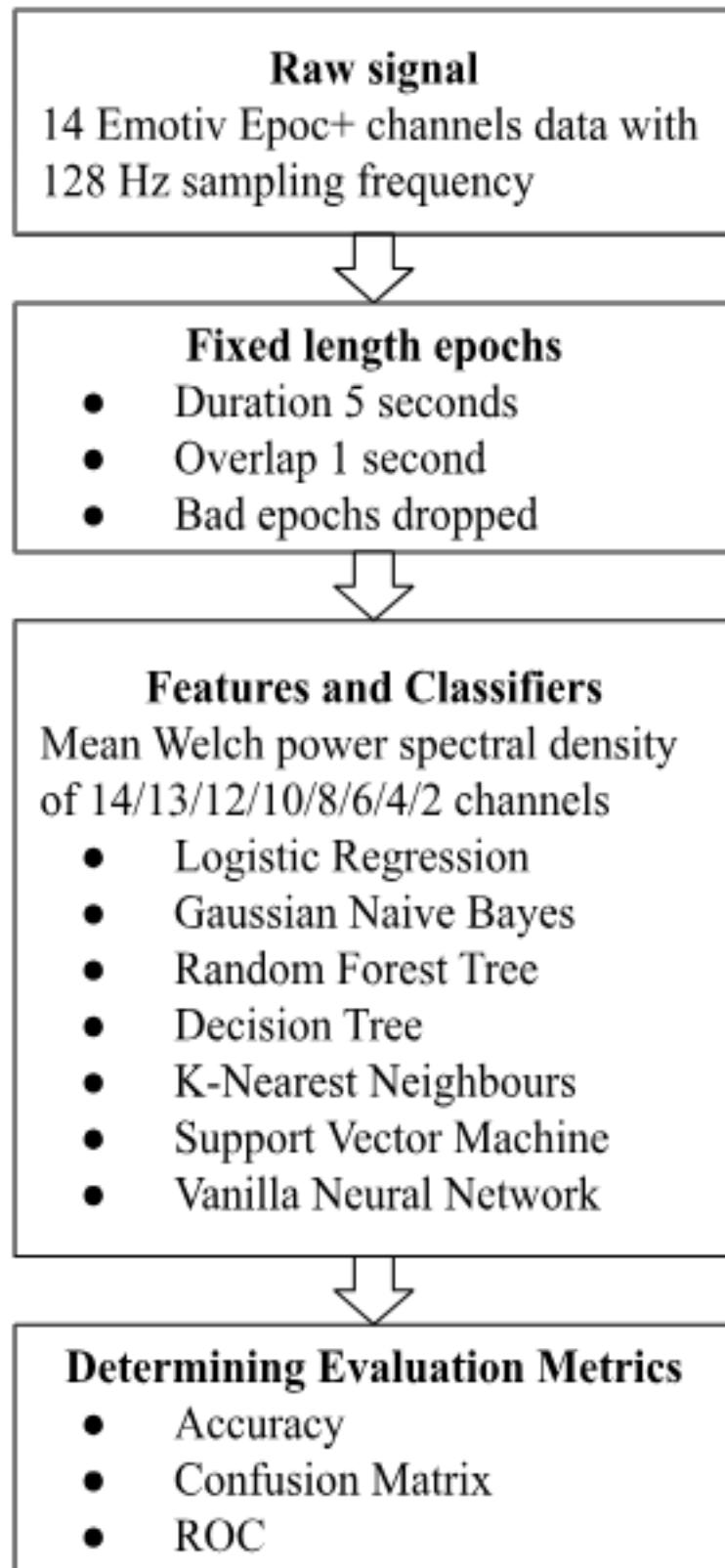


Figure 4.5: Flow chart of the machine learning models and VNN

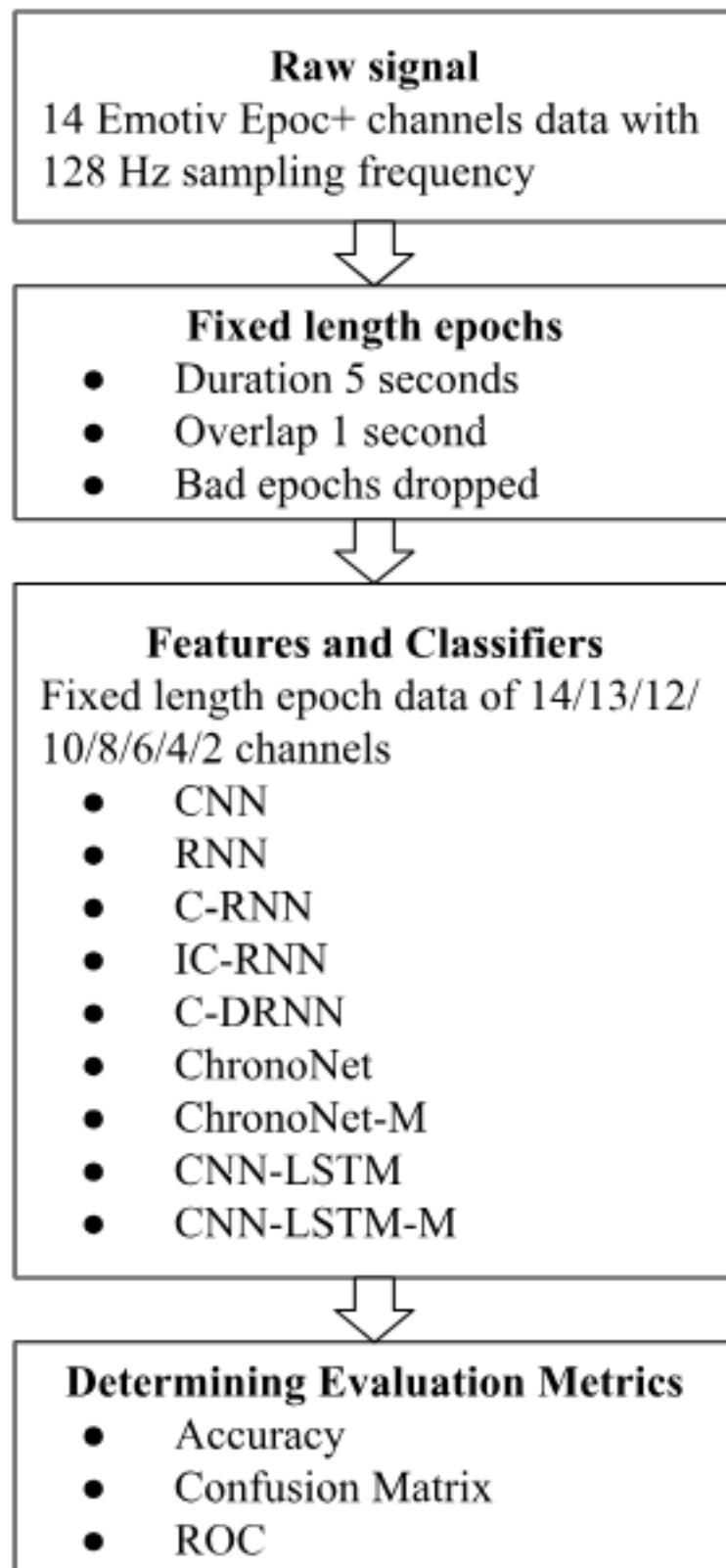


Figure 4.6: Flow chart of the deep learning models

4.4 Experimental setup and evaluation metrics

Experimental Setup for EEG Seizure Detection with Channel Reduction:

Data Collection: Two publicly available datasets of EEG recordings containing both seizure and non-seizure (normal) segments have been chosen which have a sufficient number of seizure events to draw statistically significant conclusions. The dataset include EEG recordings from multiple patients to account for individual variability.

Preprocessing: The EEG data have been preprocessed to remove artifacts, baseline drift, and noise using standard techniques like filtering, resampling, and referencing. Then the data have been segmented into fixed-length epochs, considering a suitable duration for seizure and non-seizure segments.

Channel Reduction Techniques: The following channel reduction techniques have been evaluated.

- i. **Single Channel Exclusion:** The most significant channels have been identified in a given EEG montage.
- ii. **Random Channel Selection:** A certain number of EEG channels have been discarded randomly.
- iii. **Common Spatial Pattern (CSP):** Only the most discriminative channels have been kept for seizure detection found in the literature.

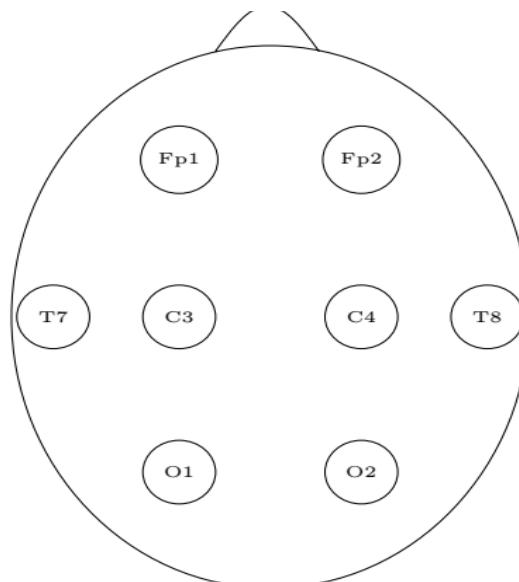


Figure 4.7: The electrodes used in the reduced montage [4].

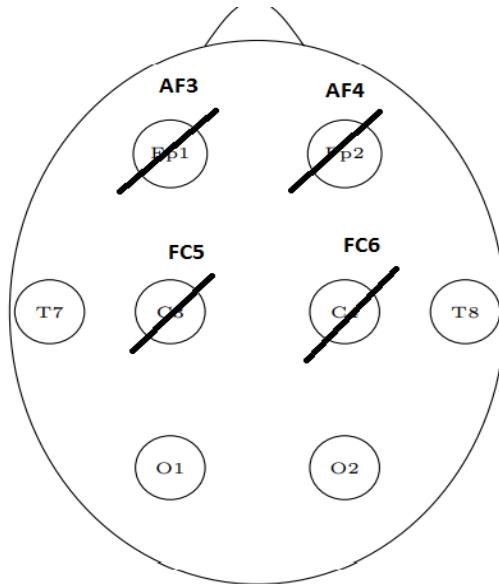


Figure 4.8: The electrodes used in the reduced montage based on CSP.

Feature Extraction: Relevant features have been extracted from the preprocessed EEG data. Commonly used features include time-domain, frequency-domain, and time-frequency domain features.

Classification Model: Suitable classification models for seizure detection have been chosen.

Experimental Design: The dataset have been randomly split into training and testing sets. The same data split is used for all channel reduction ratios and techniques to make the evaluation fair. The chosen channel reduction technique has been applied to the training set and extract features. The classification models have been trained using the reduced feature set. The performance of the testing set has been evaluated for each model and the results have been recorded.

Evaluation Metrics:

Accuracy: The overall proportion of correctly classified segments (both seizure and non-seizure).

Confusion Matrix: It provides insights into true positive, true negative, false positive, and false negative outcomes, enabling evaluation of model accuracy and error rates.

Area Under the Receiver Operating Characteristic (ROC) Curve (AUC): AUC provides an aggregate measure of the classifier's performance across various discrimination thresholds.

Chapter 5

Results and Analysis

This chapter presents the results obtained in the testing phases. Section 5.1 presents a performance comparison of all the classifiers when no channel is excluded out of 14 channels. So the best-performing classifiers can be identified easily. Then the effect of channel reduction has been observed in subsequent sections. The performance of two novel models have been presented for comparative analysis.

5.1 Performance Comparison with All Channels

A comparison of all the classifiers when all 14 channels are used, is presented here. So the best-performing classifiers can be identified easily. The results obtained here are also used as references for the comparison of the channel reduction effect.

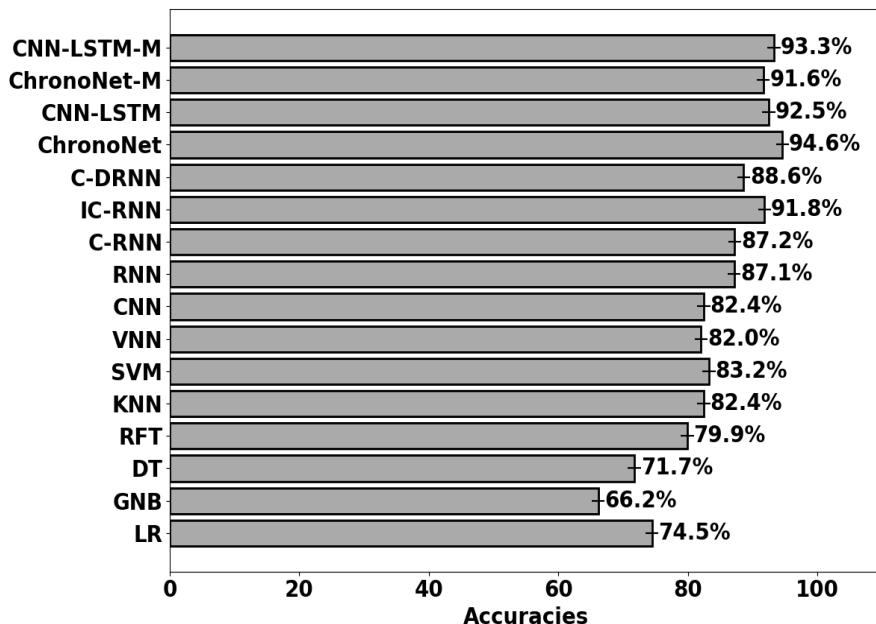


Figure 5.1: Accuracies for Guinea-Bissau dataset with no channel exclusion

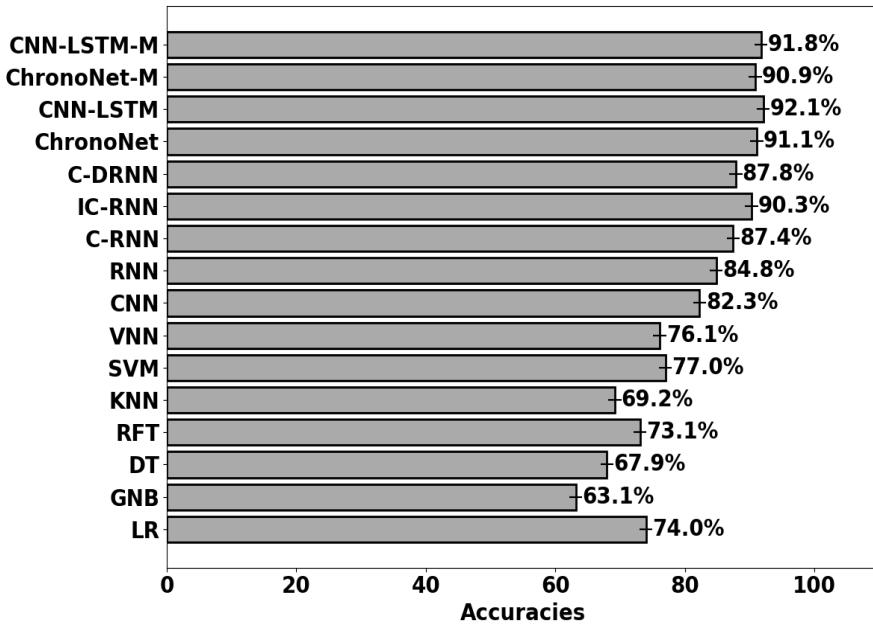


Figure 5.2: Accuracies for Nigeria dataset with no channel exclusion

5.2 Effect of Single Channel Exclusion

In this section, the performances of all the classifiers have been observed when a single channel is discarded, one at a time. It is important to find out the most significant channel that affects performance in terms of accuracy. If the accuracy falls drastically when a certain channel has been dropped, then the dropped channel certainly has significant importance. Accuracies can vary for different combinations of channels due to interference. So, the accuracies with more channels may have less values or vice versa.

5.2.1 Logistic Regression (LR):

Figure 5.3 and Figure 5.4 show the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Logistic Regression.

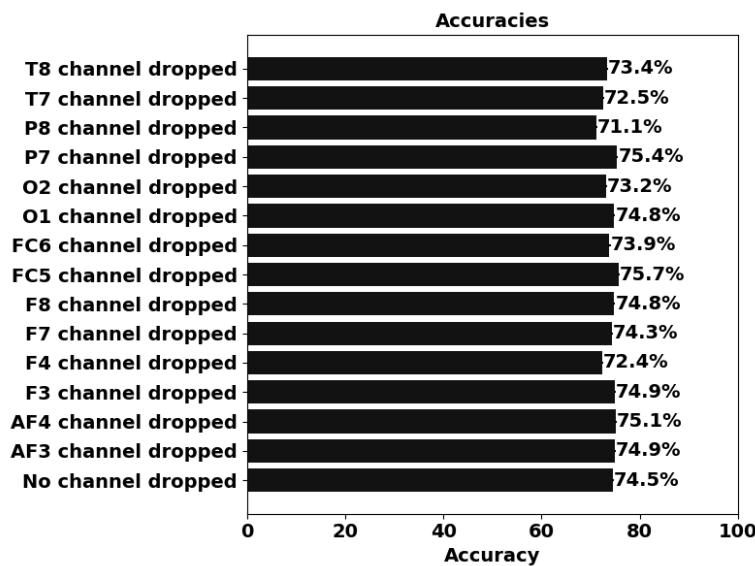


Figure 5.3: Accuracies for Guinea-Bissau dataset using LR

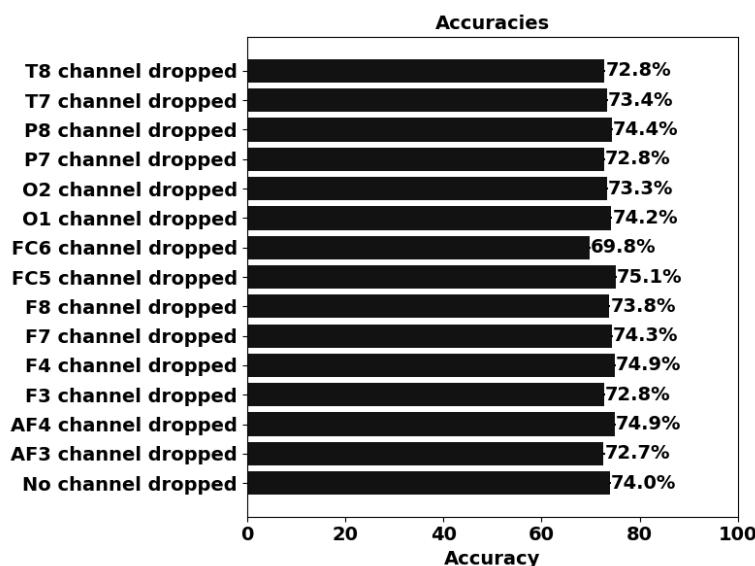


Figure 5.4: Accuracies for Nigeria dataset using LR

Figure 5.5 and Figure 5.6 represent the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Logistic Regression.

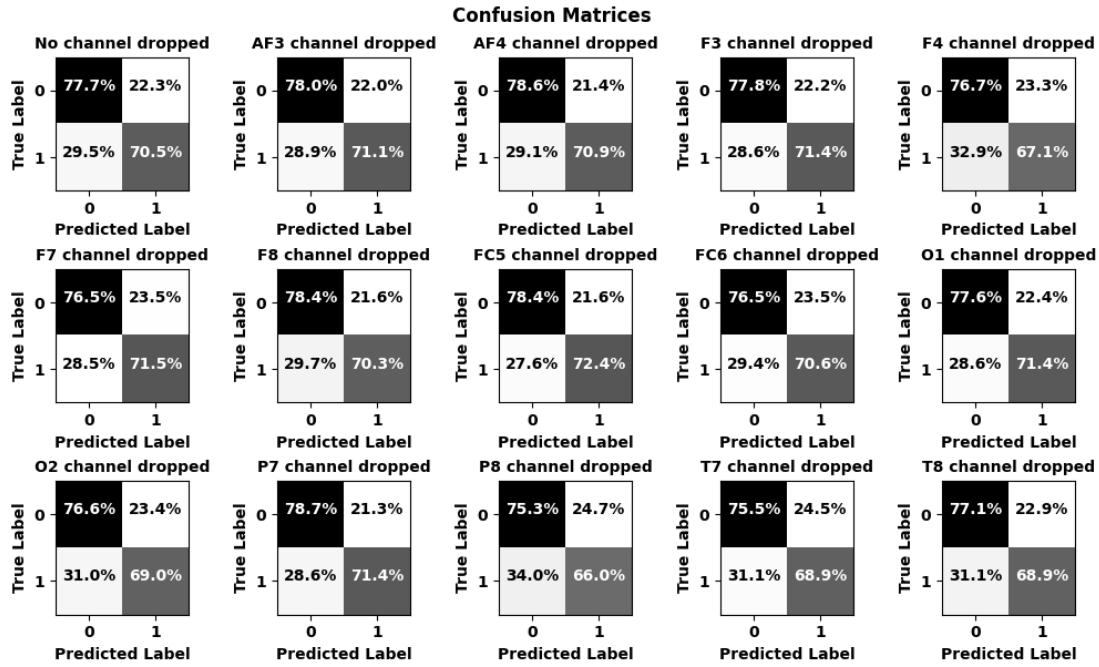


Figure 5.5: Confusion Matrices for Guinea-Bissau dataset using LR

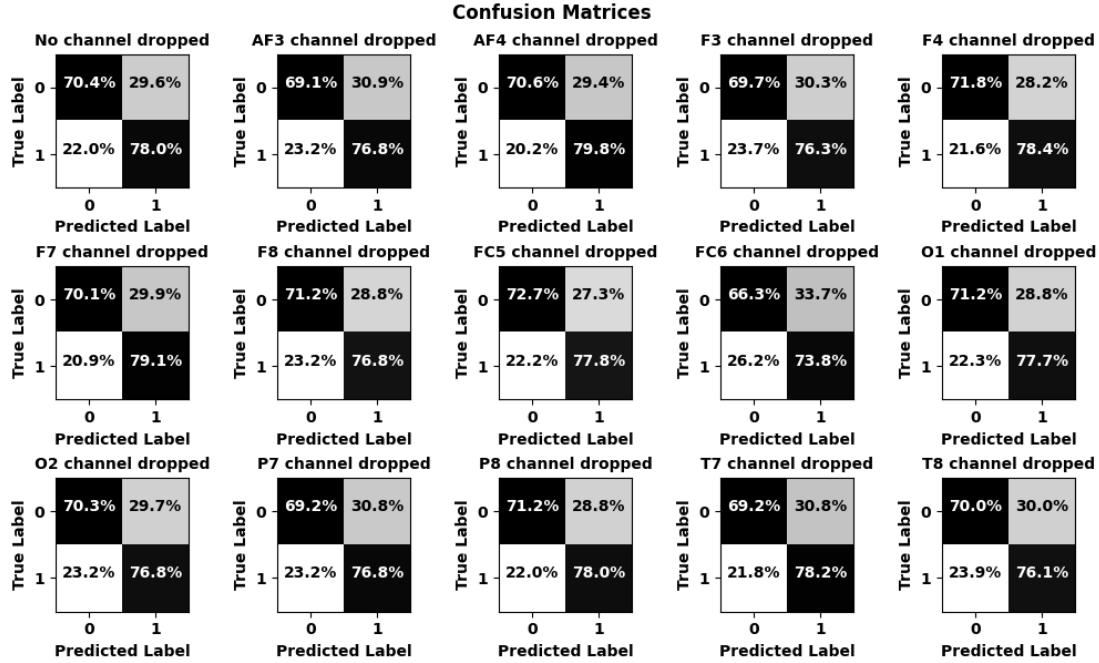


Figure 5.6: Confusion Matrices for Nigeria dataset using LR

Figure 5.7 and Figure 5.8 demonstrate the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Logistic Regression.

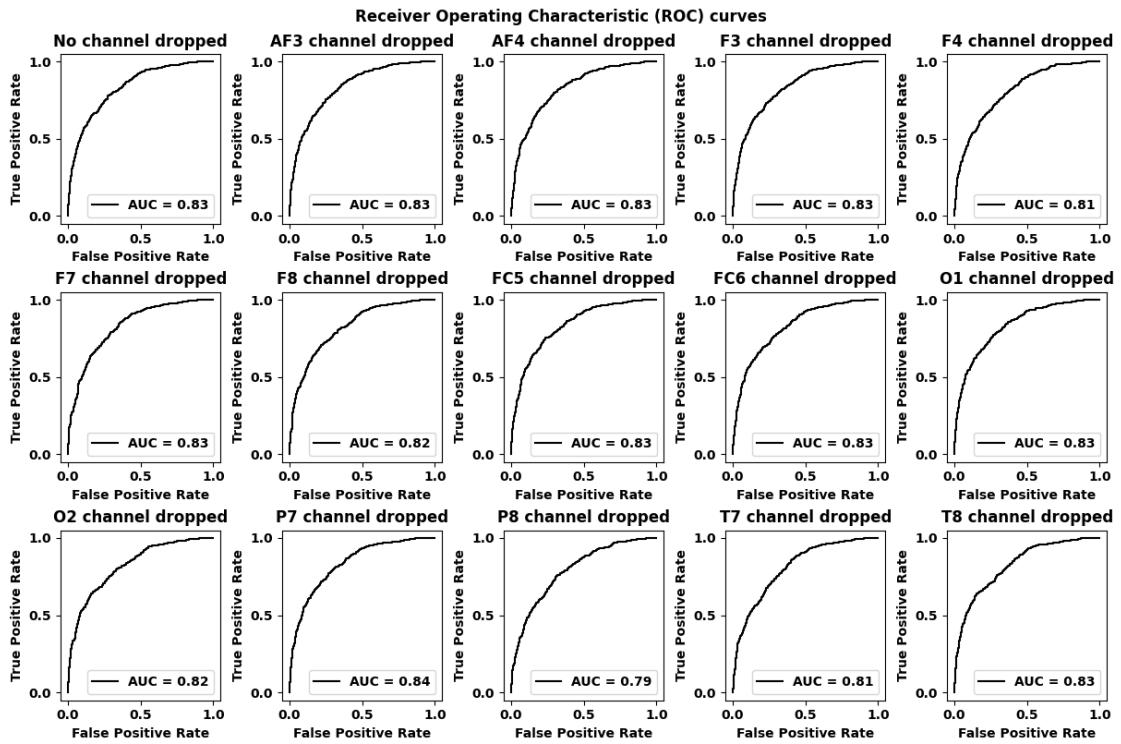


Figure 5.7: ROC curves for Guinea-Bissau dataset using LR

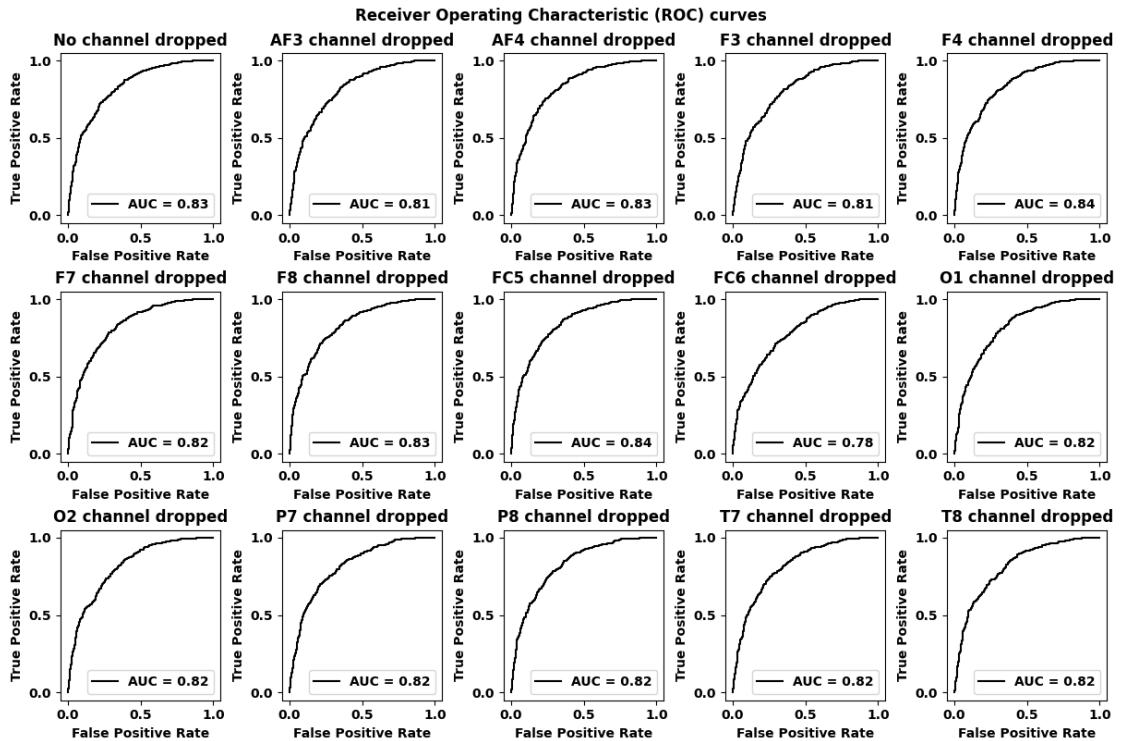


Figure 5.8: ROC curves for Nigeria dataset using LR

5.2.2 Gaussian Naive Bayes (GNB):

Figure 5.9 and Figure 5.10 illustrate the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Gaussian Naive Bayes.

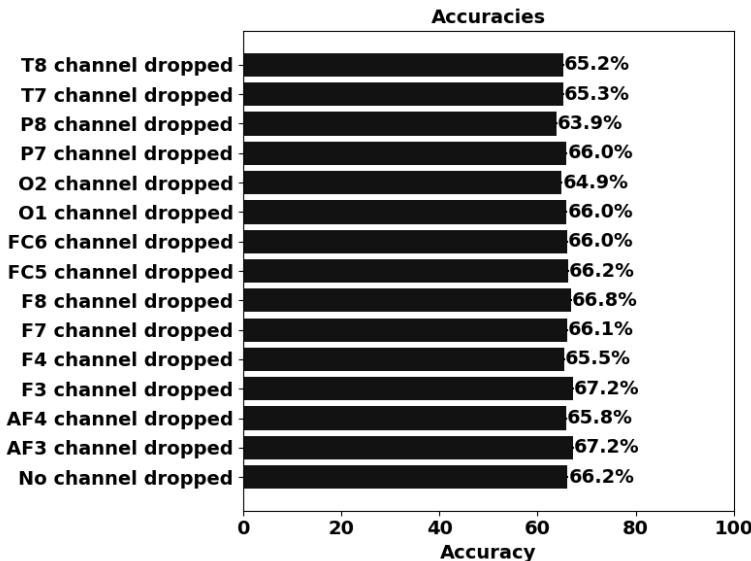


Figure 5.9: Accuracies for Guinea-Bissau dataset using GNB

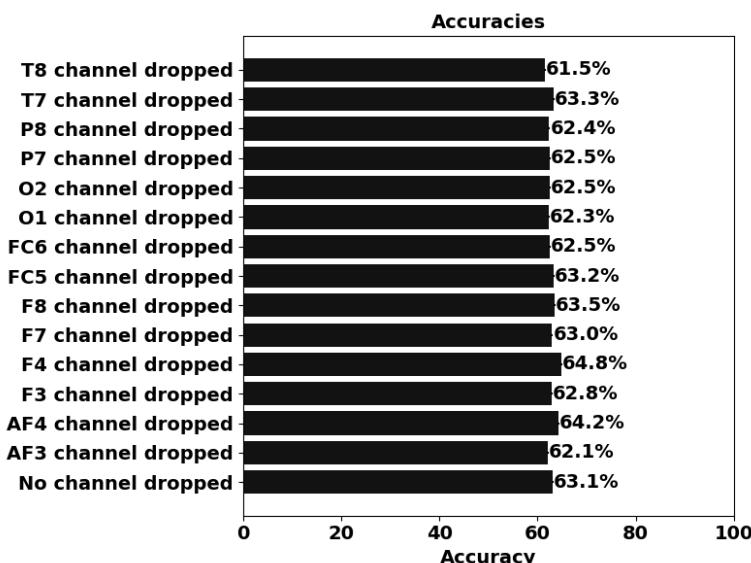


Figure 5.10: Accuracies for Nigeria dataset using GNB

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.11 and Figure 5.12 respectively using Gaussian Naive Bayes.

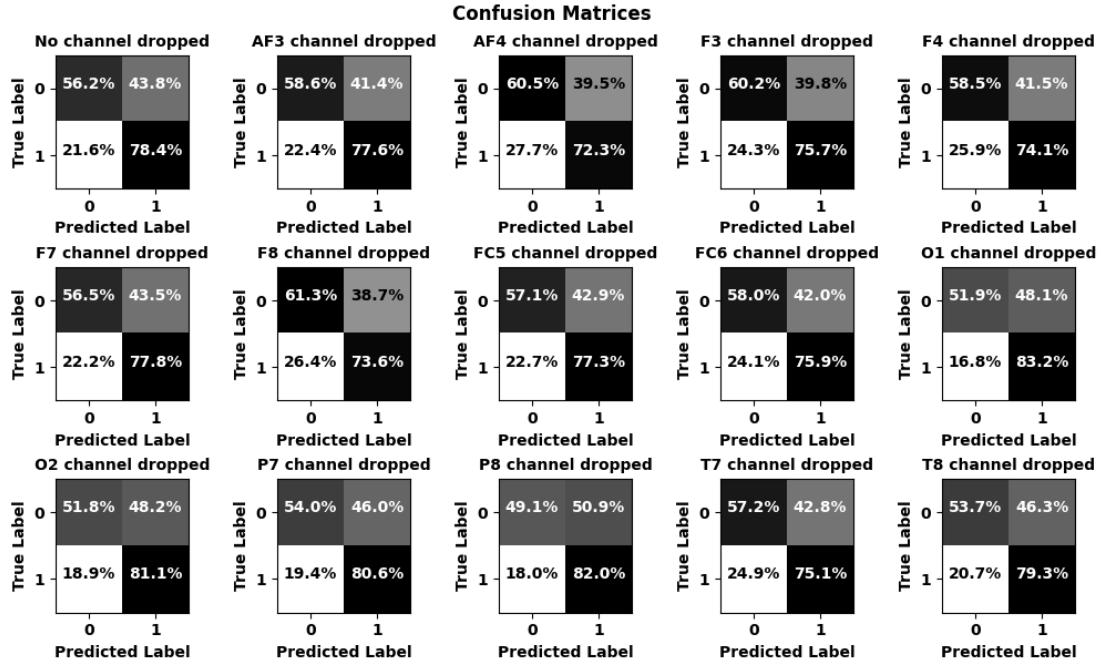


Figure 5.11: Confusion Matrices for Guinea-Bissau dataset using GNB

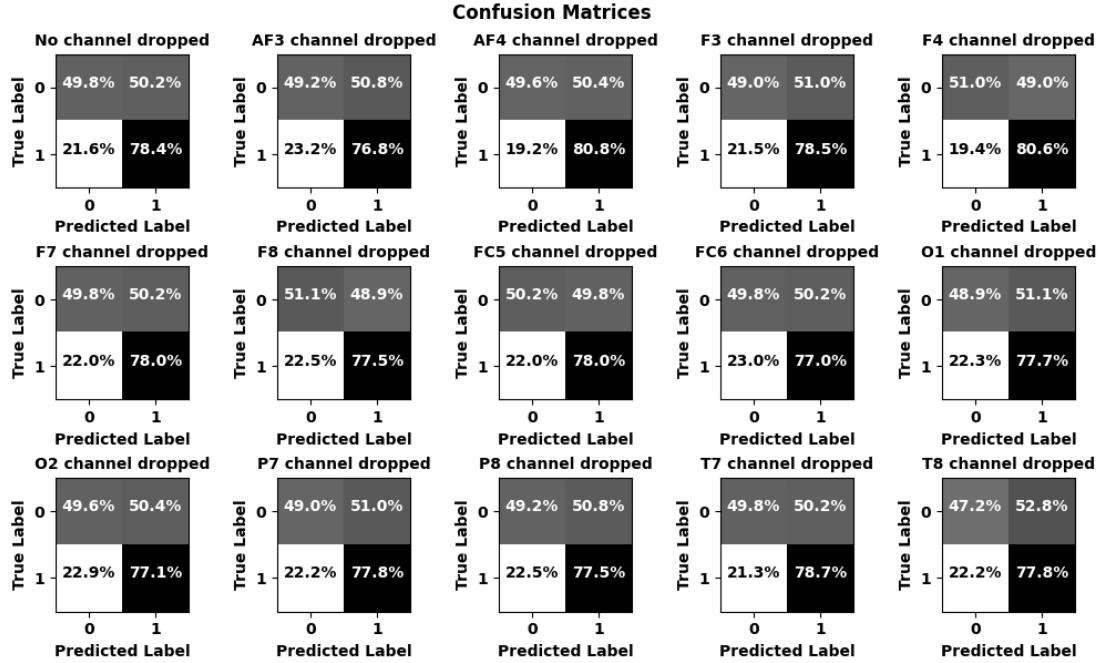


Figure 5.12: Confusion Matrices for Nigeria dataset using GNB

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.13 and Figure 5.14 respectively using Gaussian Naive Bayes.

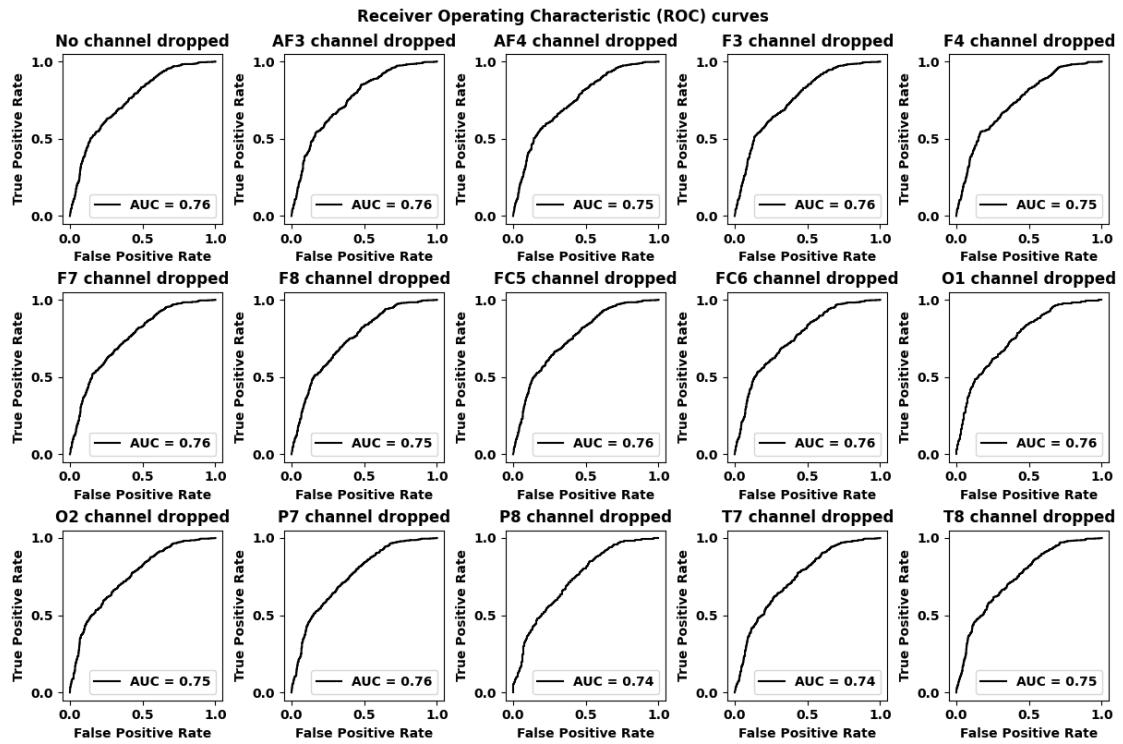


Figure 5.13: ROC curves for Guinea-Bissau dataset using GNB

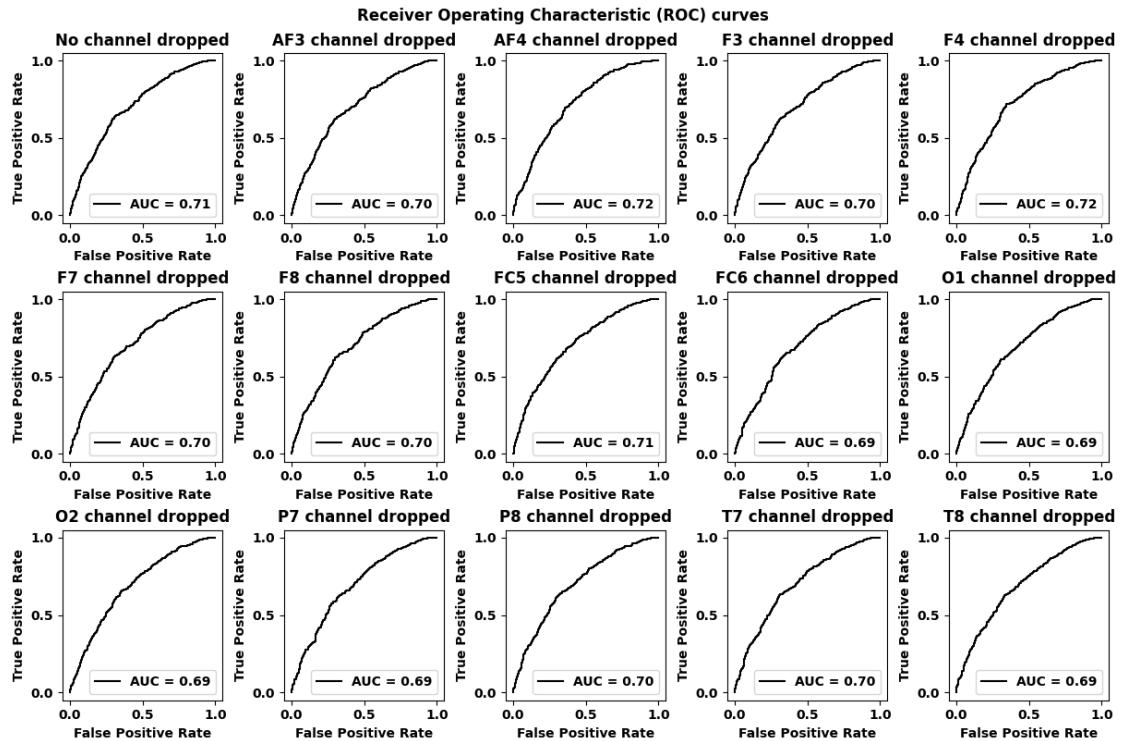


Figure 5.14: ROC curves for Nigeria dataset using GNB

5.2.3 Decision Tree (DT):

Figure 5.15 and Figure 5.16 display the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Decision Tree.

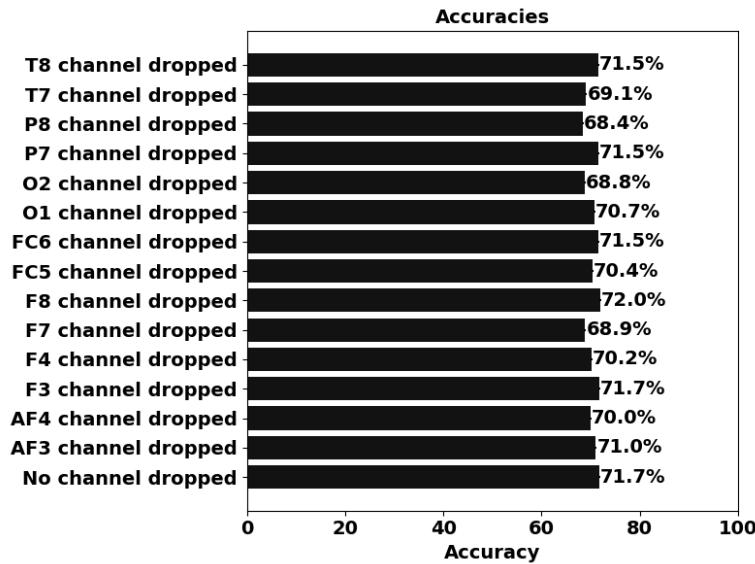


Figure 5.15: Accuracies for Guinea-Bissau dataset using DT

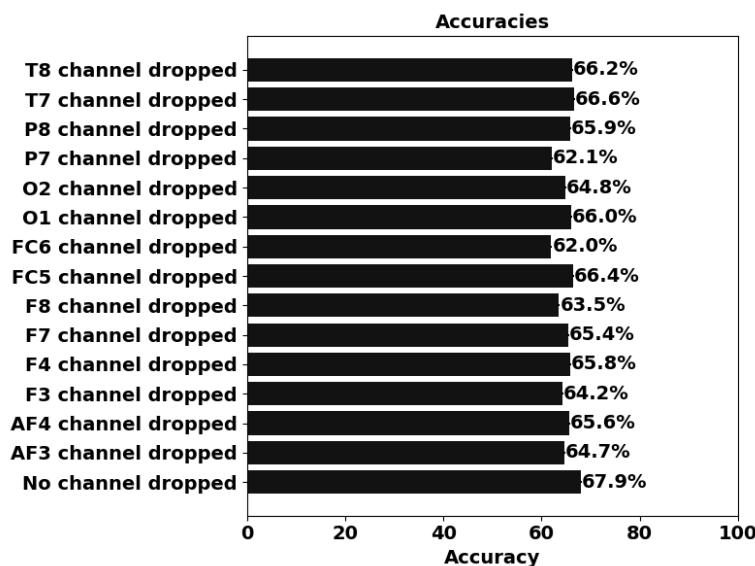


Figure 5.16: Accuracies for Nigeria dataset using DT

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.17 and Figure 5.18 respectively using Decision Tree.

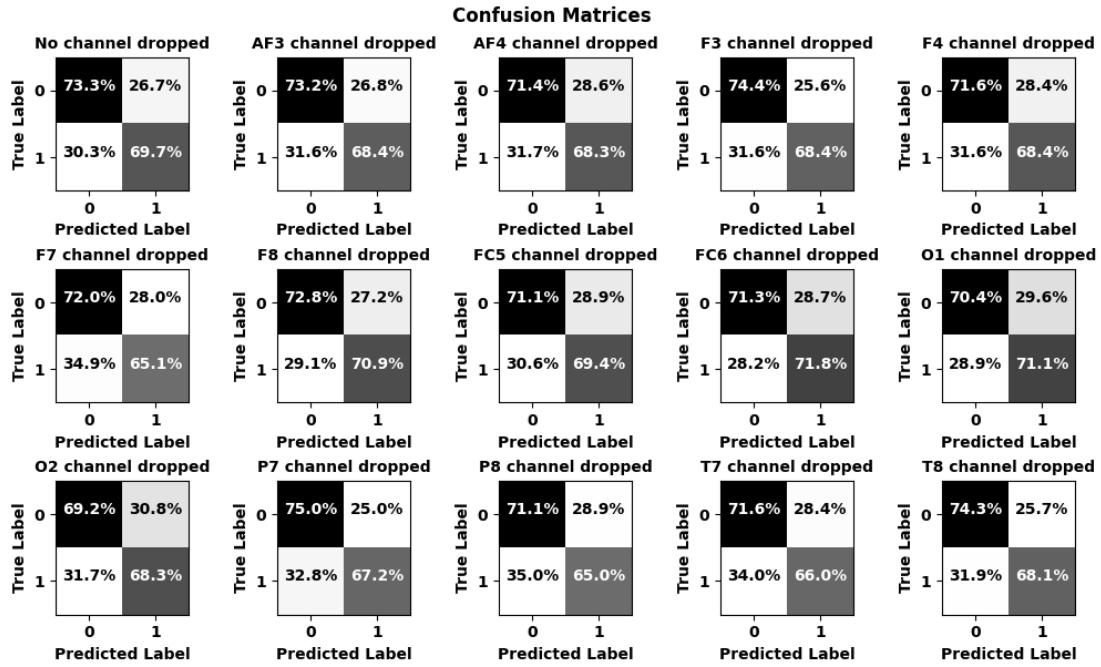


Figure 5.17: Confusion Matrices for Guinea-Bissau dataset using DT

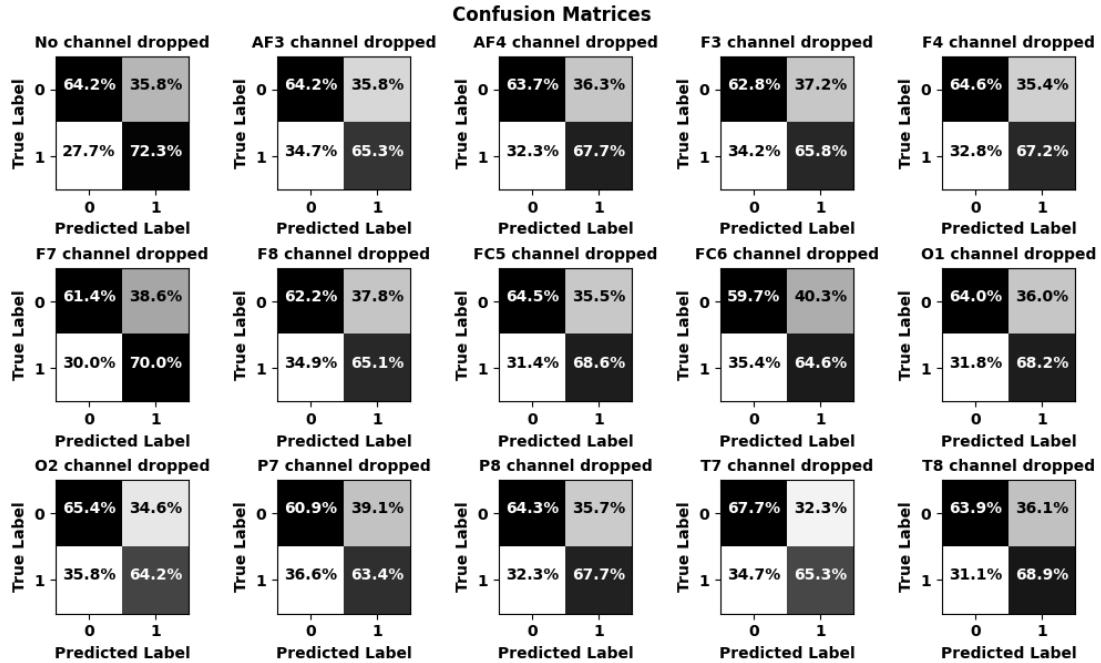


Figure 5.18: Confusion Matrices for Nigeria dataset using DT

Figure 5.19 and Figure 5.20 show the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Decision Tree.

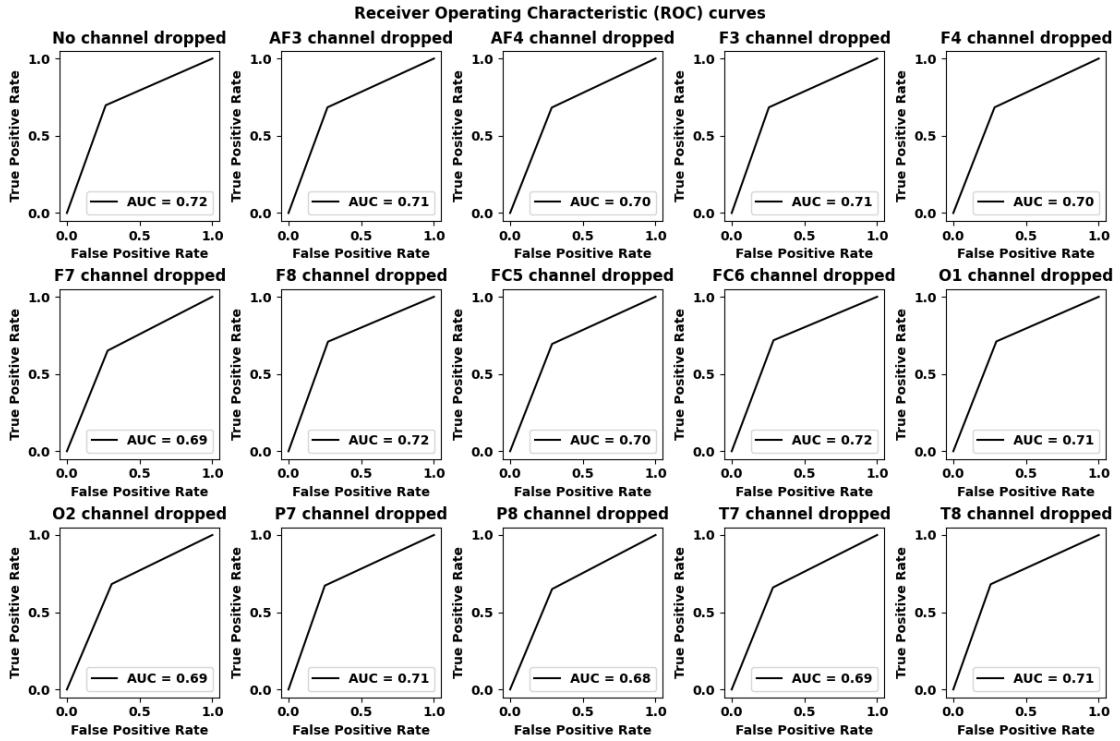


Figure 5.19: ROC curves for Guinea-Bissau dataset using DT

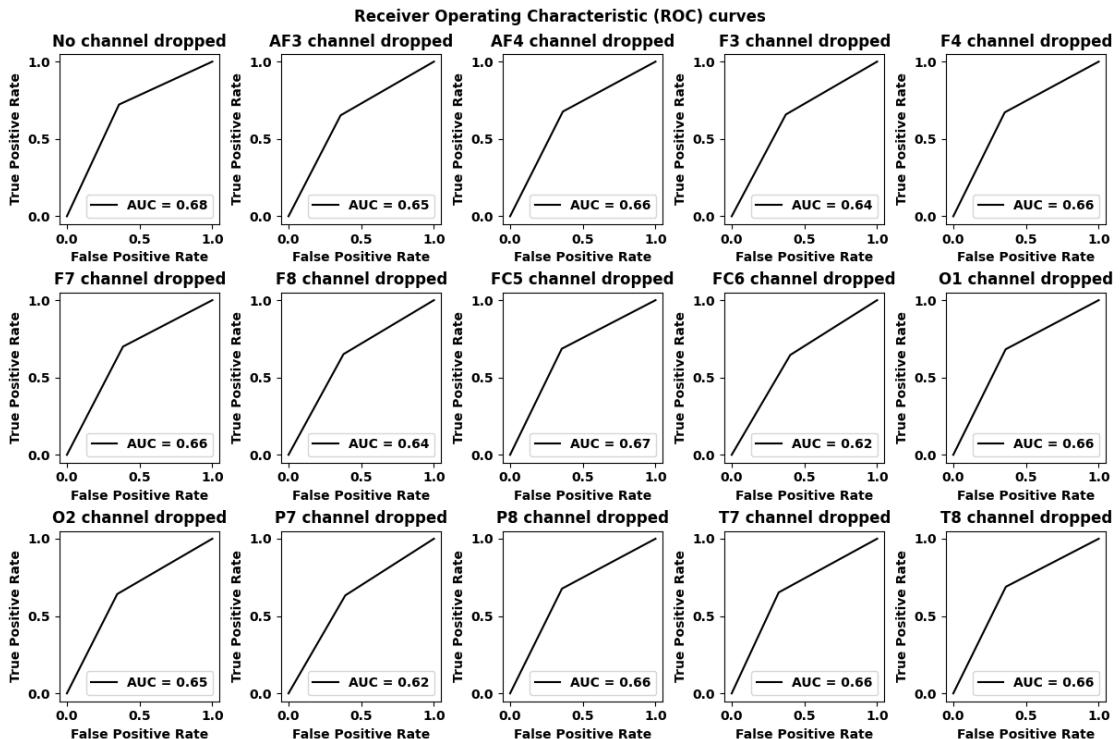


Figure 5.20: ROC curves for Nigeria dataset using DT

The number of points in an ROC curve for a DT is typically limited because decision trees inherently make binary predictions. Unlike some other classifiers that can output

probabilities for multiple thresholds, decision trees usually yield discrete class labels (e.g., 0 or 1).

5.2.4 Random Forest Tree (RFT):

Figure 5.21 and Figure 5.22 represent the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Random Forest Tree.

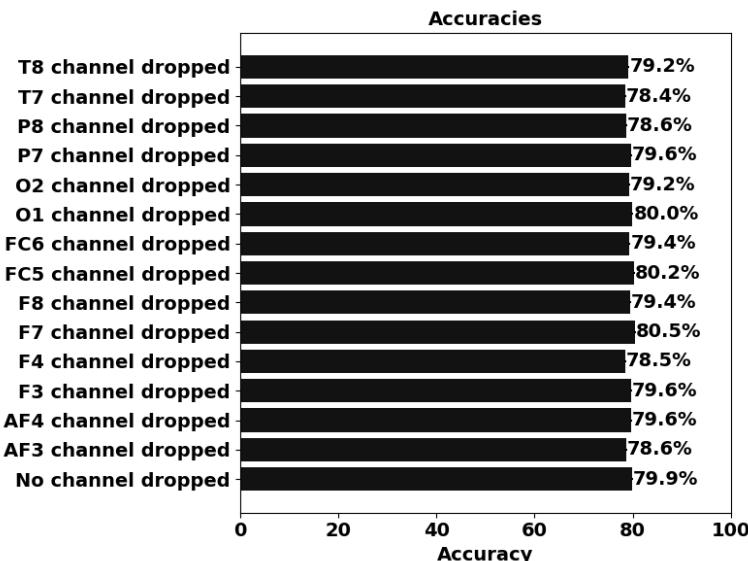


Figure 5.21: Accuracies for Guinea-Bissau dataset using RFT

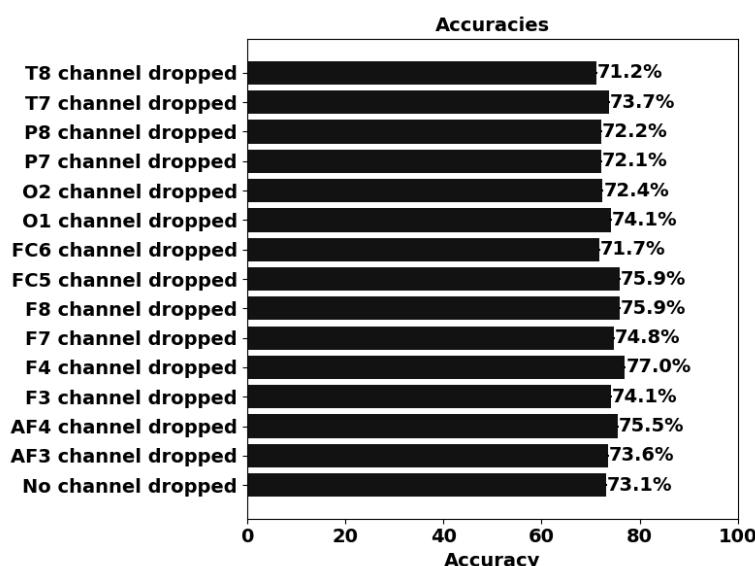


Figure 5.22: Accuracies for Nigeria dataset using RFT

Figure 5.23 and Figure 5.24 demonstrate the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Random Forest Tree.

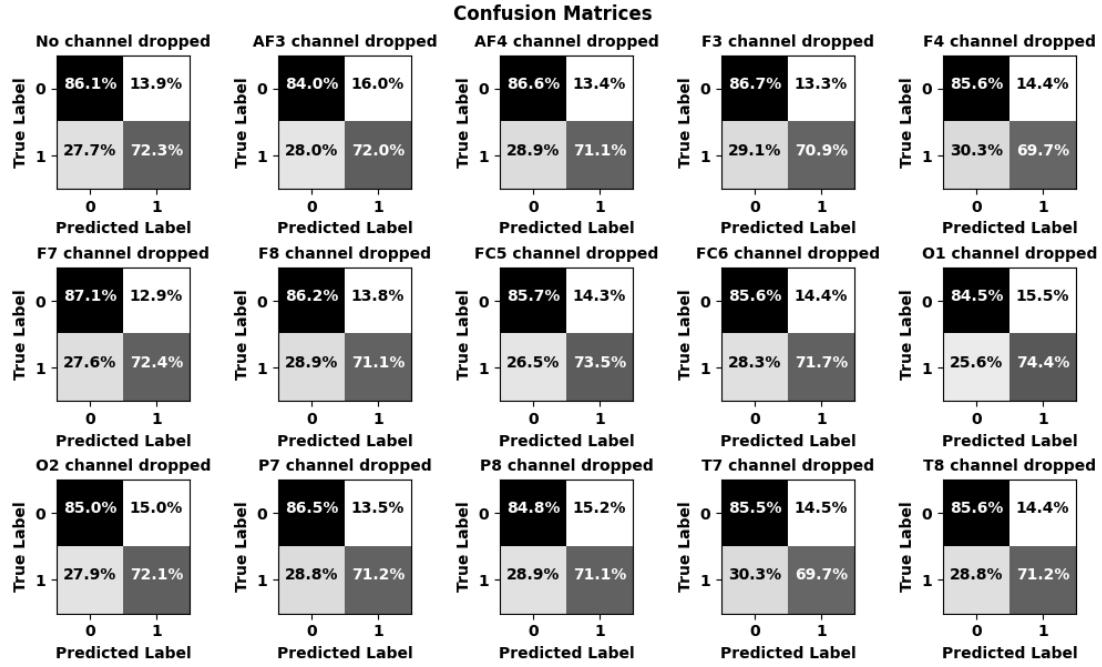


Figure 5.23: Confusion Matrices for Guinea-Bissau dataset using RFT

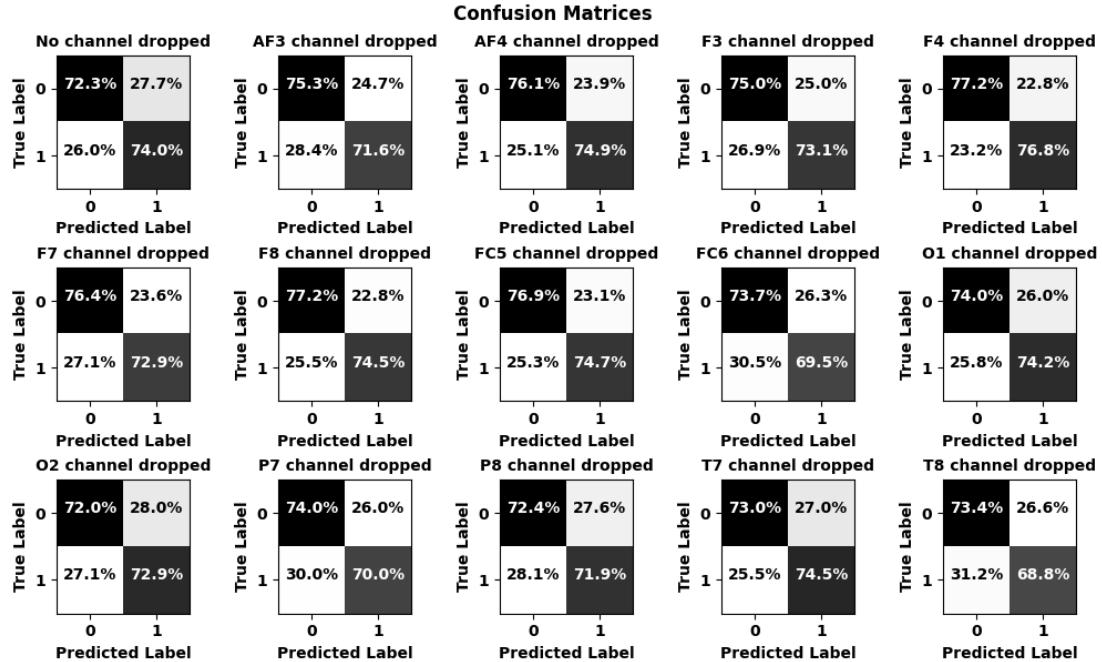


Figure 5.24: Confusion Matrices for Nigeria dataset using RFT

Figure 5.25 and Figure 5.26 illustrate the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Random Forest Tree.

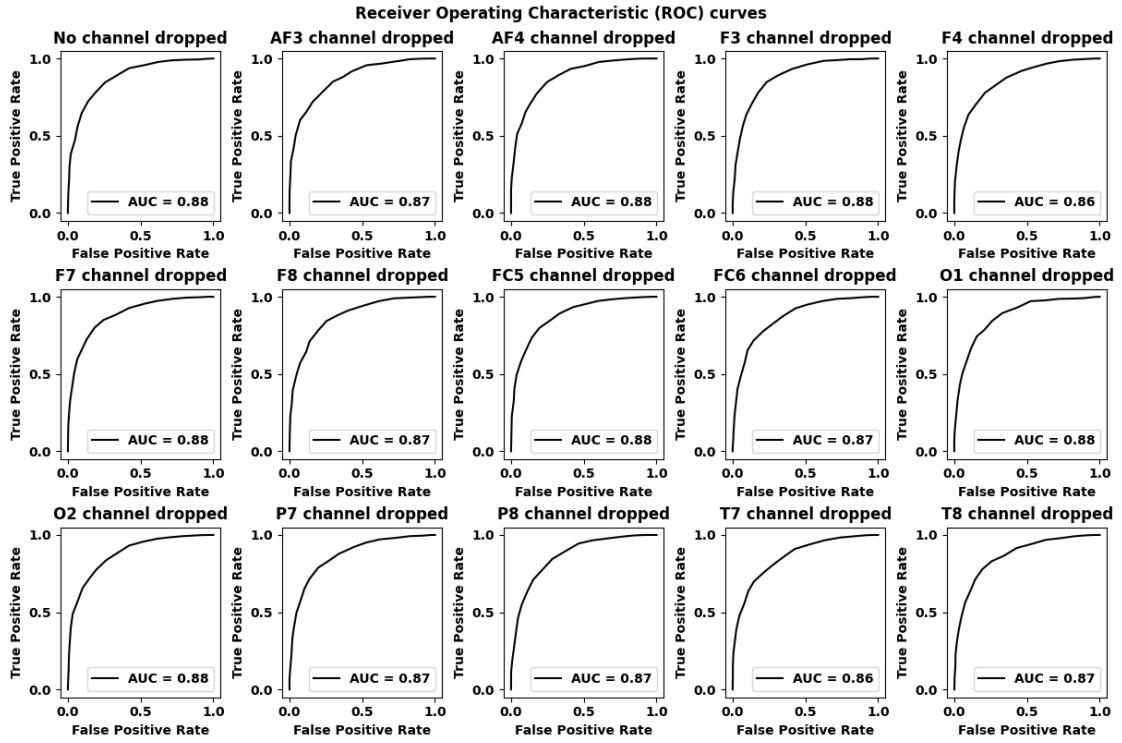


Figure 5.25: ROC curves for Guinea-Bissau dataset using RFT

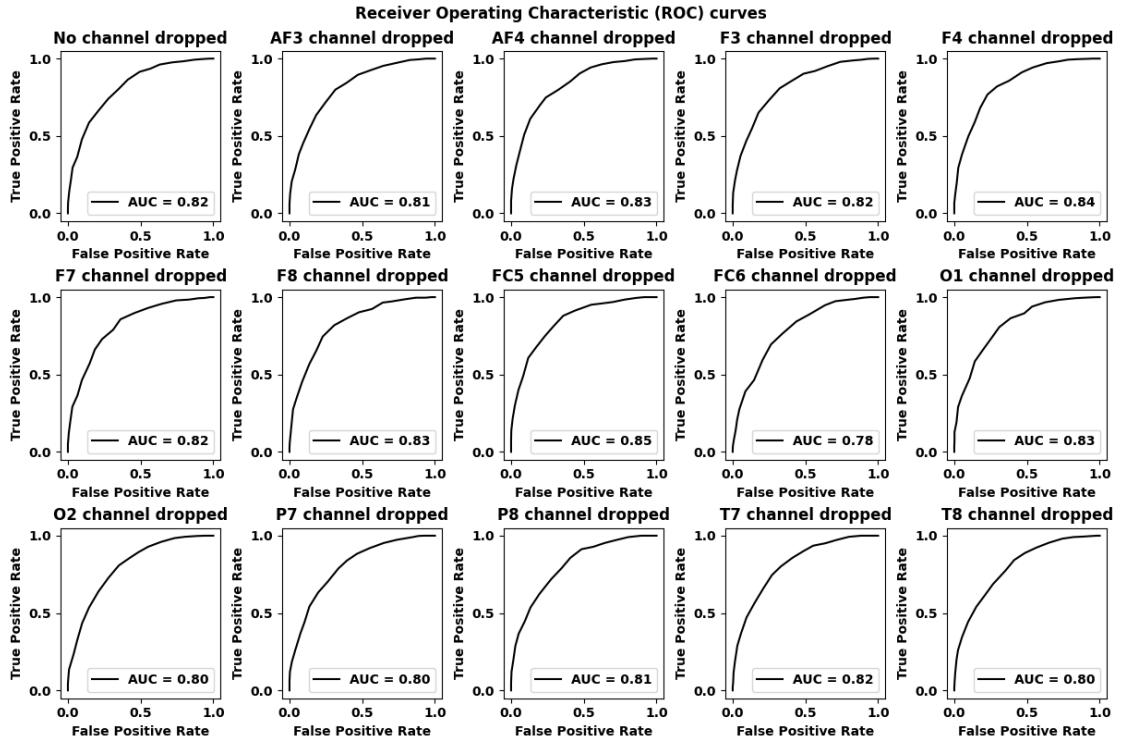


Figure 5.26: ROC curves for Nigeria dataset using RFT

5.2.5 K-Nearest Neighbour (KNN):

The accuracies for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.27 and Figure 5.28 respectively using k-Nearest Neighbour.

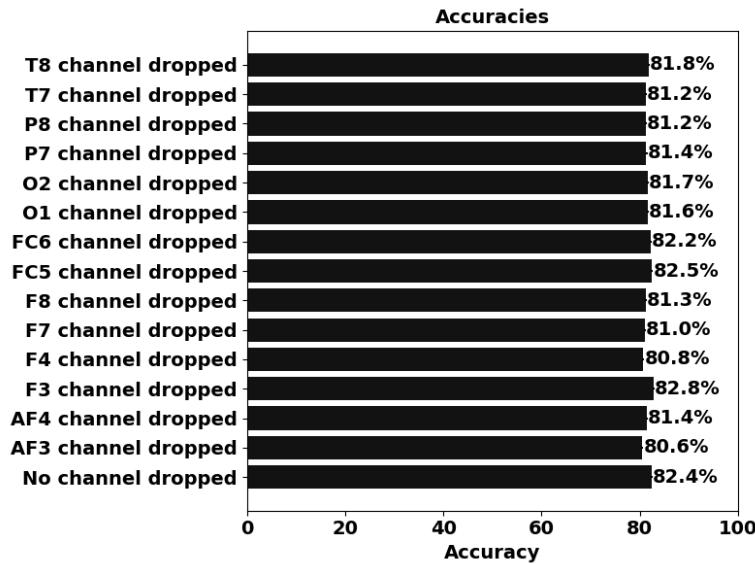


Figure 5.27: Accuracies for Guinea-Bissau dataset using KNN

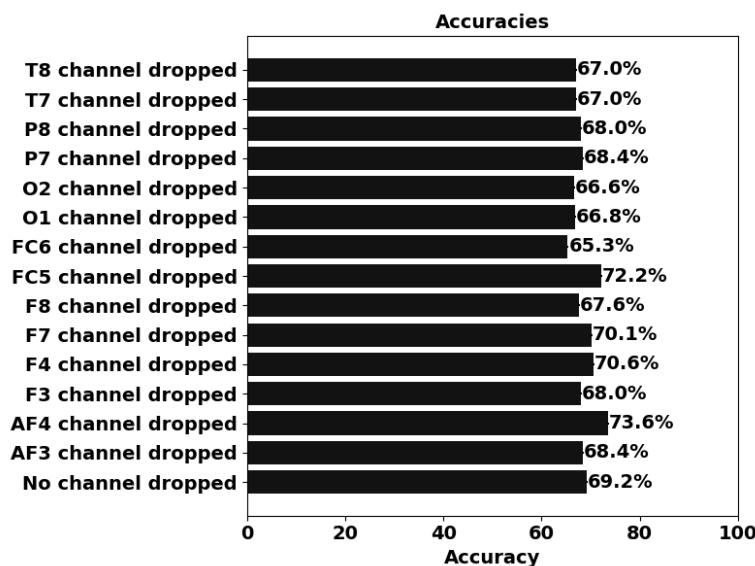


Figure 5.28: Accuracies for Nigeria dataset using KNN

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.29 and Figure 5.30 respectively using k-Nearest Neighbour.

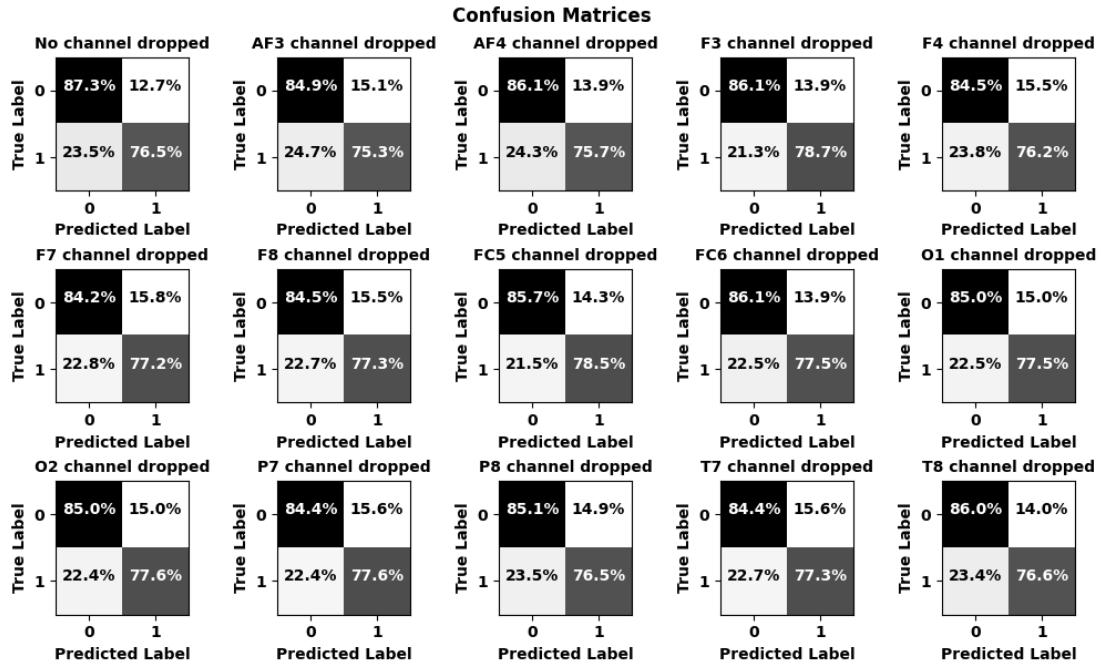


Figure 5.29: Confusion Matrices for Guinea-Bissau dataset using KNN

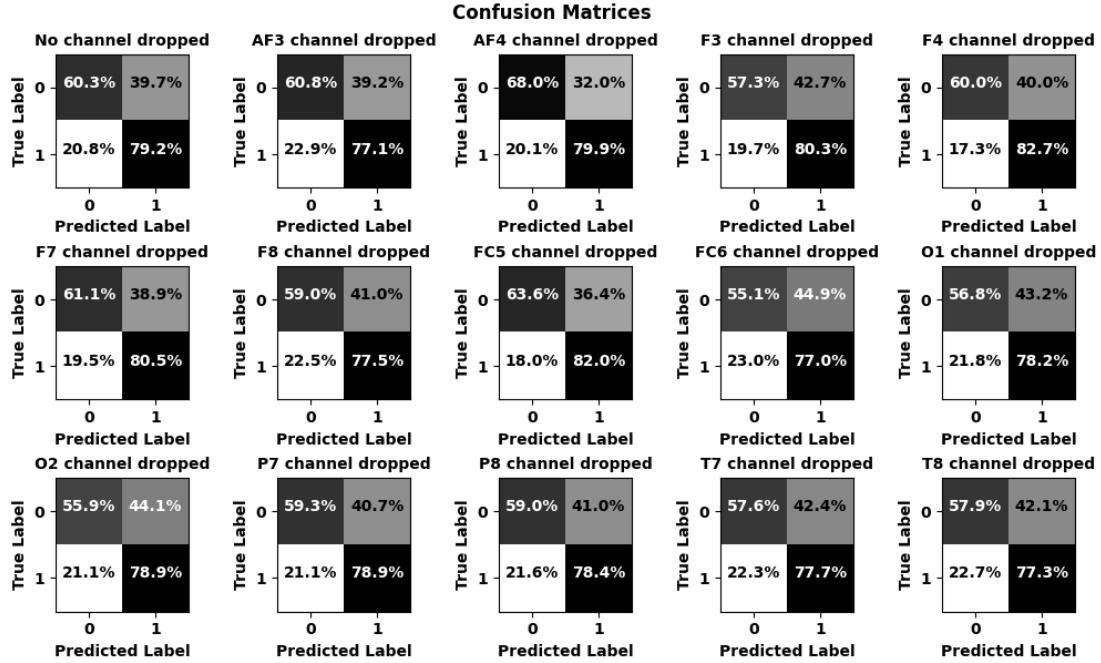


Figure 5.30: Confusion Matrices for Nigeria dataset using KNN

Figure 5.31 and Figure 5.32 display the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using k-Nearest Neighbour.

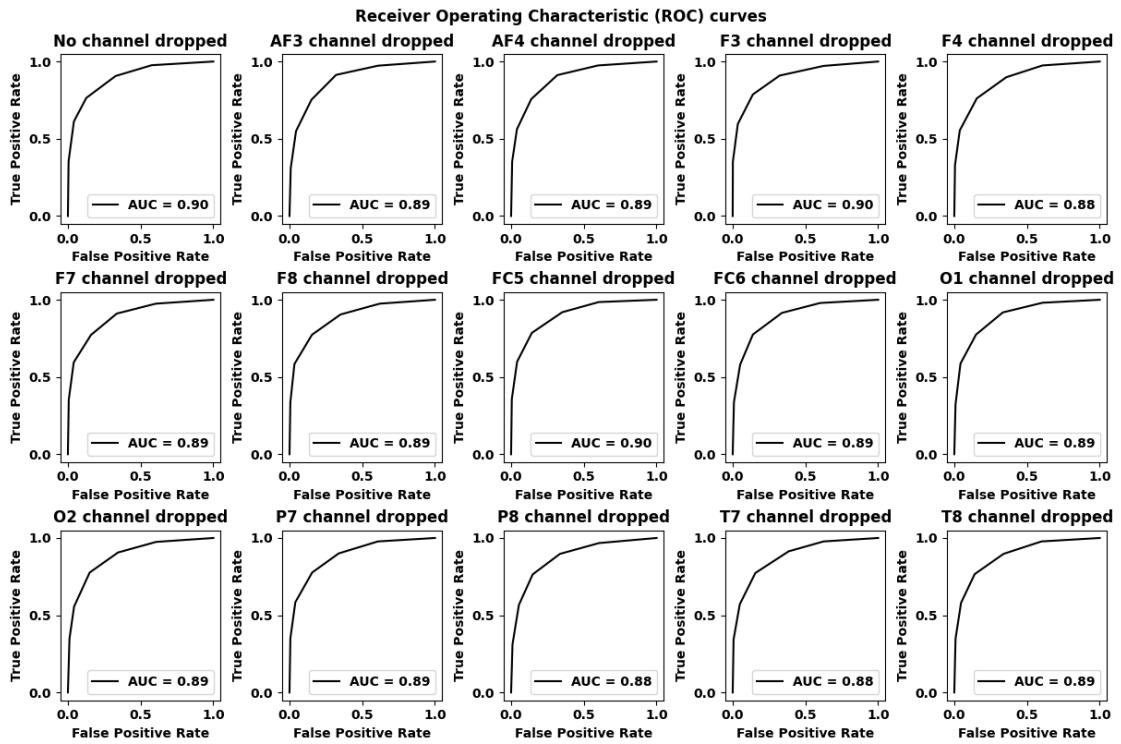


Figure 5.31: ROC curves for Guinea-Bissau dataset using KNN

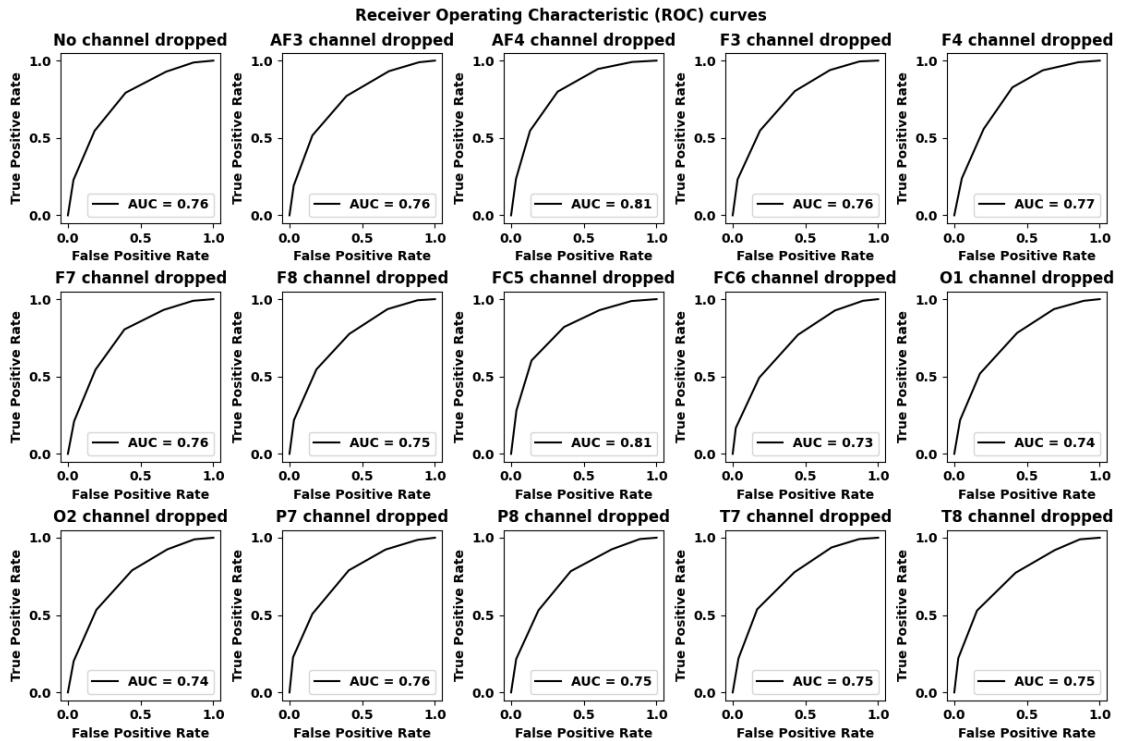


Figure 5.32: ROC curves for Nigeria dataset using KNN

5.2.6 Support Vector Machine (SVM):

The accuracies for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.33 and Figure 5.34 respectively using Support Vector Machine.

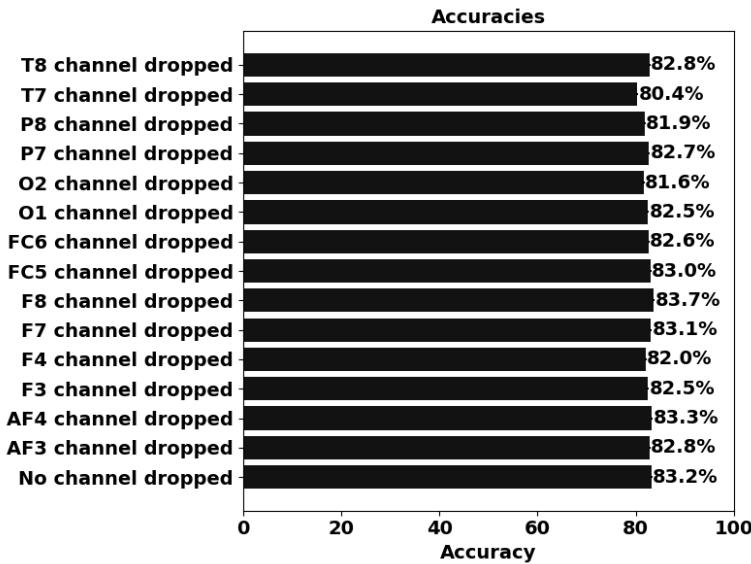


Figure 5.33: Accuracies for Guinea-Bissau dataset using SVM

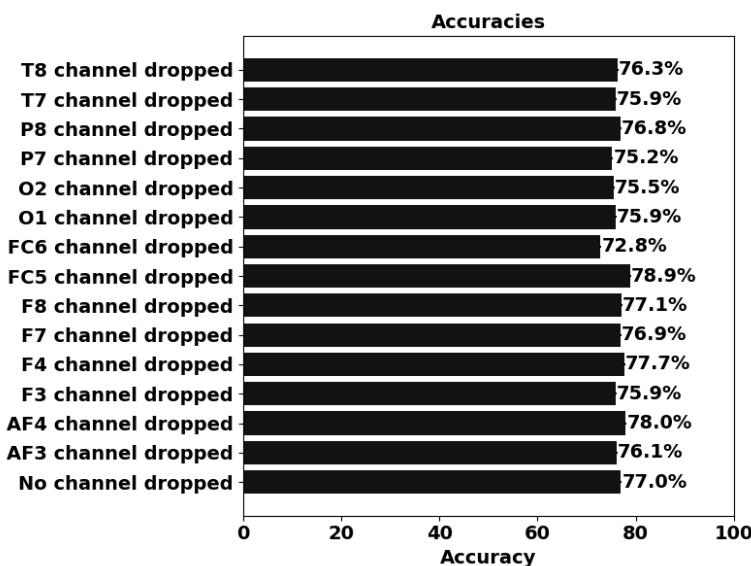


Figure 5.34: Accuracies for Nigeria dataset using SVM

Figure 5.35 and Figure 5.36 show the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Support Vector Machine.

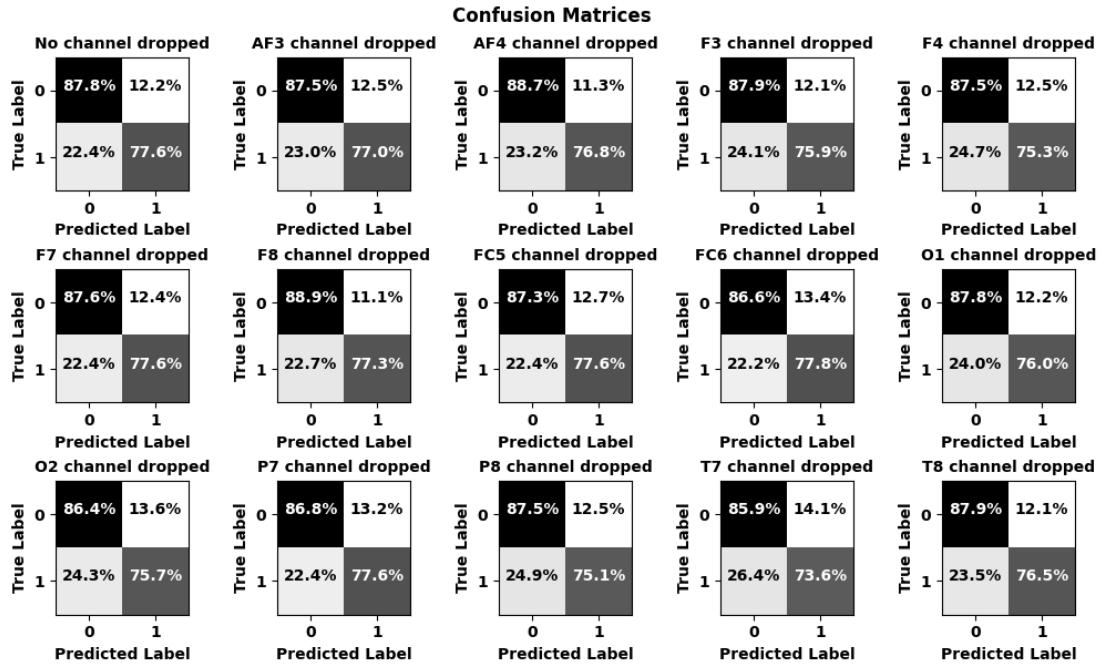


Figure 5.35: Confusion Matrices for Guinea-Bissau dataset using SVM

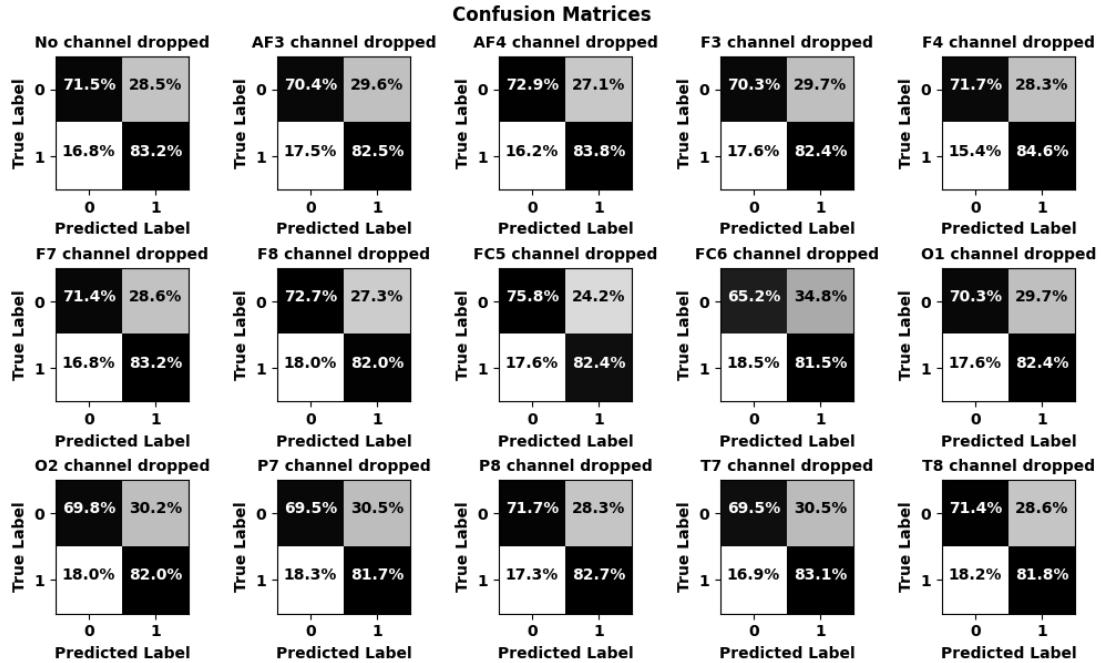


Figure 5.36: Confusion Matrices for Nigeria dataset using SVM

Figure 5.37 and Figure 5.38 represent the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Support Vector Machine.

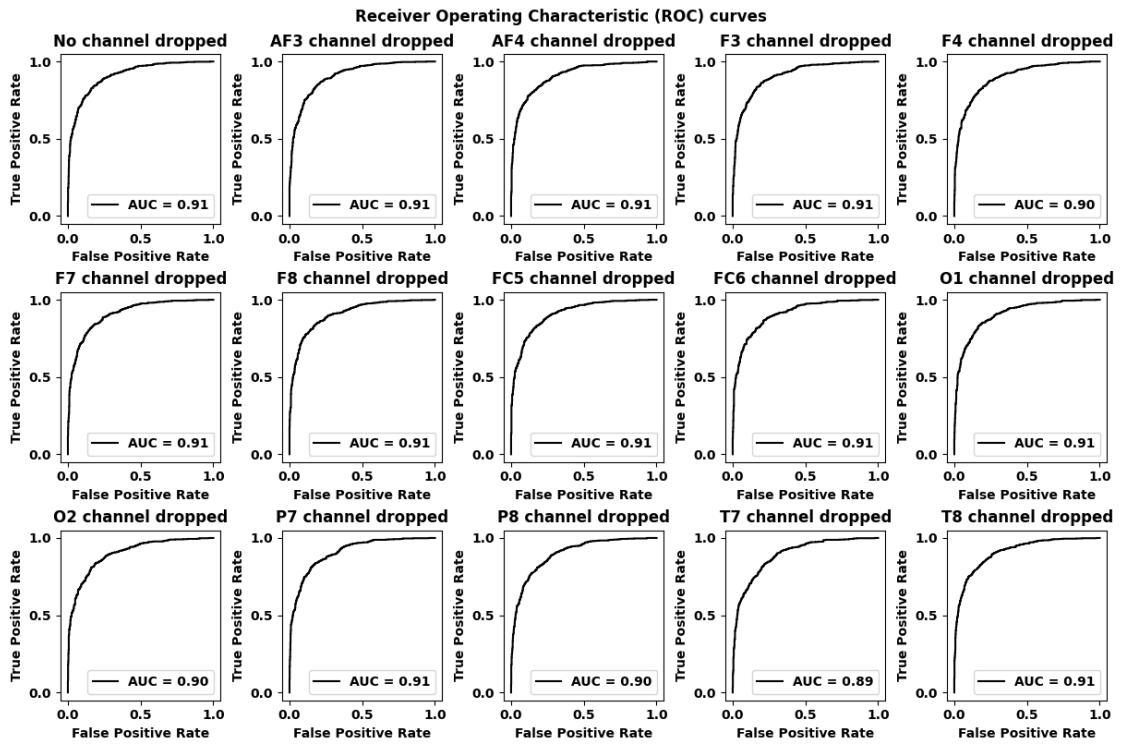


Figure 5.37: ROC curves for Guinea-Bissau dataset using SVM

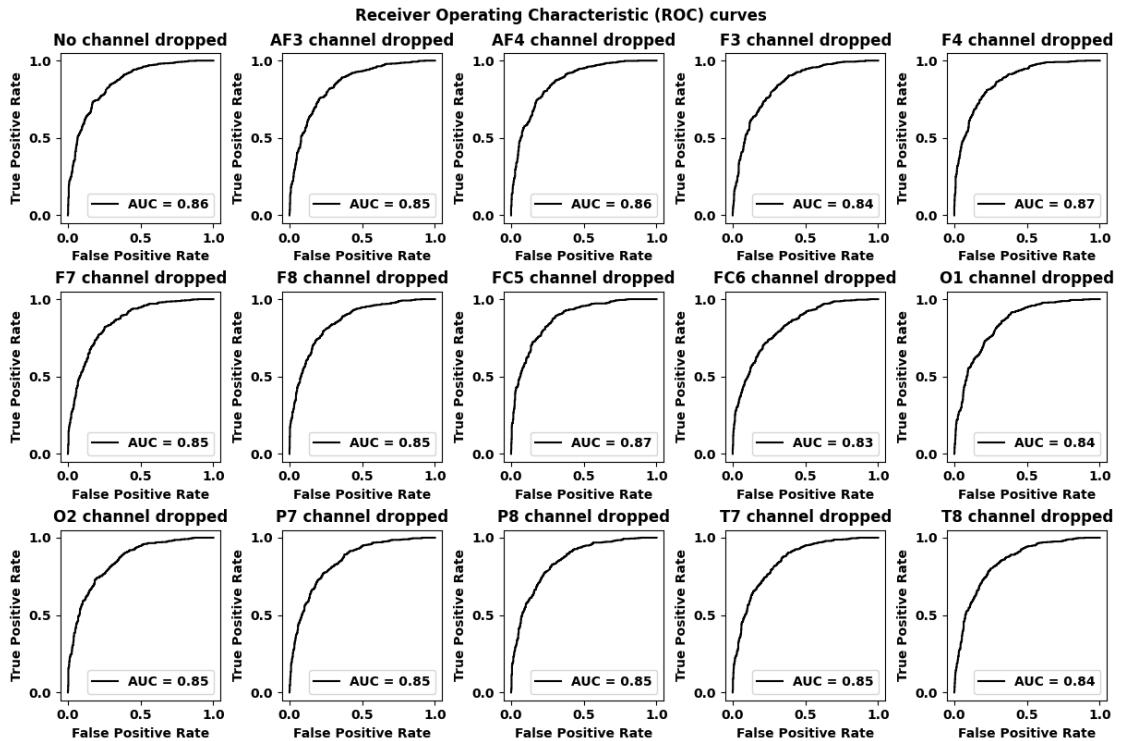


Figure 5.38: ROC curves for Nigeria dataset using SVM

5.2.7 Vanilla Neural Network (VNN):

Figure 5.39 and Figure 5.40 demonstrate the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Vanilla Neural Network.

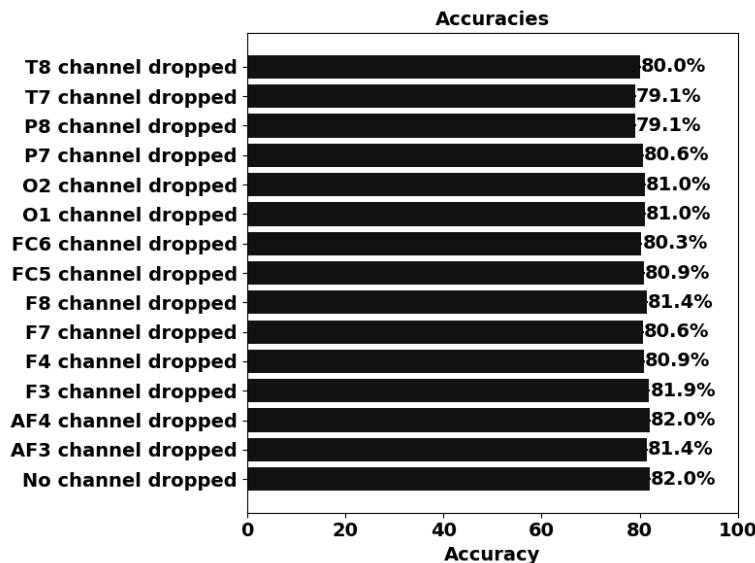


Figure 5.39: Accuracies for Guinea-Bissau dataset using VNN

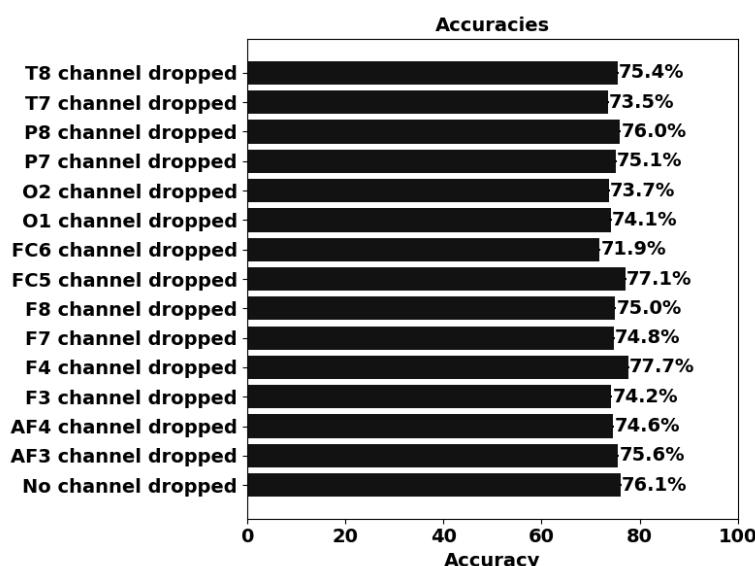


Figure 5.40: Accuracies for Nigeria dataset using VNN

Figure 5.41 and Figure 5.42 illustrate the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Vanilla Neural Network.

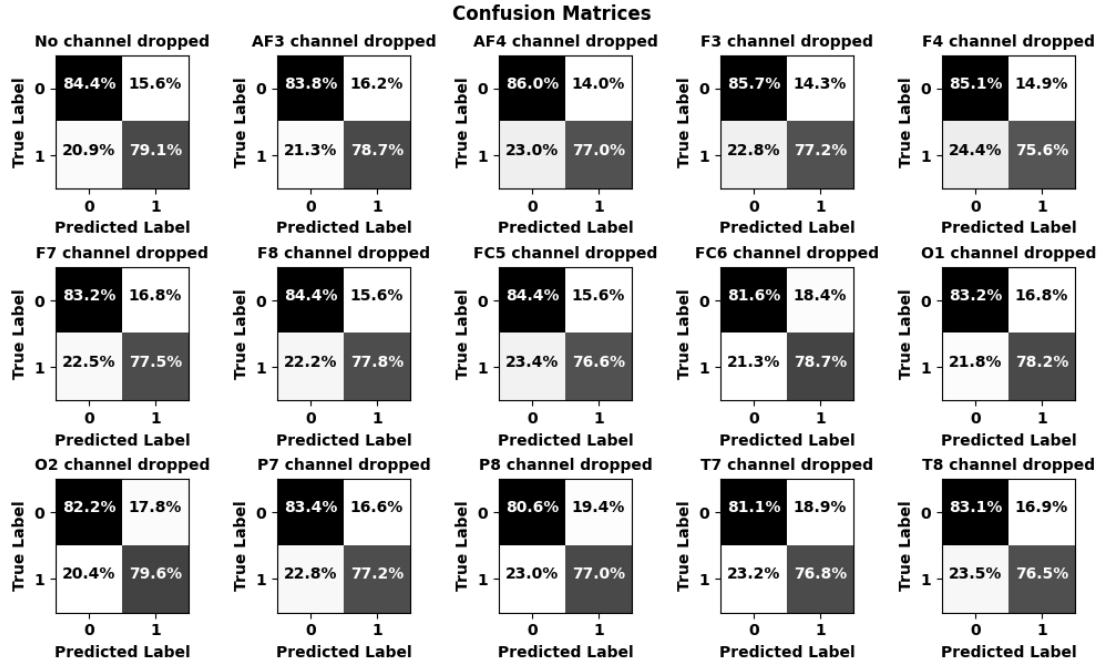


Figure 5.41: Confusion Matrices for Guinea-Bissau dataset using VNN

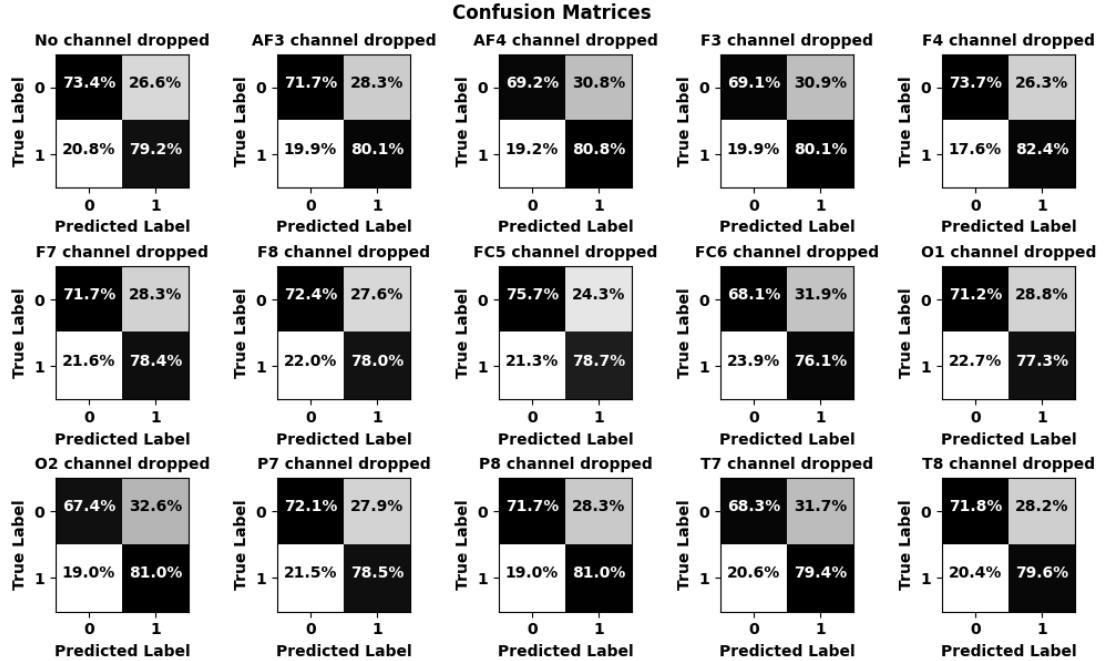


Figure 5.42: Confusion Matrices for Nigeria dataset using VNN

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.43 and Figure 5.44 respectively using Vanilla Neural Network.

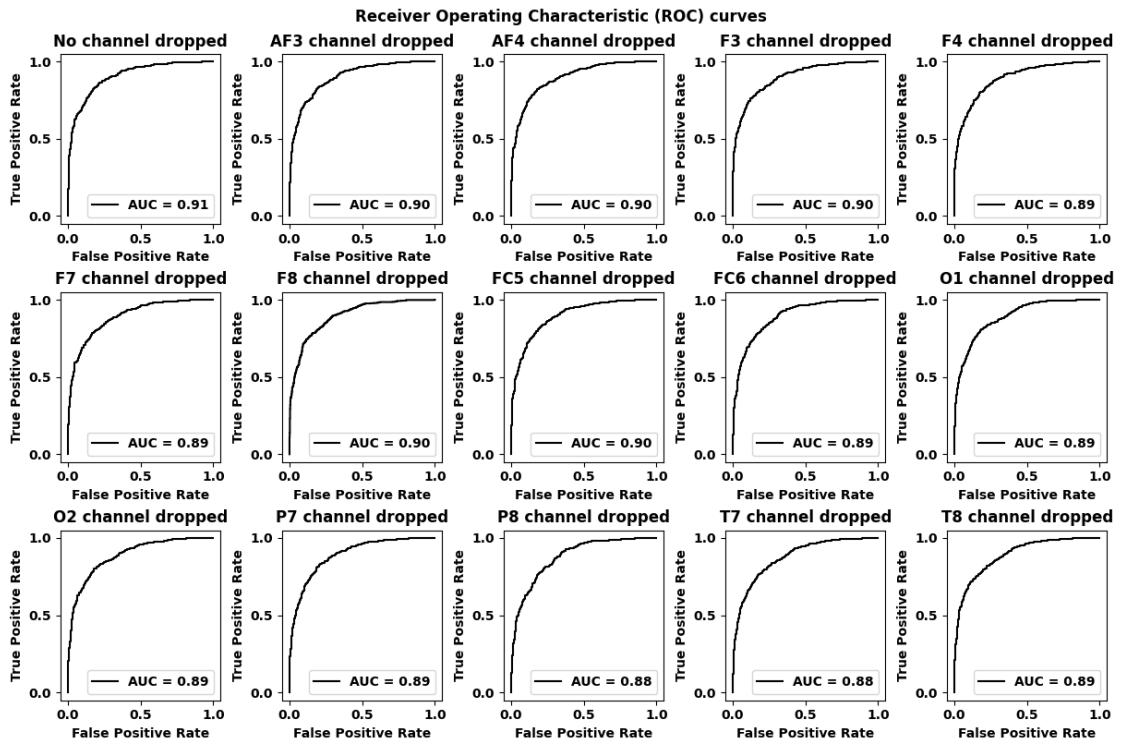


Figure 5.43: ROC curves for Guinea-Bissau dataset using VNN

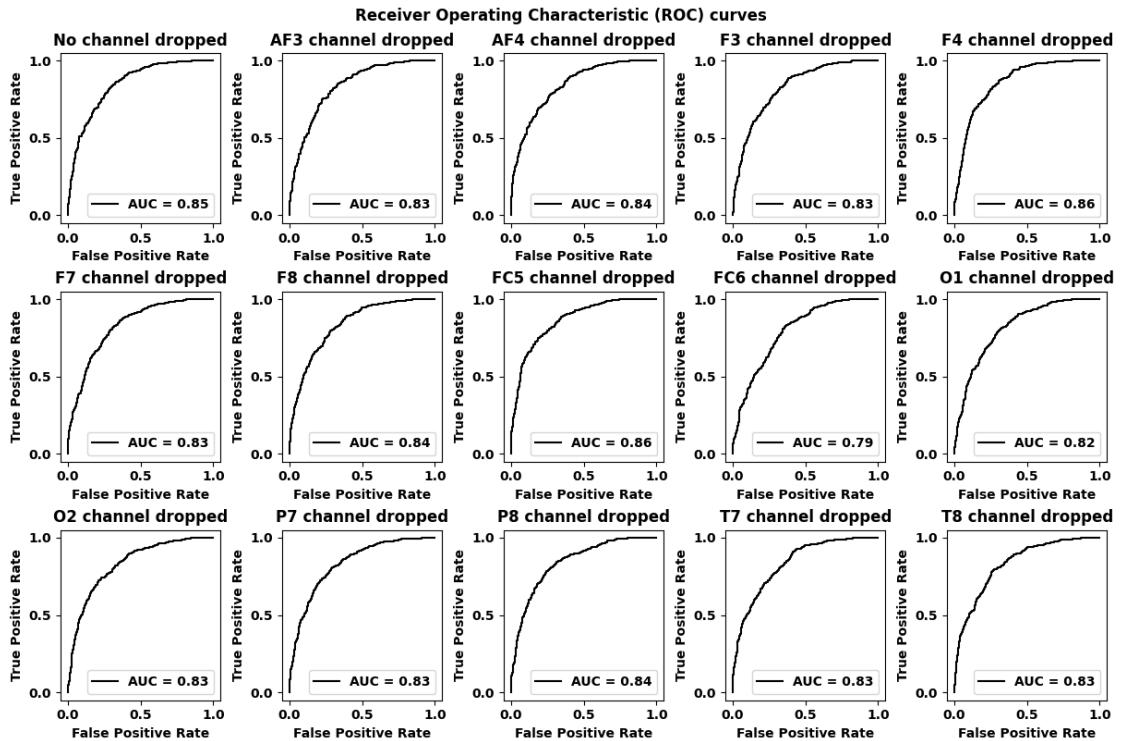


Figure 5.44: ROC curves for Nigeria dataset using VNN

5.2.8 Convolutional Neural Network (CNN):

The accuracies for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.45 and Figure 5.46 respectively using Convolutional Neural Network.

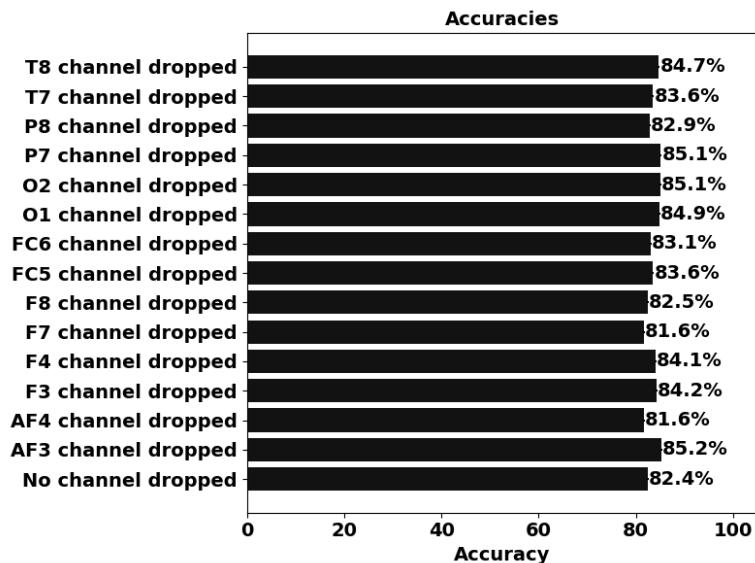


Figure 5.45: Accuracies for Guinea-Bissau dataset using CNN

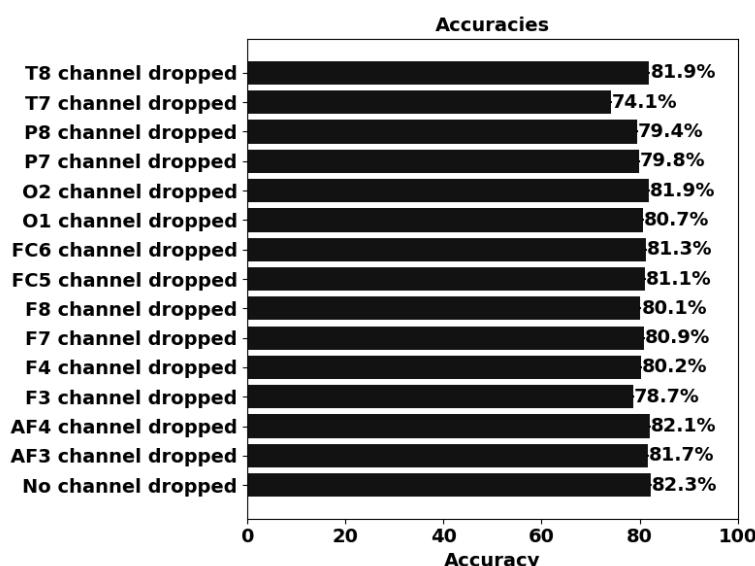


Figure 5.46: Accuracies for Nigeria dataset using CNN

Figure 5.47 and Figure 5.48 display the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Neural Network.

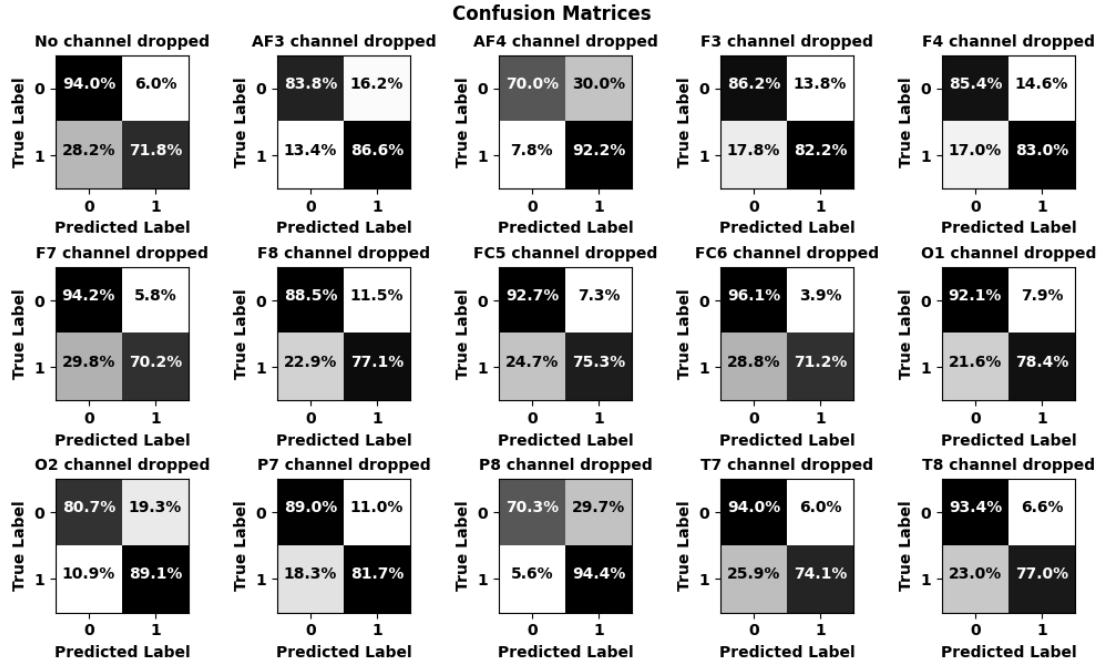


Figure 5.47: Confusion Matrices for Guinea-Bissau dataset using CNN

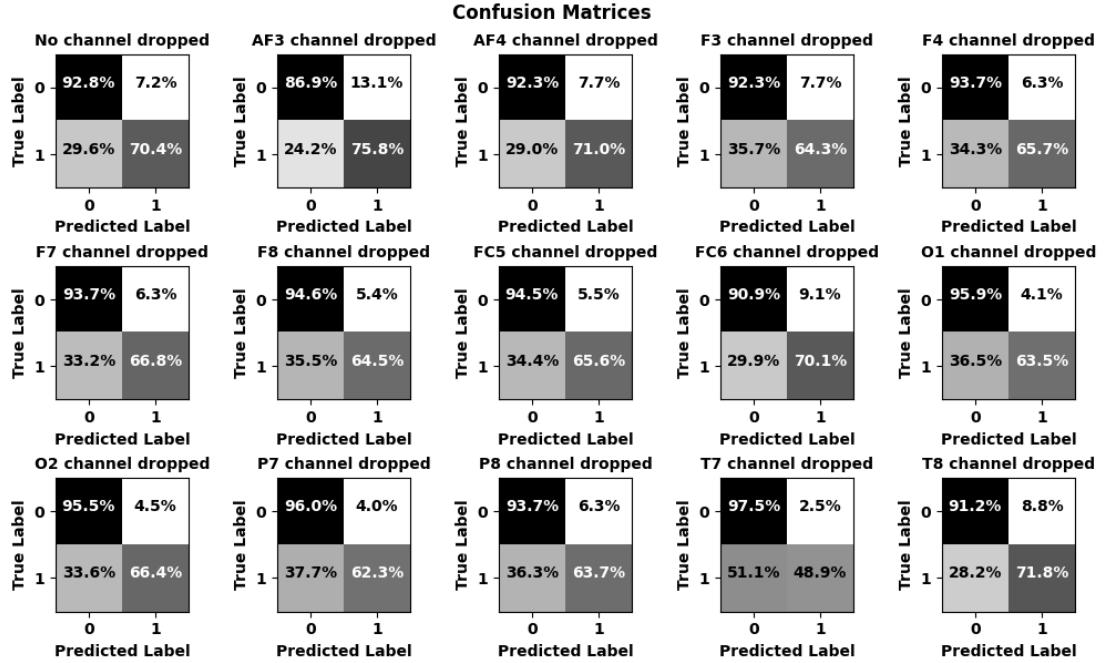


Figure 5.48: Confusion Matrices for Nigeria dataset using CNN

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.49 and Figure 5.50 respectively using Convolutional Neural Network.

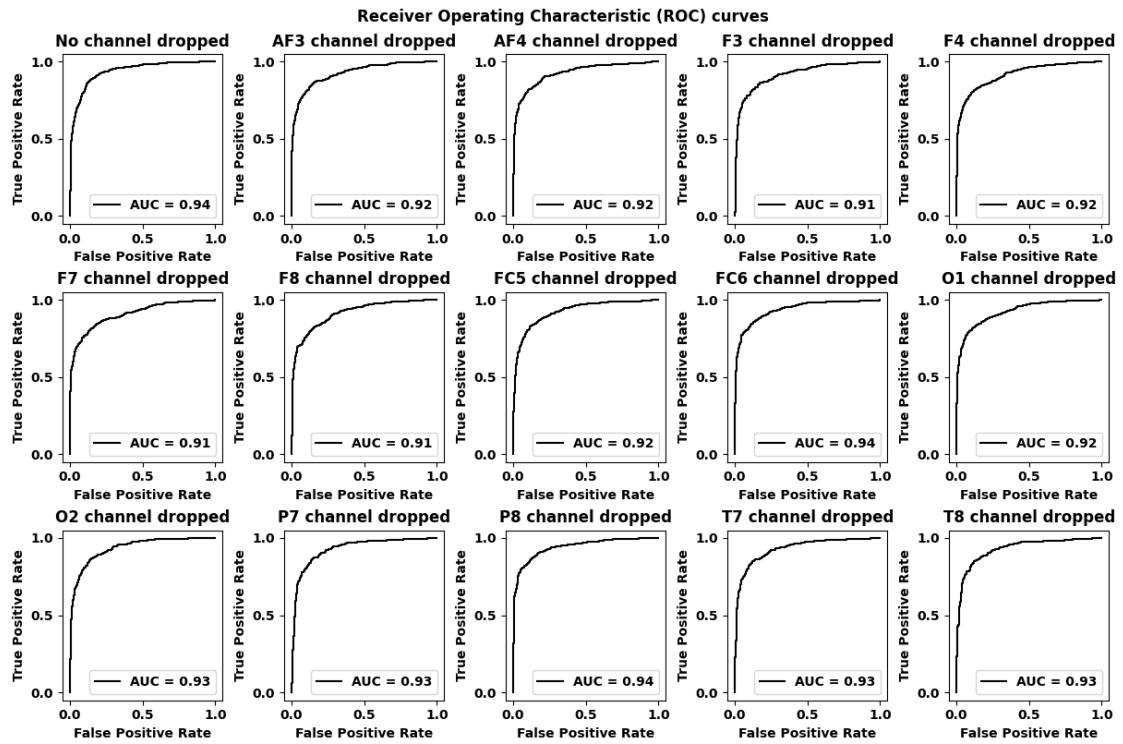


Figure 5.49: ROC curves for Guinea-Bissau dataset using CNN

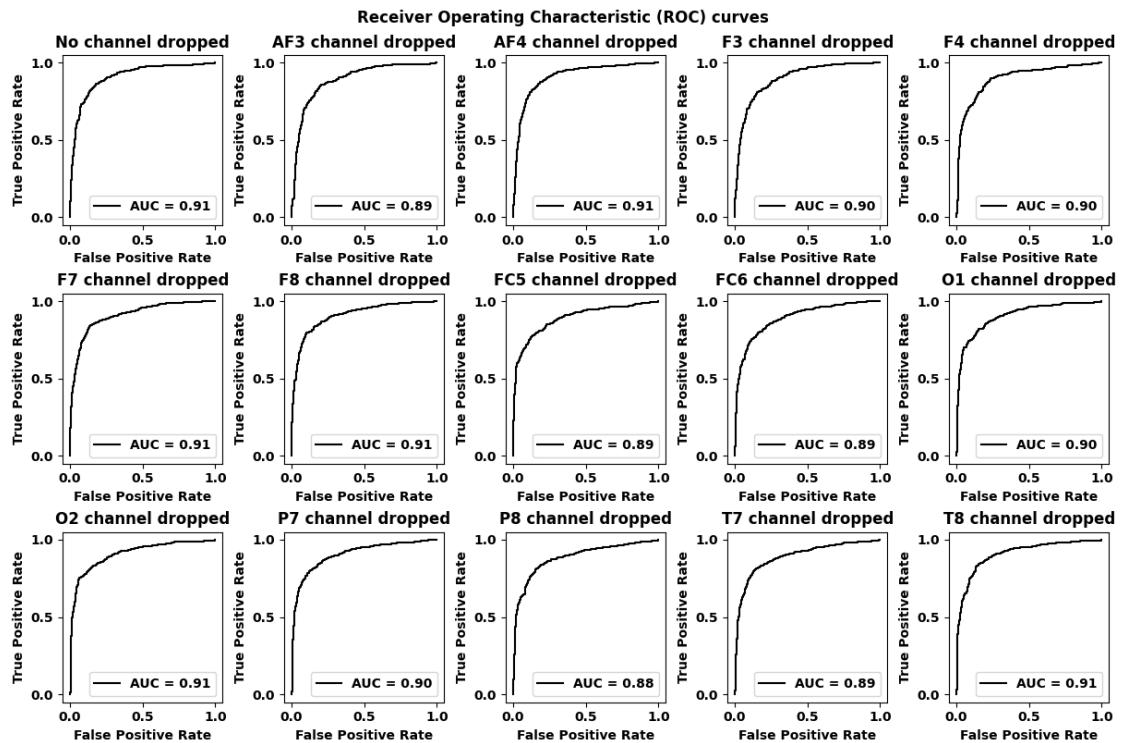


Figure 5.50: ROC curves for Nigeria dataset using CNN

5.2.9 Gated Recurrent Neural Network (RNN):

Figure 5.51 and Figure 5.52 show the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Gated Recurrent Neural Network.

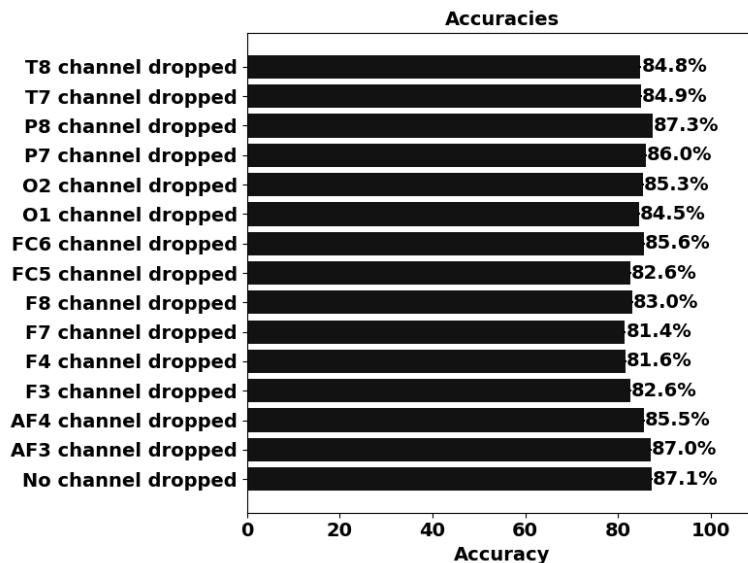


Figure 5.51: Accuracies for Guinea-Bissau dataset using RNN

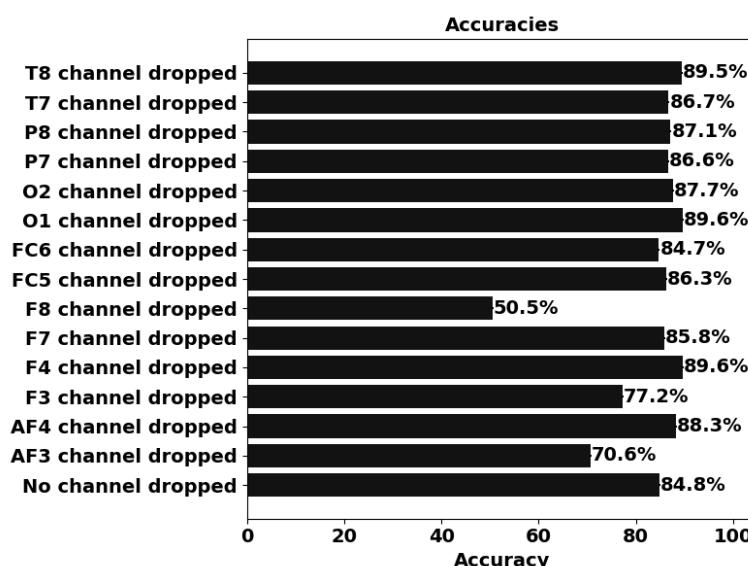


Figure 5.52: Accuracies for Nigeria dataset using RNN

Figure 5.53 and Figure 5.54 represent the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Gated Recurrent Neural Network.

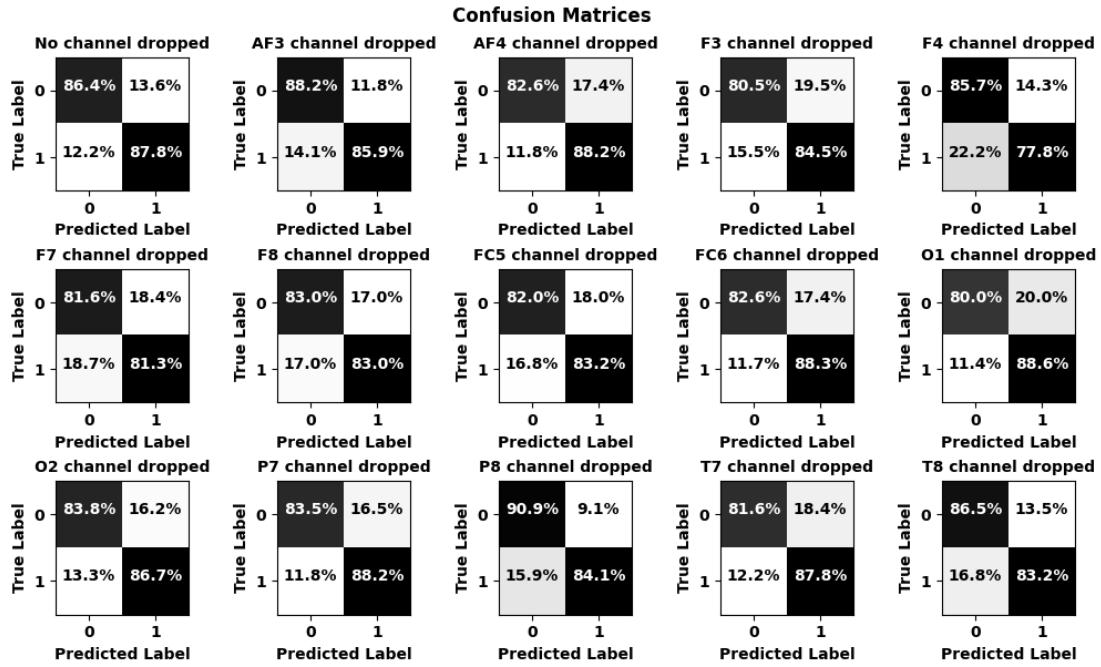


Figure 5.53: Confusion Matrices for Guinea-Bissau dataset using RNN

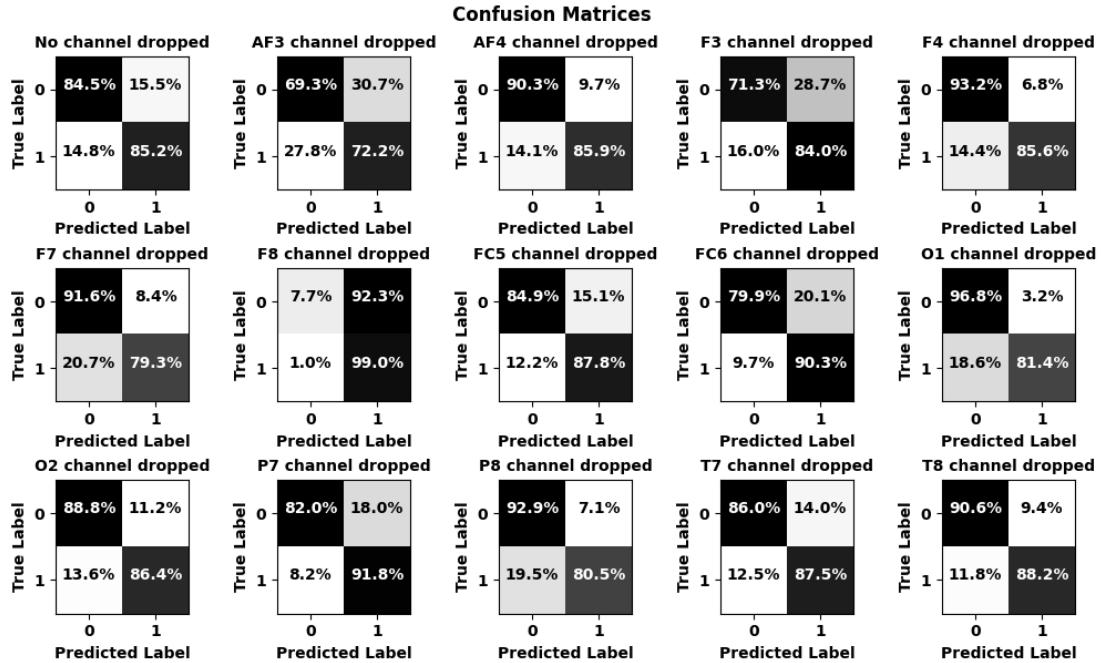


Figure 5.54: Confusion Matrices for Nigeria dataset using RNN

Figure 5.55 and Figure 5.56 demonstrate the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Gated Recurrent Neural Network.

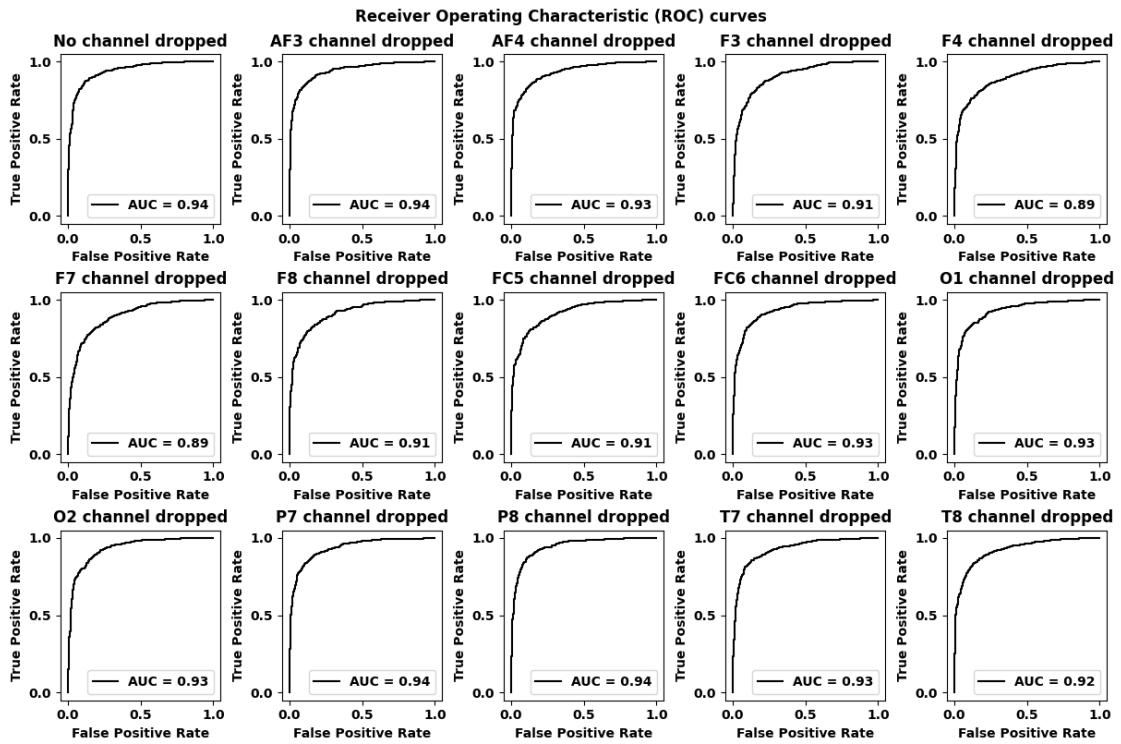


Figure 5.55: ROC curves for Guinea-Bissau dataset using RNN

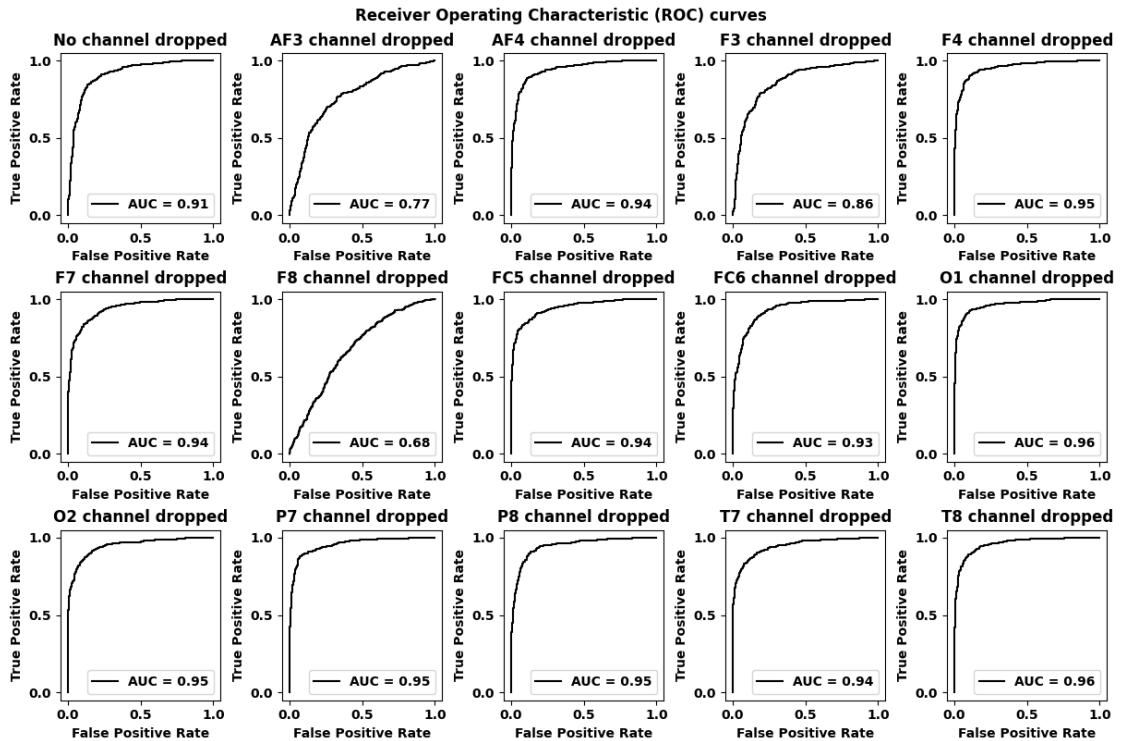


Figure 5.56: ROC curves for Nigeria dataset using RNN

5.2.10 Convolutional Gated Recurrent Neural Network (C-RNN):

Figure 5.57 and Figure 5.58 illustrate the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Gated Recurrent Neural Network.

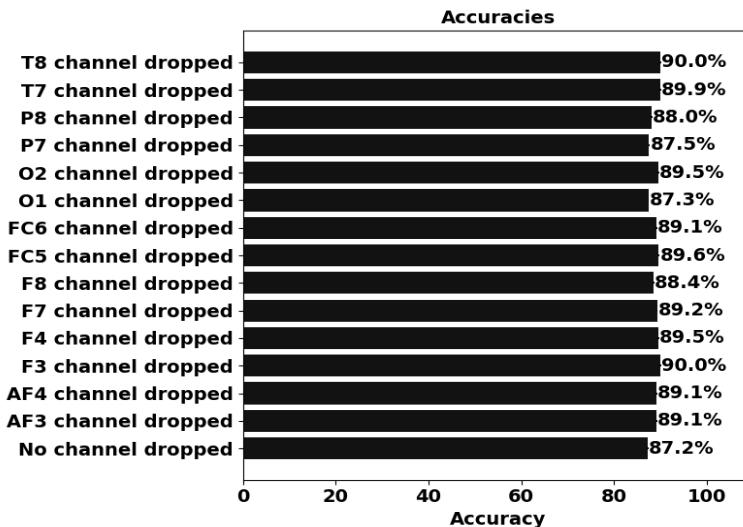


Figure 5.57: Accuracies for Guinea-Bissau dataset using C-RNN

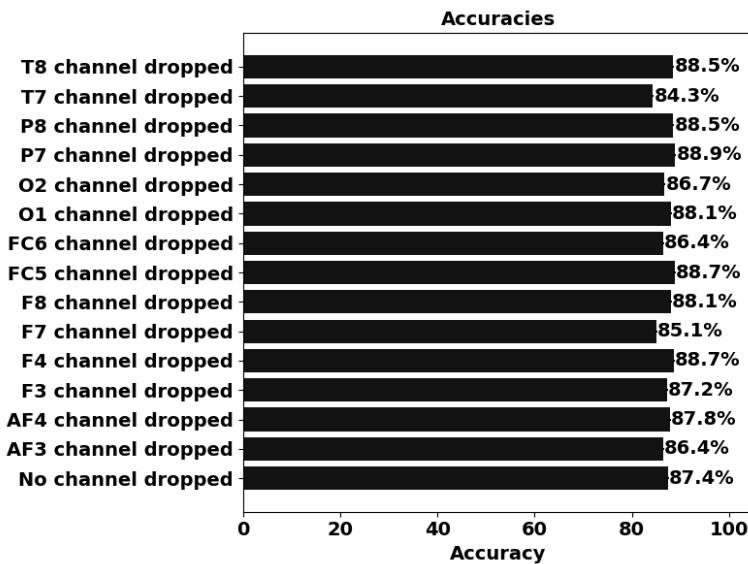


Figure 5.58: Accuracies for Nigeria dataset using C-RNN

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.59 and Figure 5.60 respectively using Convolutional Gated Recurrent Neural Network.

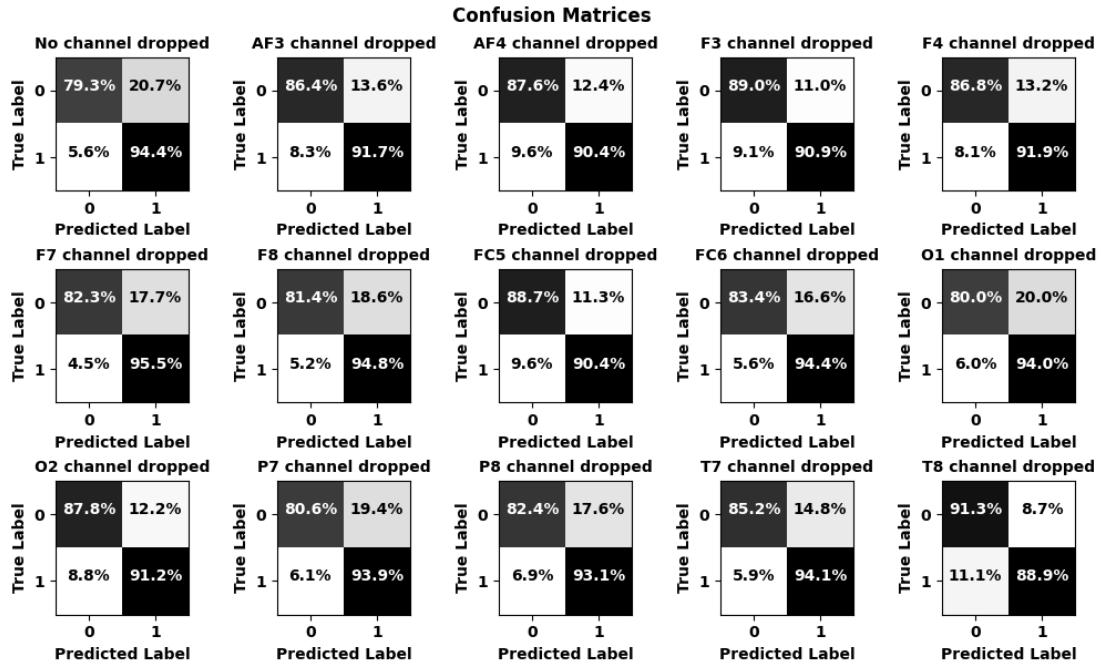


Figure 5.59: Confusion Matrices for Guinea-Bissau dataset using C-RNN

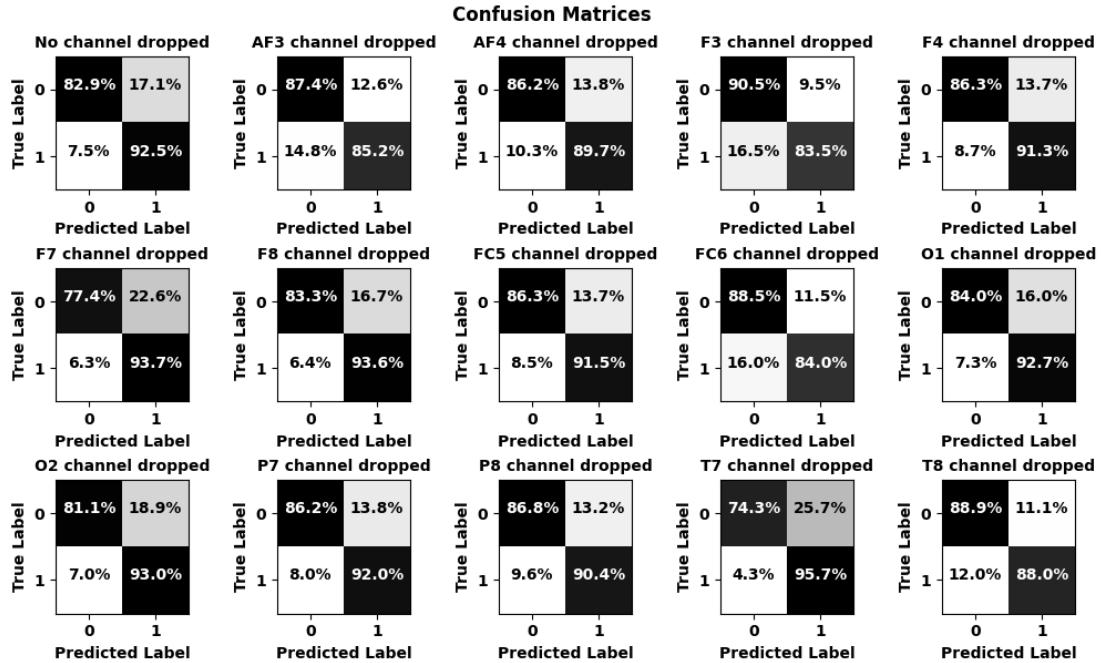


Figure 5.60: Confusion Matrices for Nigeria dataset using C-RNN

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.61 and Figure 5.62 respectively using Convolutional Gated Recurrent Neural Network.

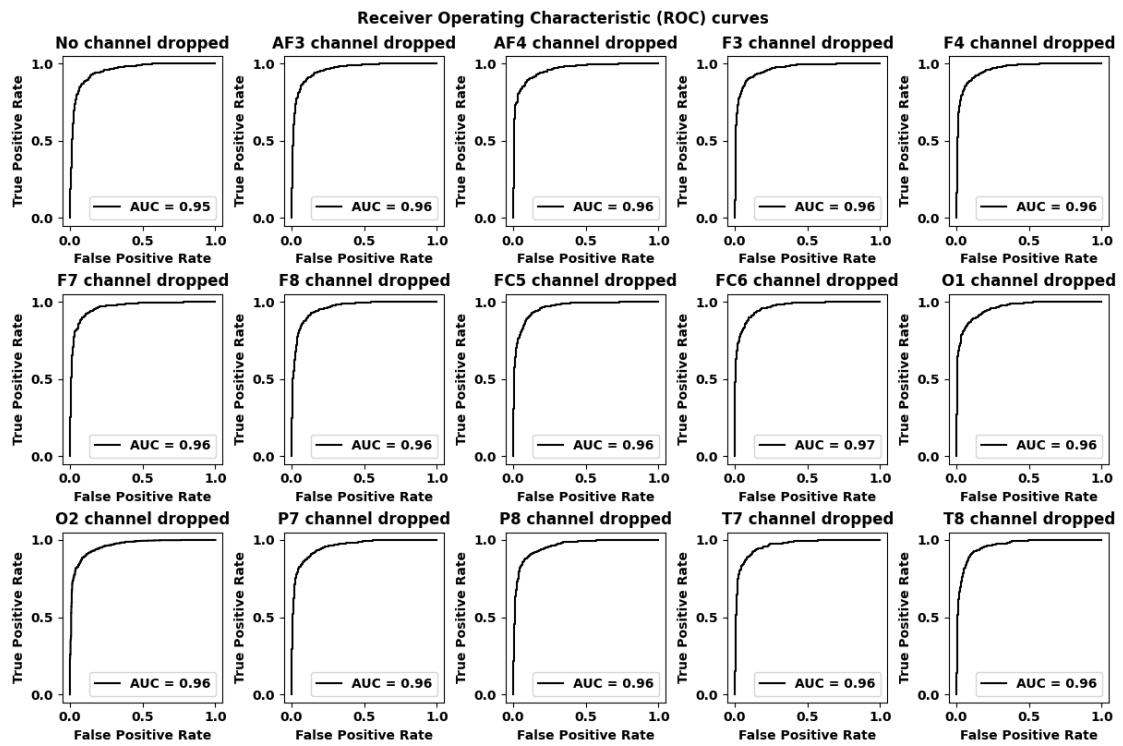


Figure 5.61: ROC curves for Guinea-Bissau dataset using C-RNN

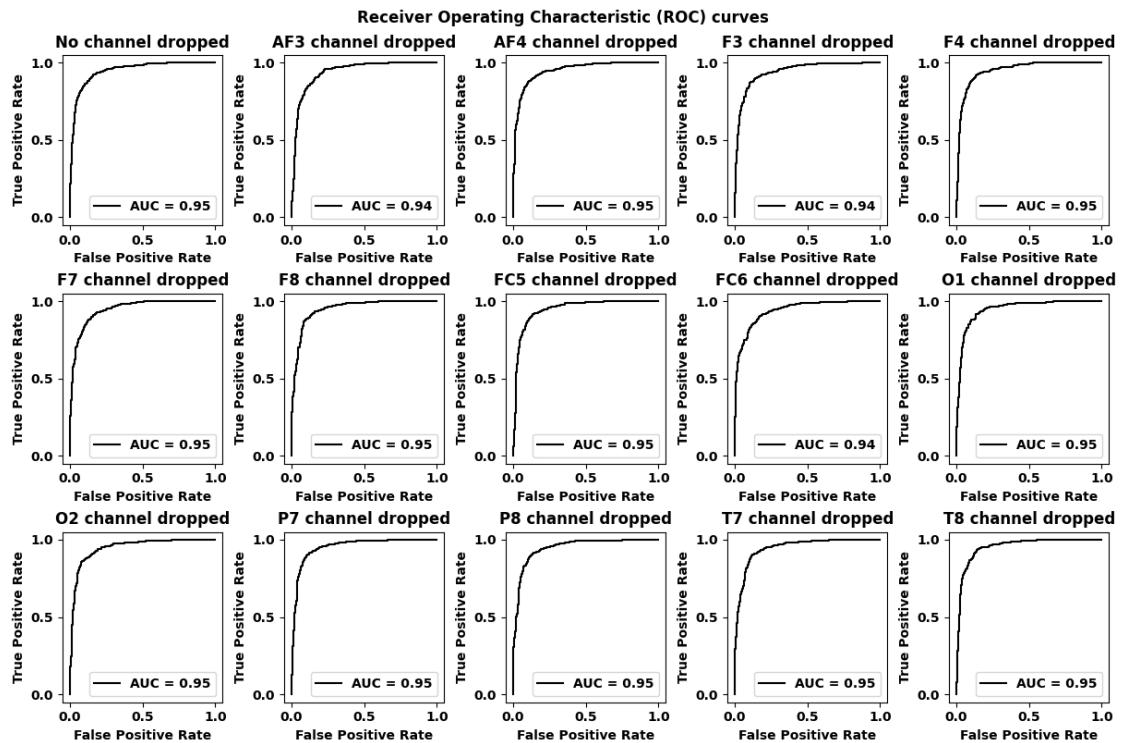


Figure 5.62: ROC curves for Nigeria dataset using C-RNN

5.2.11 Inception Convolutional Gated Recurrent Neural Network (IC-RNN):

Figure 5.63 and Figure 5.64 display the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Inception Convolutional Gated Recurrent Neural Network.

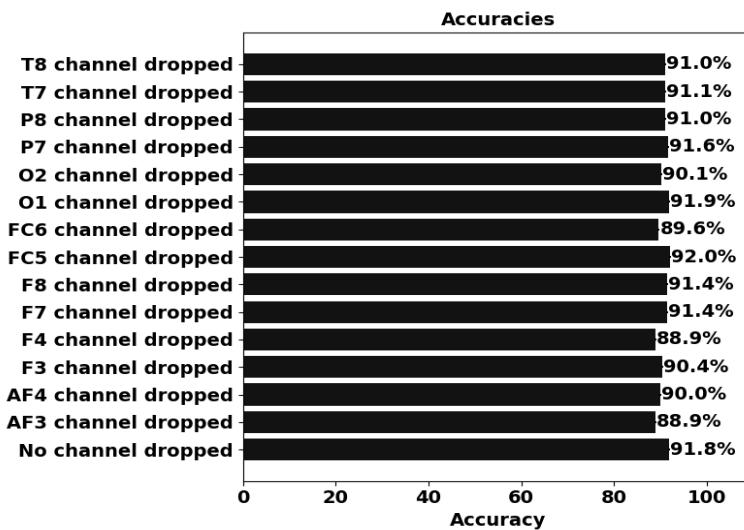


Figure 5.63: Accuracies for Guinea-Bissau dataset using IC-RNN

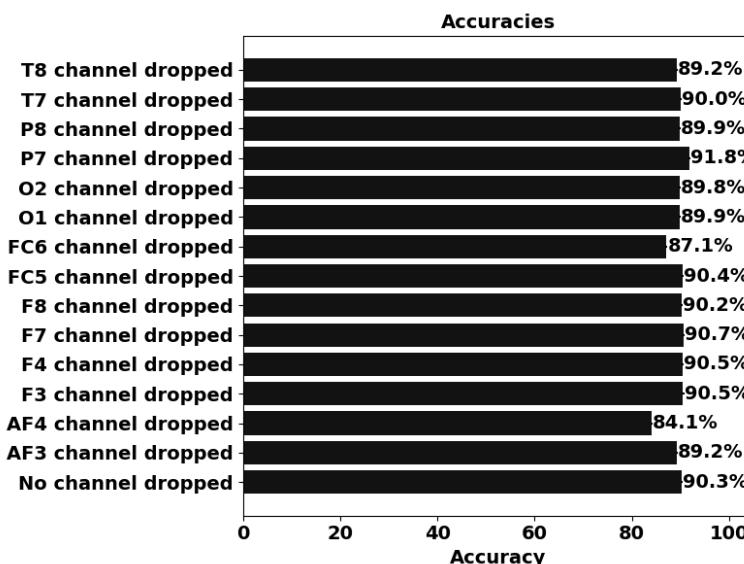


Figure 5.64: Accuracies for Nigeria dataset using IC-RNN

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.65 and Figure 5.66 respectively using Inception Convolutional Gated Recurrent Neural Network.

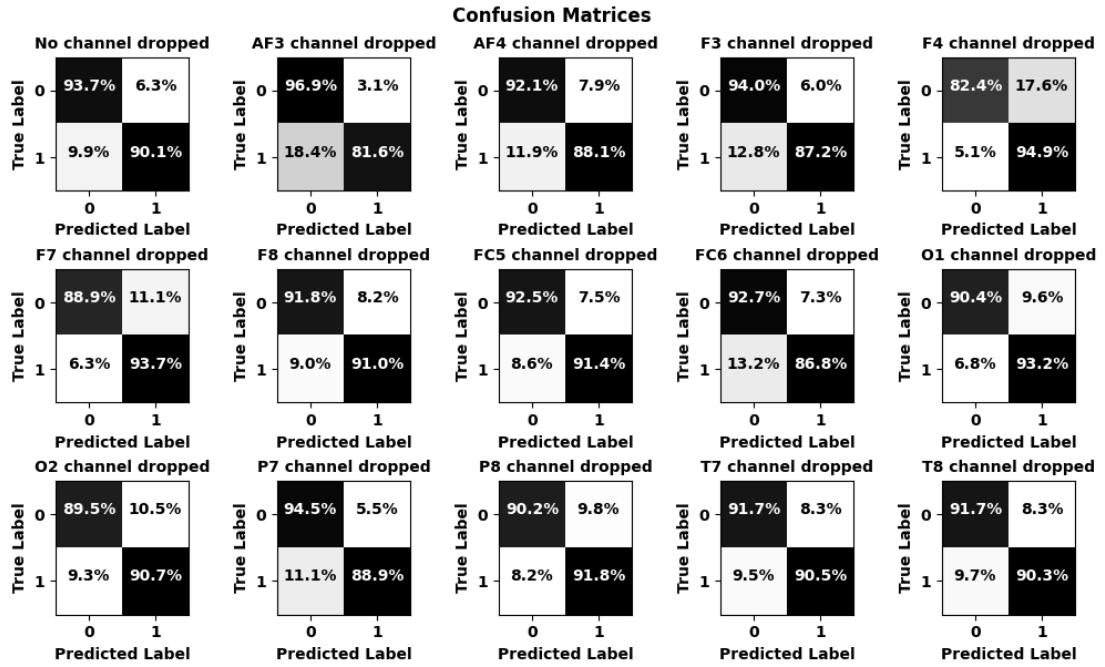


Figure 5.65: Confusion Matrices for Guinea-Bissau dataset using IC-RNN

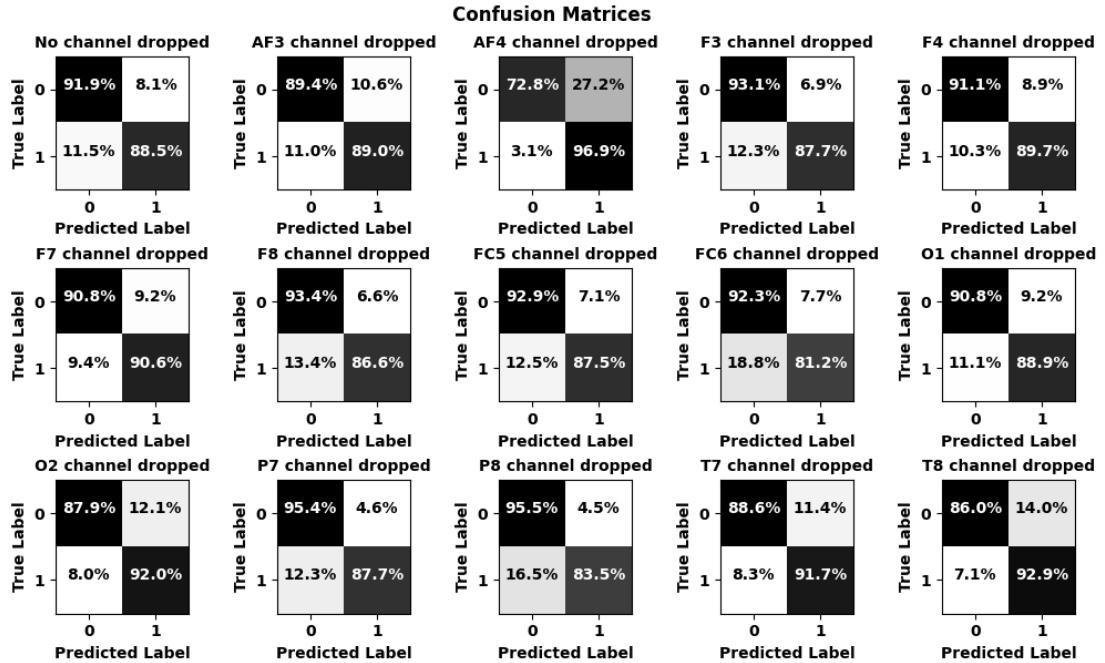


Figure 5.66: Confusion Matrices for Nigeria dataset using IC-RNN

Figure 5.67 and Figure 5.68 show the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Inception Convolutional Gated Recurrent Neural Network.

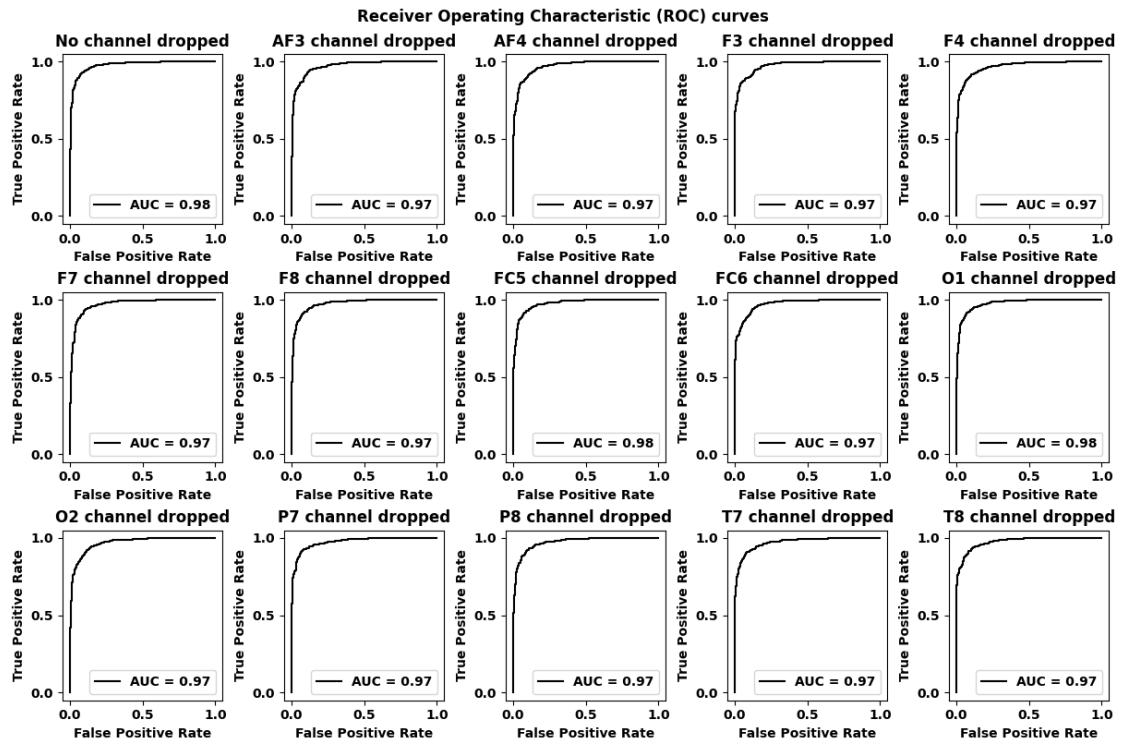


Figure 5.67: ROC curves for Guinea-Bissau dataset using IC-RNN

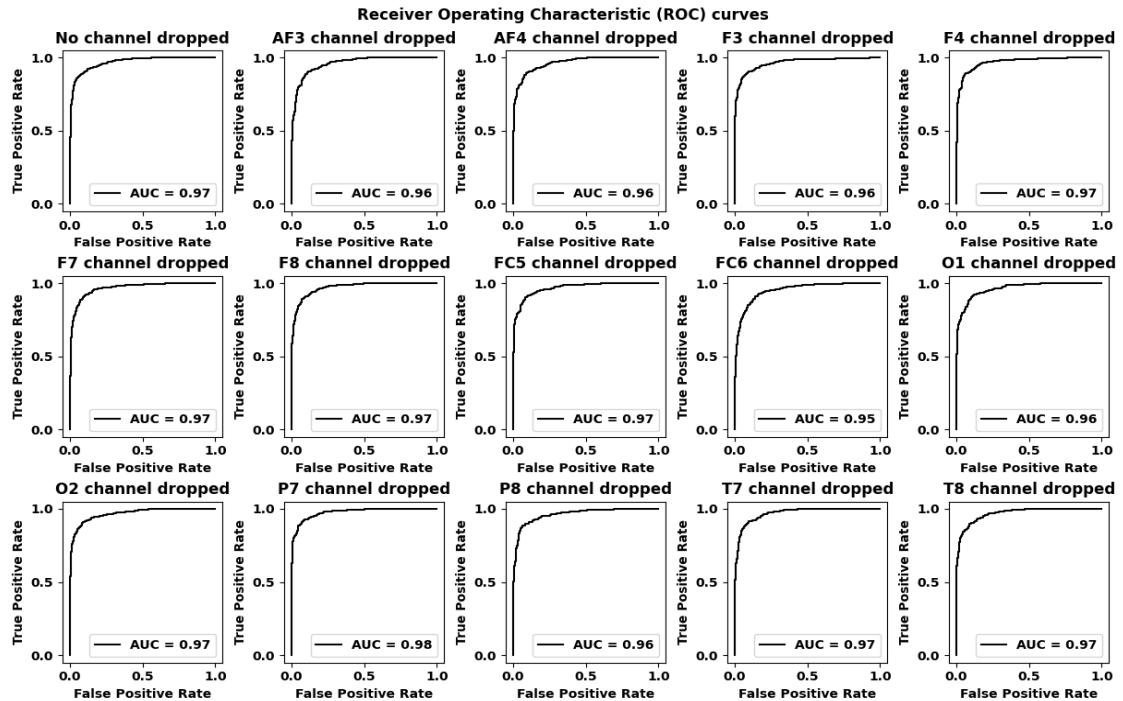


Figure 5.68: ROC curves for Nigeria dataset using IC-RNN

5.2.12 Convolutional Densely Connected Gated Recurrent Neural Network (C-DRNN):

Figure 5.69 and Figure 5.70 represent the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Densely Connected Gated Recurrent Neural Network.

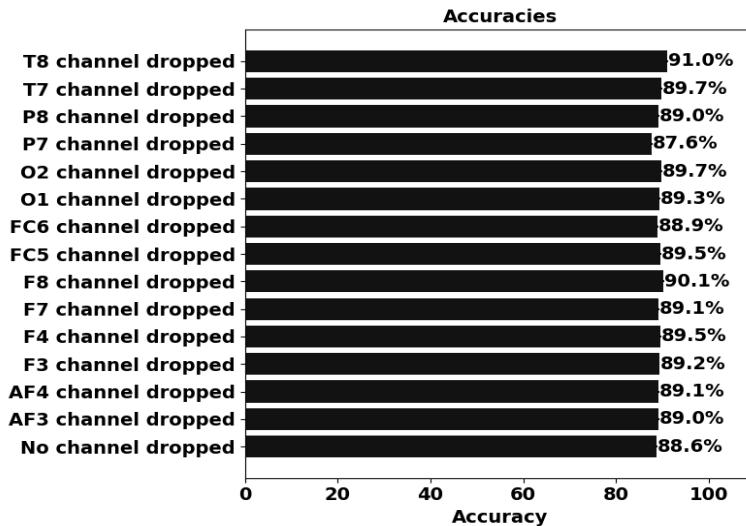


Figure 5.69: Accuracies for Guinea-Bissau dataset using C-DRNN

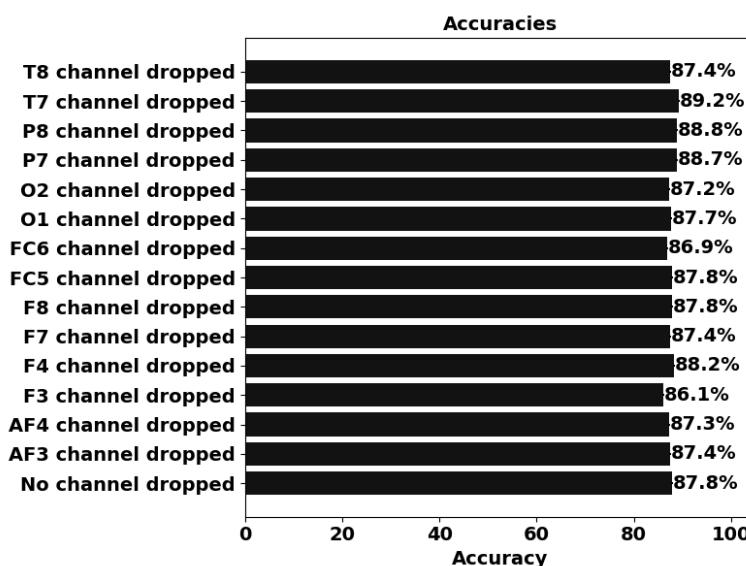


Figure 5.70: Accuracies for Nigeria dataset using C-DRNN

Figure 5.71 and Figure 5.72 demonstrate the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Densely Connected Gated Recurrent Neural Network.

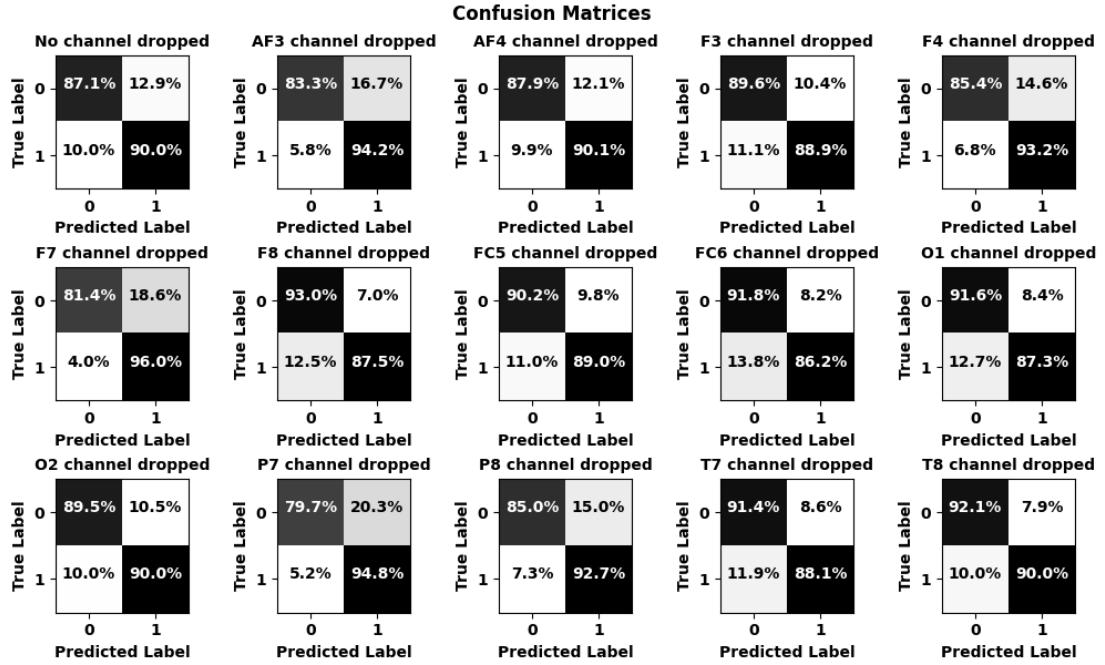


Figure 5.71: Confusion Matrices for Guinea-Bissau dataset using C-DRNN

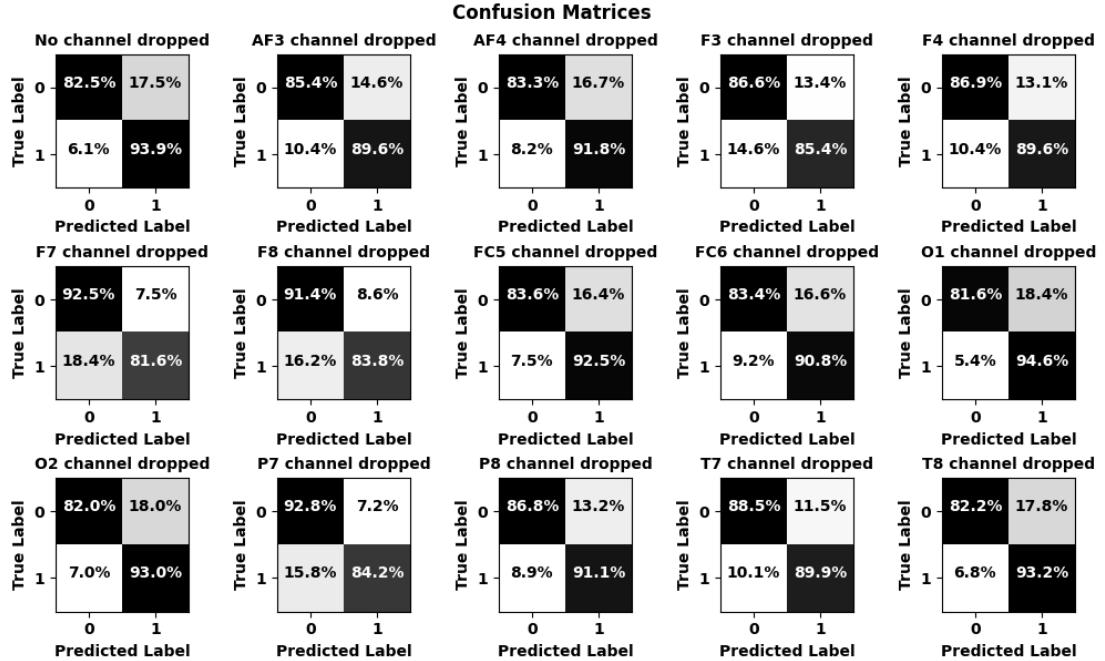


Figure 5.72: Confusion Matrices for Nigeria dataset using C-DRNN

Figure 5.73 and Figure 5.74 illustrate the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Densely Connected Gated Recurrent Neural Network.

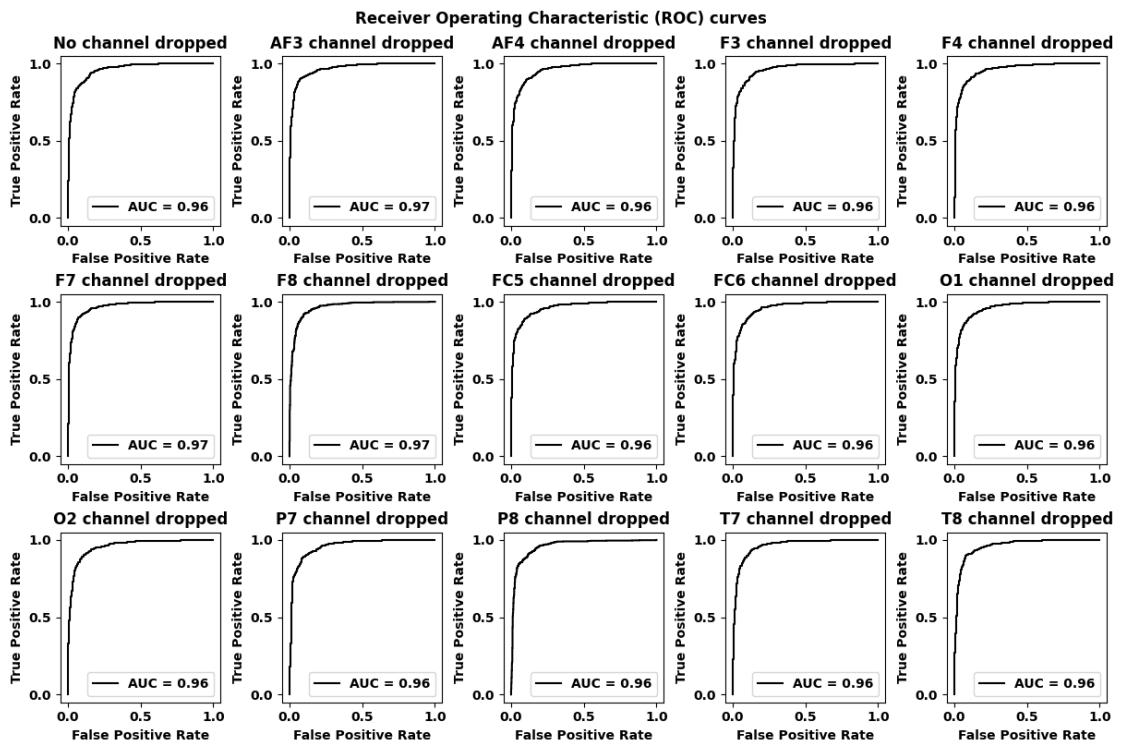


Figure 5.73: ROC curves for Guinea-Bissau dataset using C-DRNN

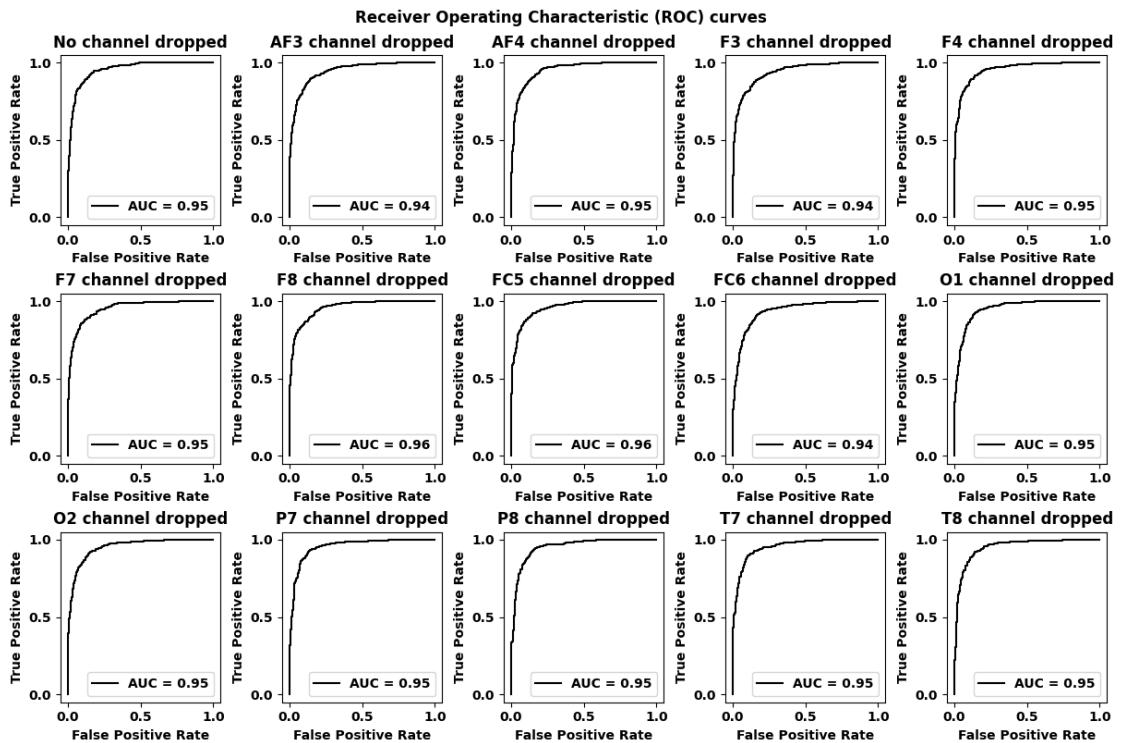


Figure 5.74: ROC curves for Nigeria dataset using C-DRNN

5.2.13 ChronoNet:

The accuracies for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.75 and Figure 5.76 respectively using ChronoNet.

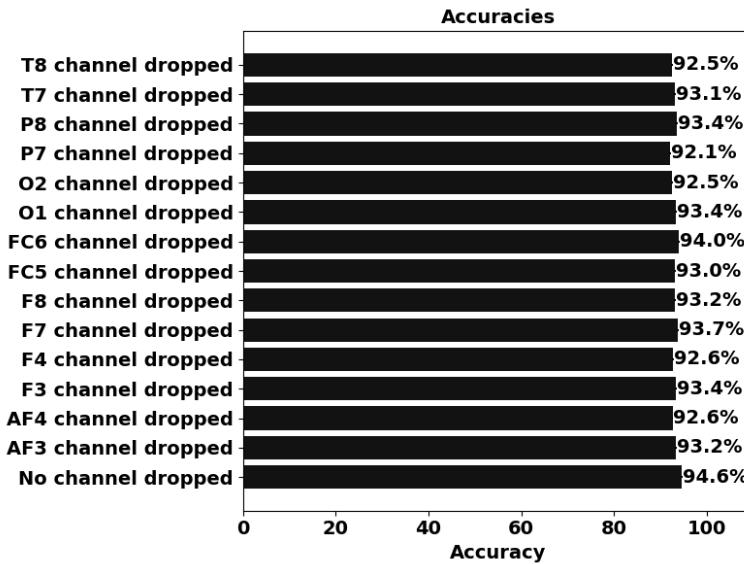


Figure 5.75: Accuracies for Guinea-Bissau dataset using ChronoNet

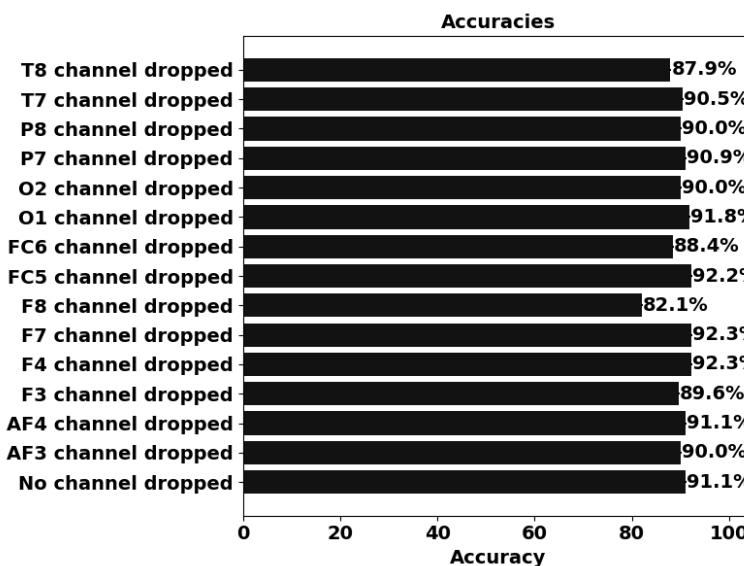


Figure 5.76: Accuracies for Nigeria dataset using ChronoNet

The confusion matrices for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.77 and Figure 5.78 respectively using ChronoNet.

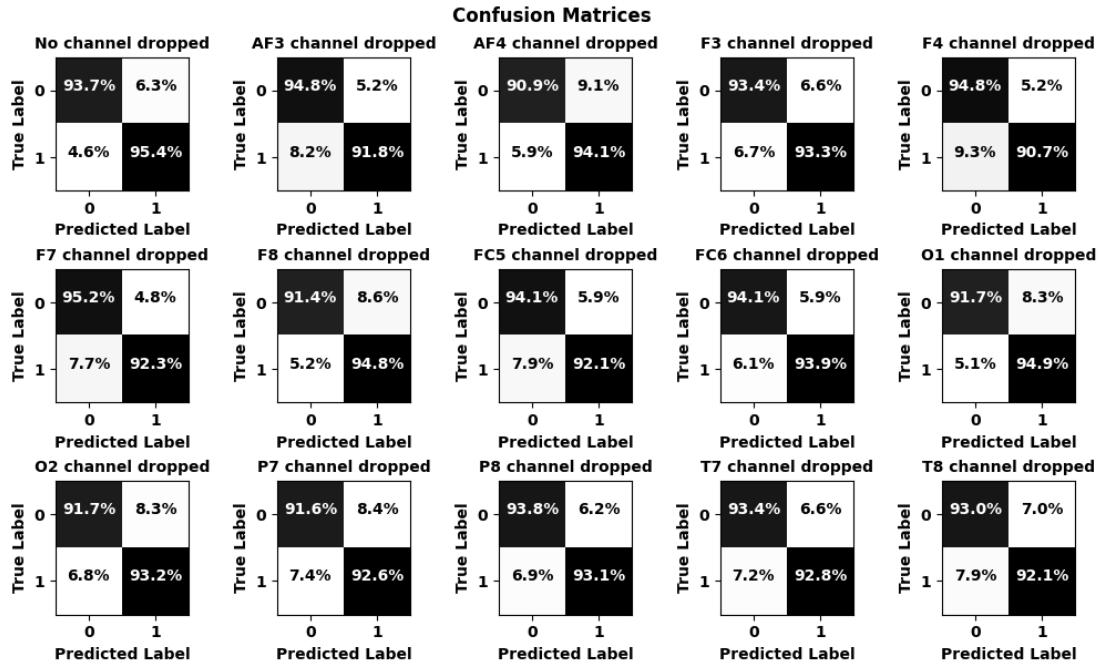


Figure 5.77: Confusion Matrices for Guinea-Bissau dataset using ChronoNet

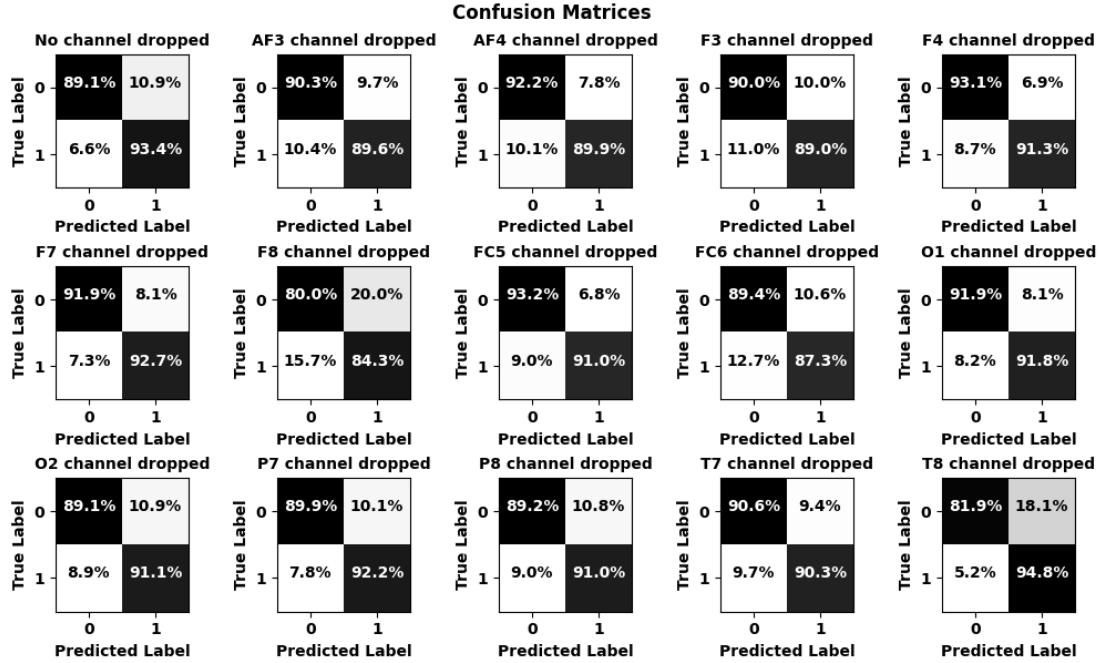


Figure 5.78: Confusion Matrices for Nigeria dataset using ChronoNet

Figure 5.79 and Figure 5.80 display the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using ChronoNet.

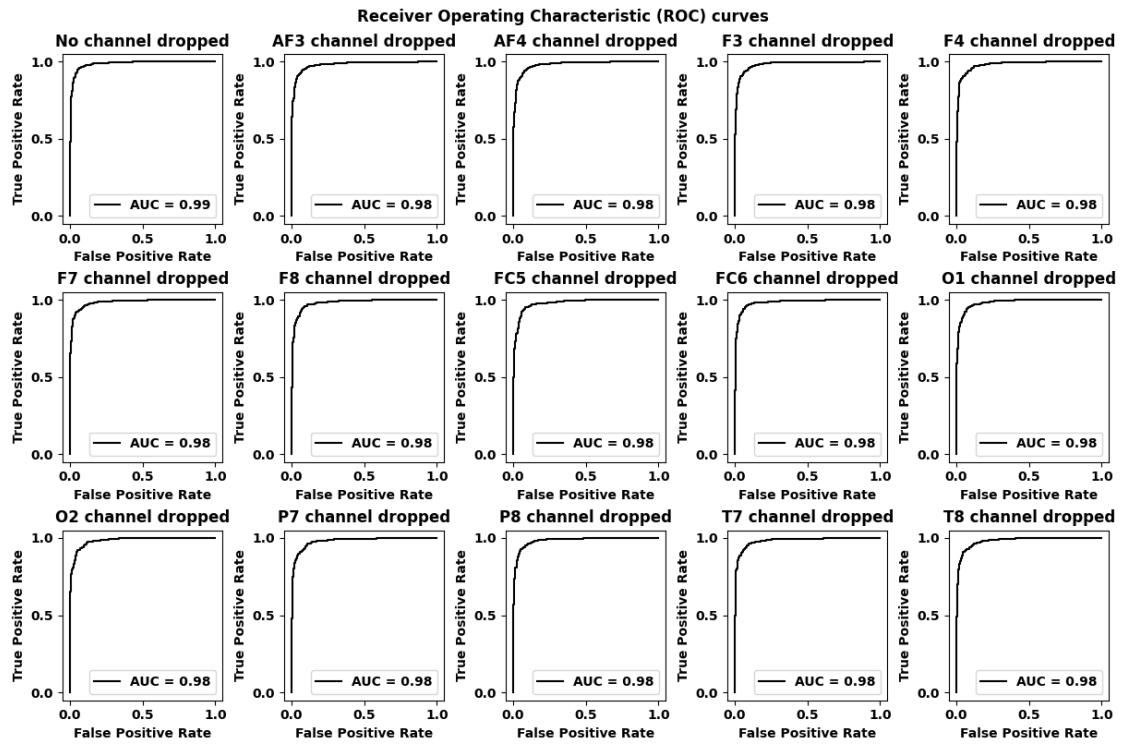


Figure 5.79: ROC curves for Guinea-Bissau dataset using ChronoNet

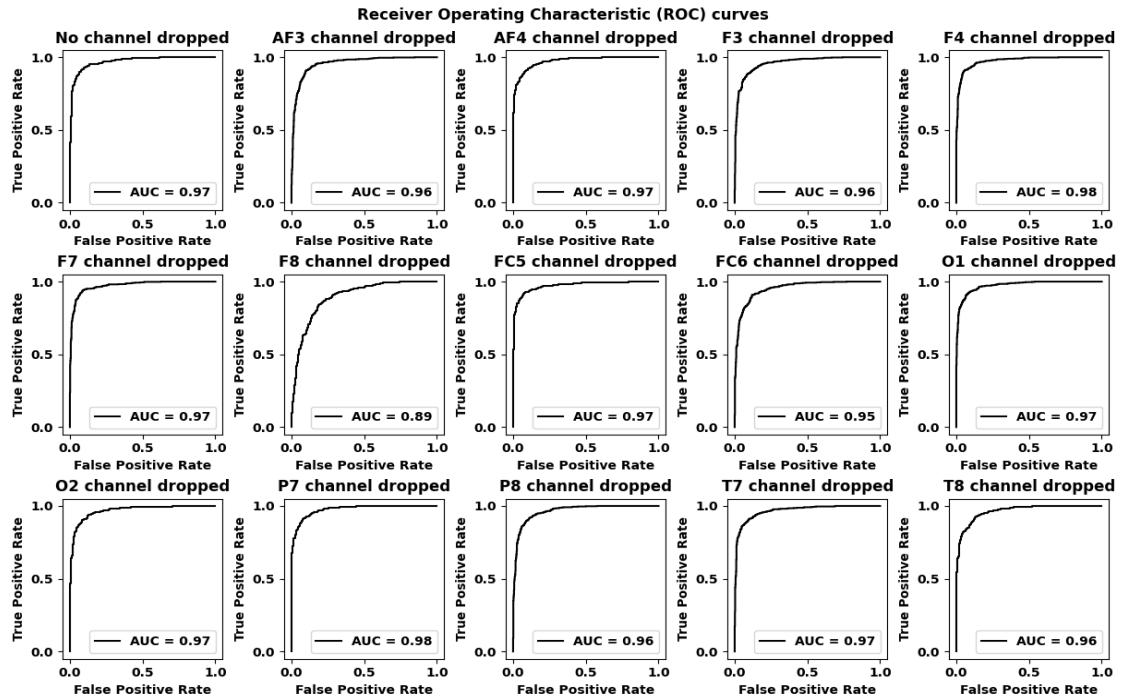


Figure 5.80: ROC curves for Nigeria dataset using ChronoNet

5.2.14 Convolutional Neural Network - Long Short Term Memory (CNN-LSTM):

The accuracies for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.81 and Figure 5.82 respectively using Convolutional Neural Network - Long Short Term Memory.

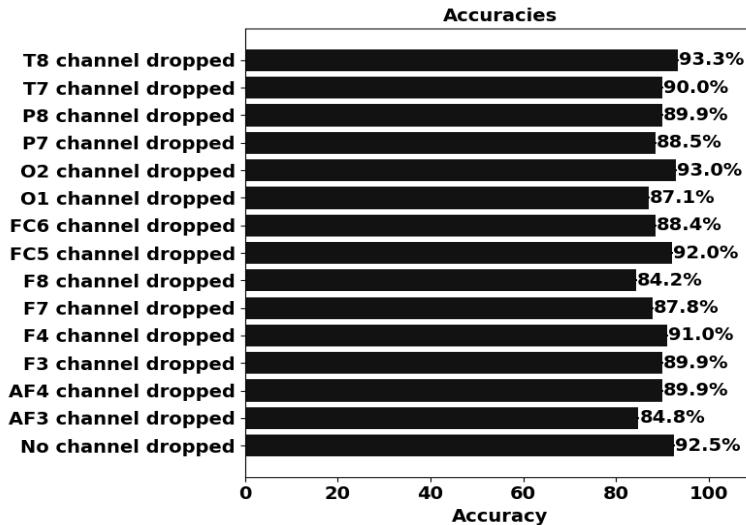


Figure 5.81: Accuracies for Guinea-Bissau dataset using CNN-LSTM

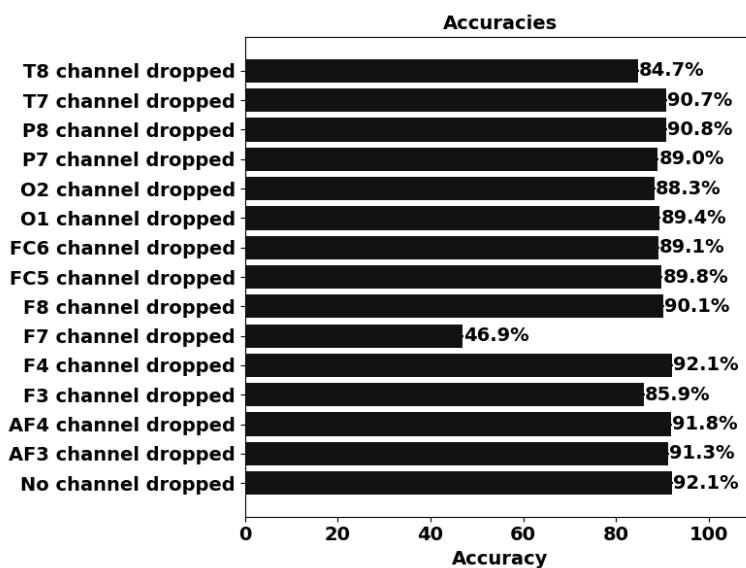


Figure 5.82: Accuracies for Nigeria dataset using CNN-LSTM

Figure 5.83 and Figure 5.84 show the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Neural Network - Long Short Term Memory.

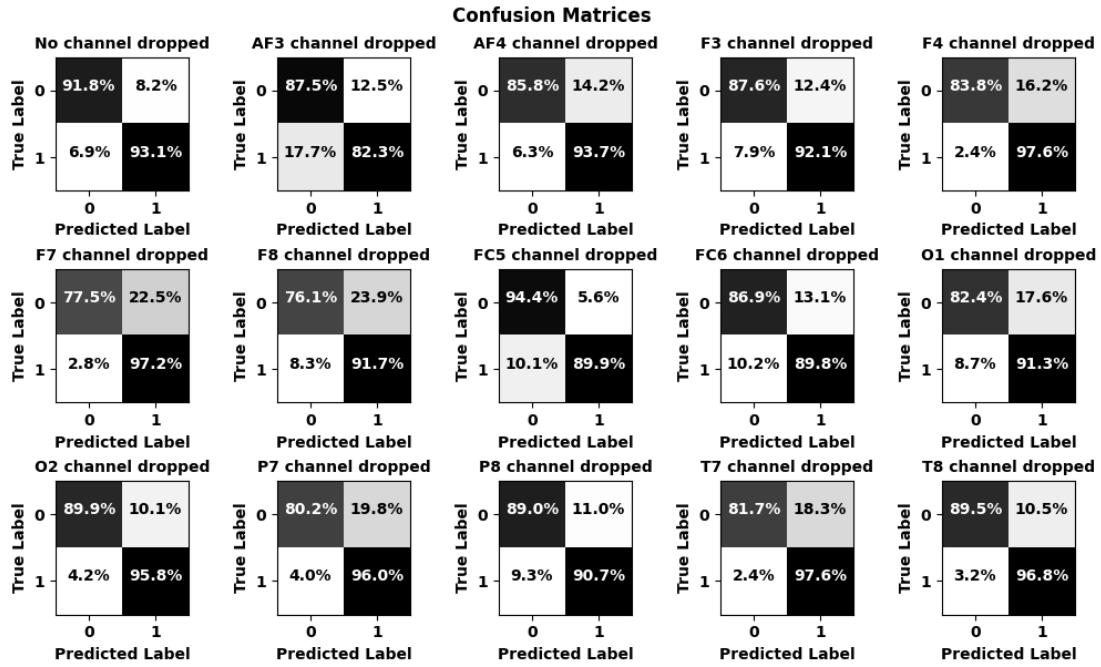


Figure 5.83: Confusion Matrices for Guinea-Bissau dataset using CNN-LSTM

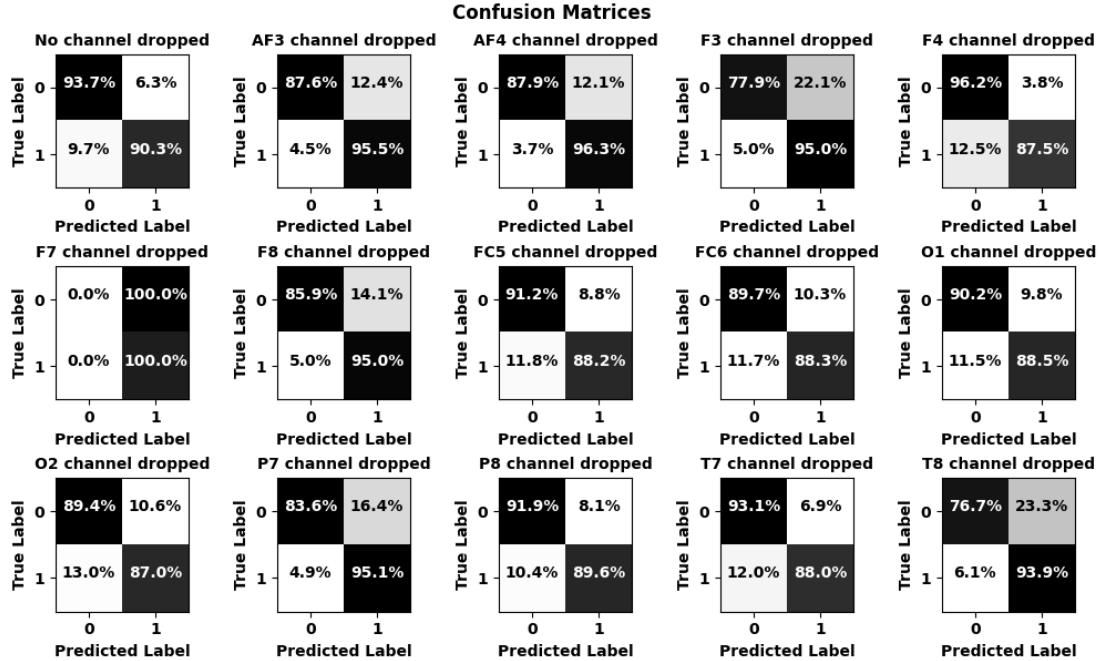


Figure 5.84: Confusion Matrices for Nigeria dataset using CNN-LSTM

Figure 5.85 and Figure 5.86 represent the receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset respectively using Convolutional Neural Network - Long Short Term Memory.

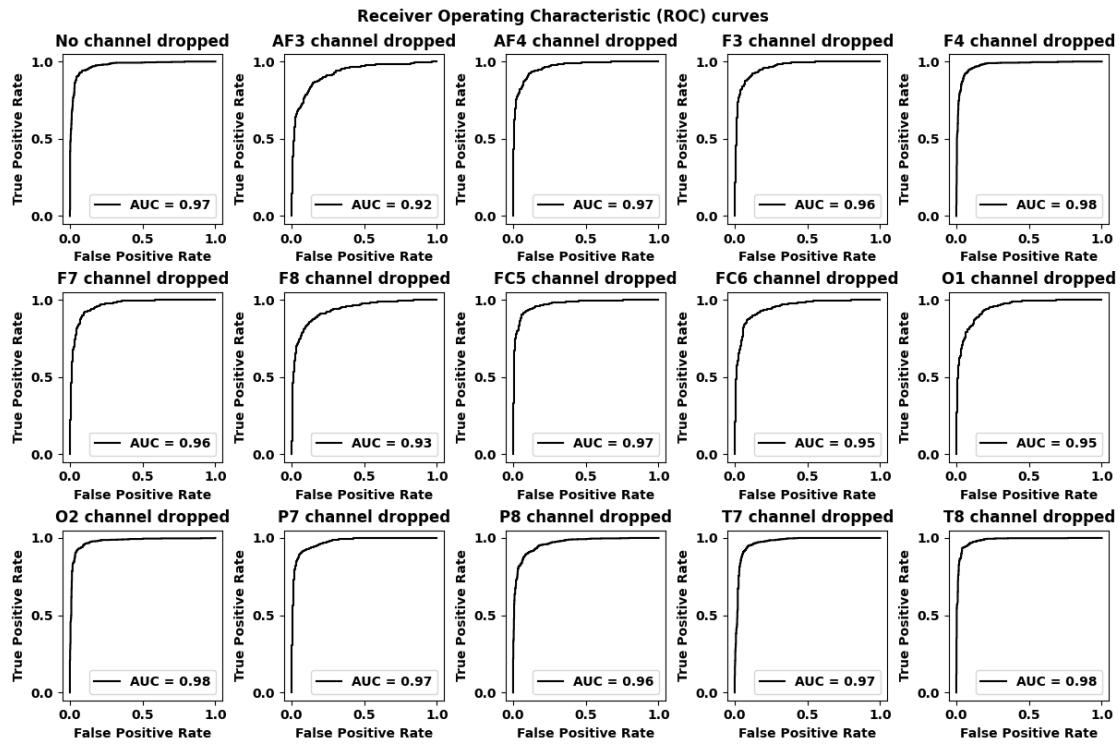


Figure 5.85: ROC curves for Guinea-Bissau dataset using CNN-LSTM

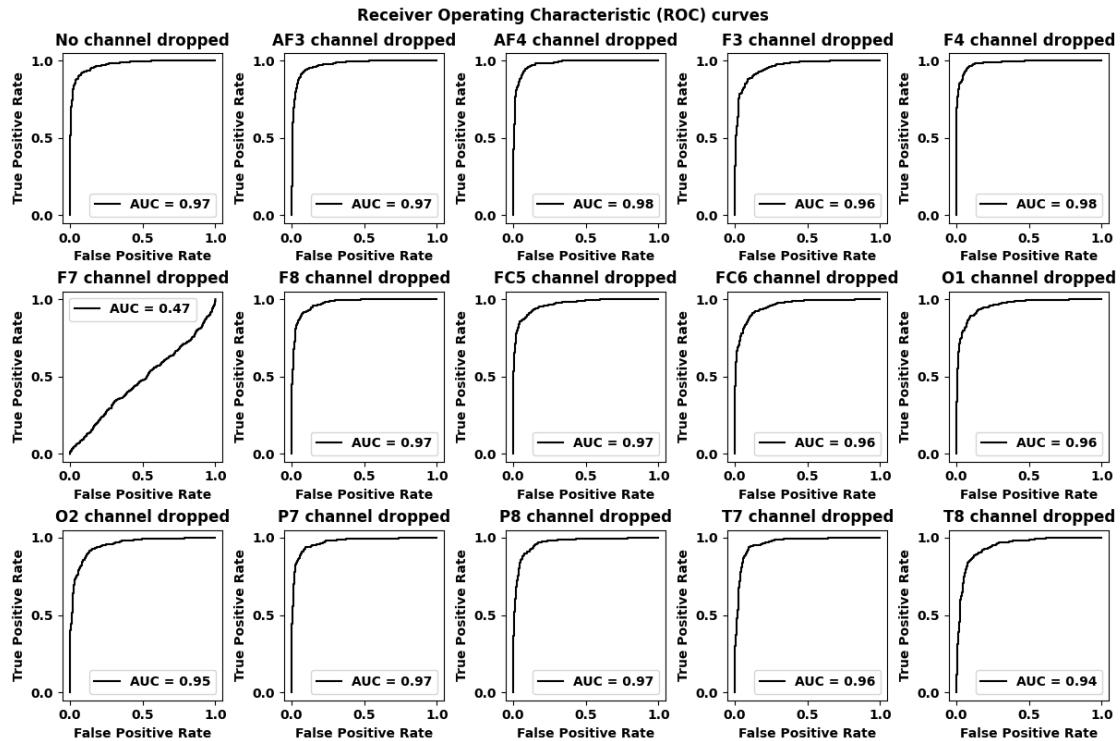


Figure 5.86: ROC curves for Nigeria dataset using CNN-LSTM

5.2.15 Proposed Model 1 (ChronoNet-M):

Figure 5.87 and Figure 5.88 demonstrate the accuracies for Guinea-Bissau dataset and Nigeria dataset respectively using ChronoNet-M.

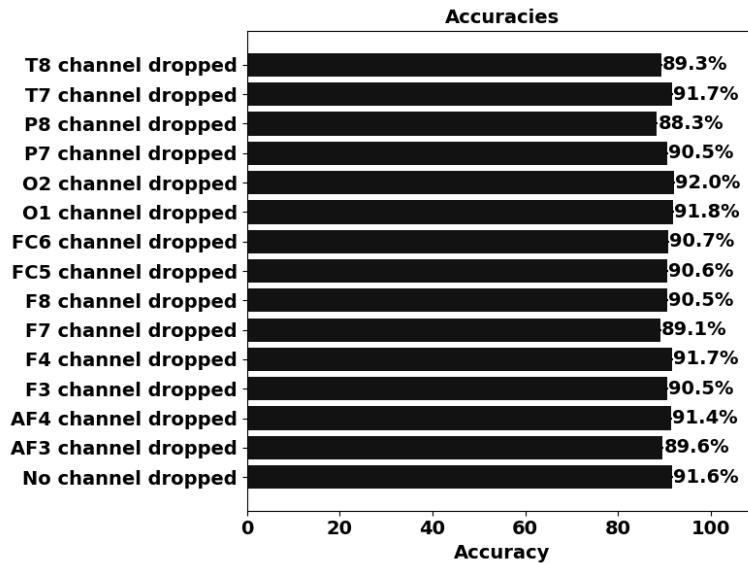


Figure 5.87: Accuracies for Guinea-Bissau dataset using ChronoNet-M

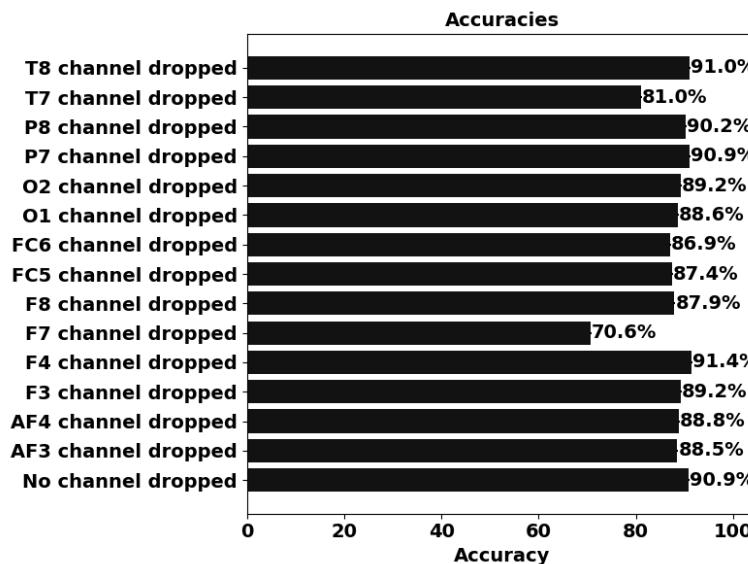


Figure 5.88: Accuracies for Nigeria dataset using ChronoNet-M

Figure 5.89 and Figure 5.90 illustrate the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using ChronoNet-M.

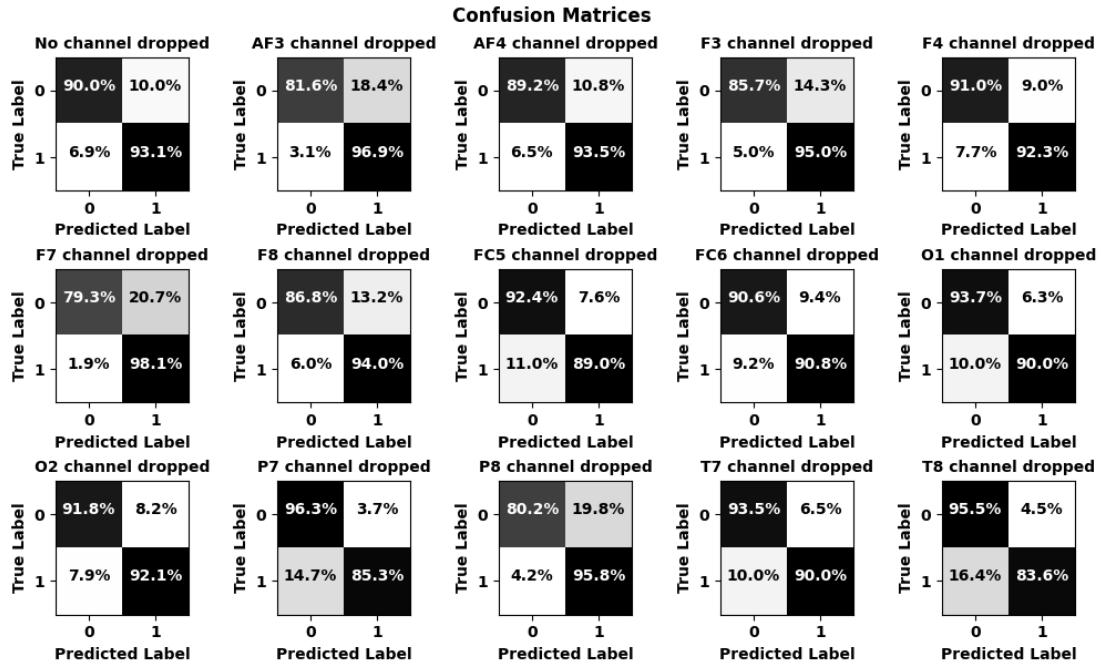


Figure 5.89: Confusion Matrices for Guinea-Bissau dataset using ChronoNet-M

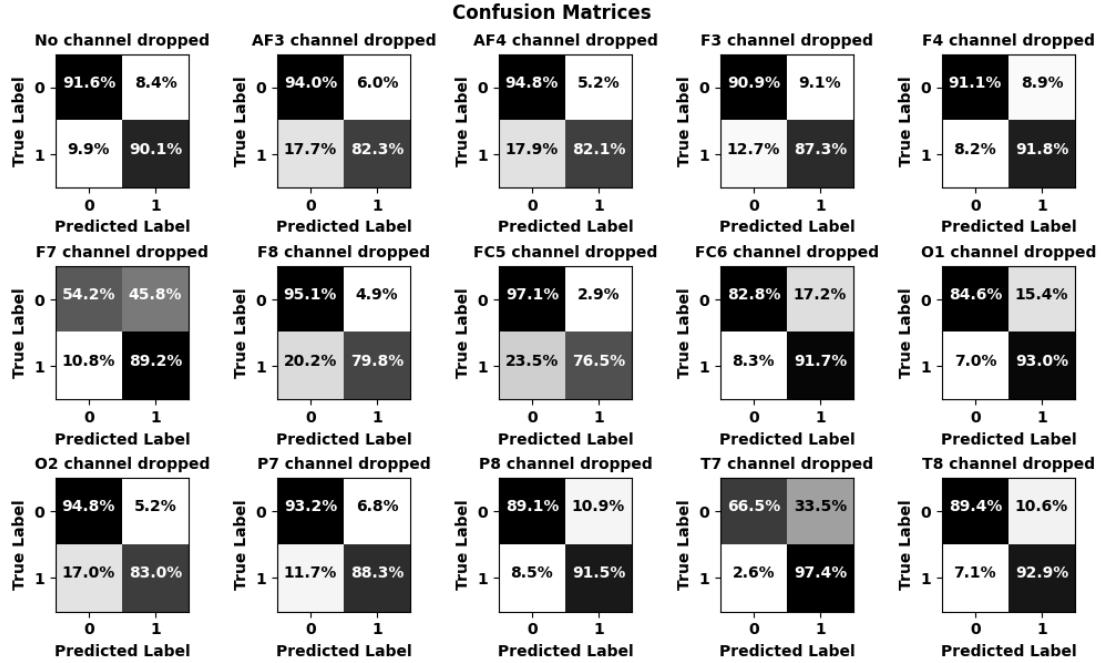


Figure 5.90: Confusion Matrices for Nigeria dataset using ChronoNet-M

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset can be observed in Figure 5.91 and Figure 5.92 respectively using ChronoNet-M.

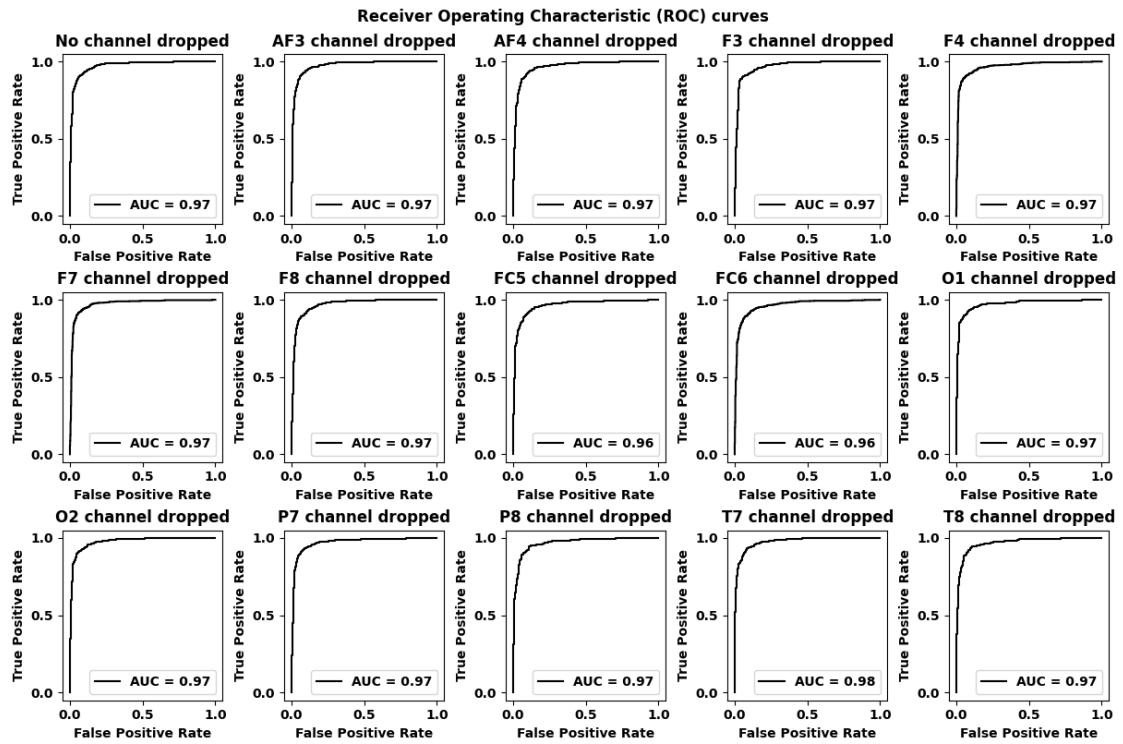


Figure 5.91: ROC curves for Guinea-Bissau dataset using ChronoNet-M

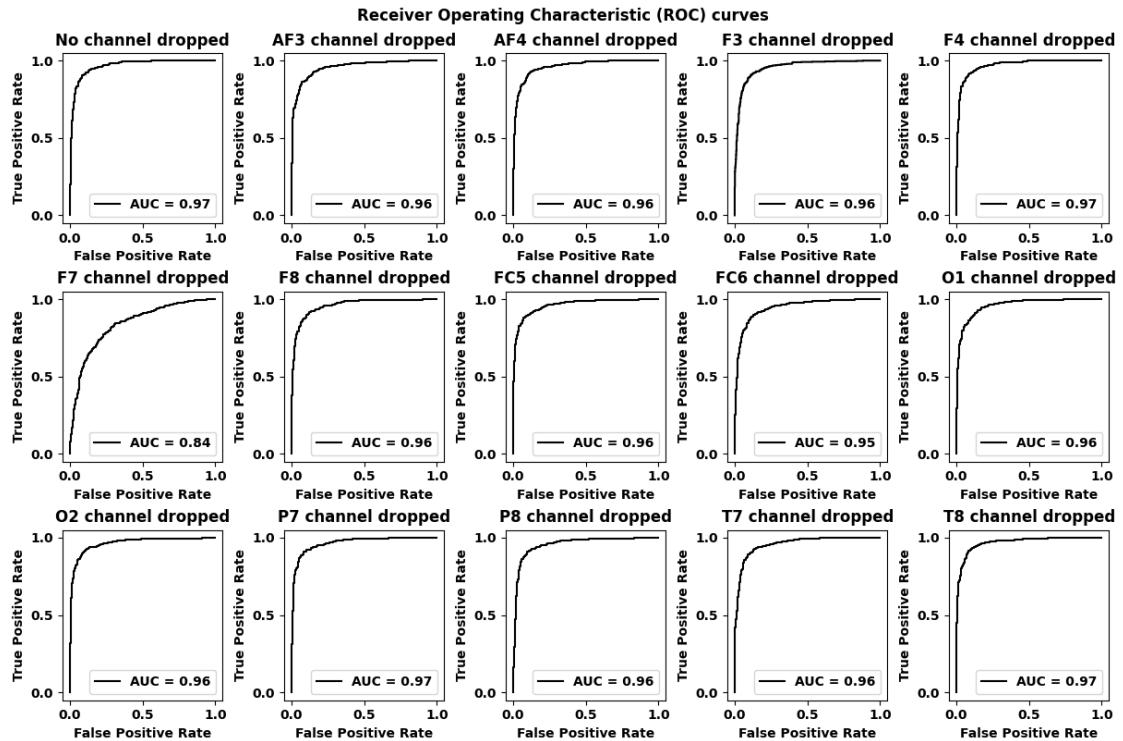


Figure 5.92: ROC curves for Nigeria dataset using ChronoNet-M

5.2.16 Proposed Model (CNN-LSTM-M):

The accuracies for Guinea-Bissau dataset and Nigeria dataset are depicted in Figure 5.93 and Figure 5.94 respectively using CNN-LSTM-M.

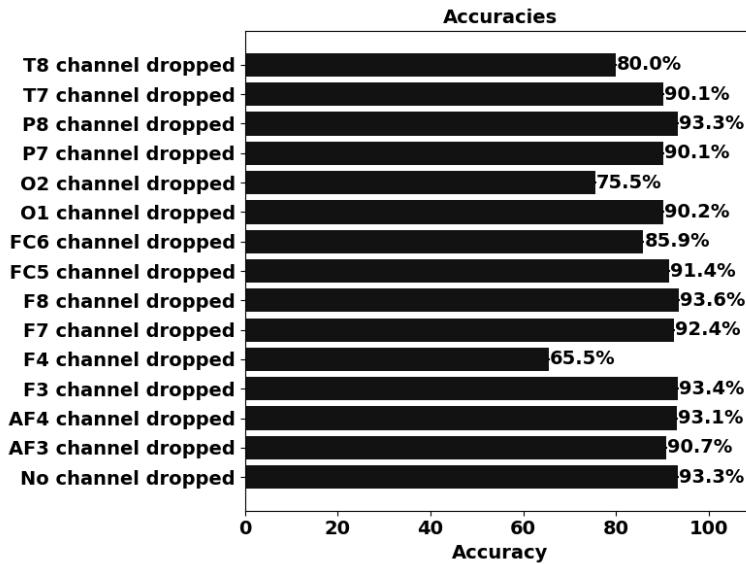


Figure 5.93: Accuracies for Guinea-Bissau dataset using mod CNN-LSTM-M

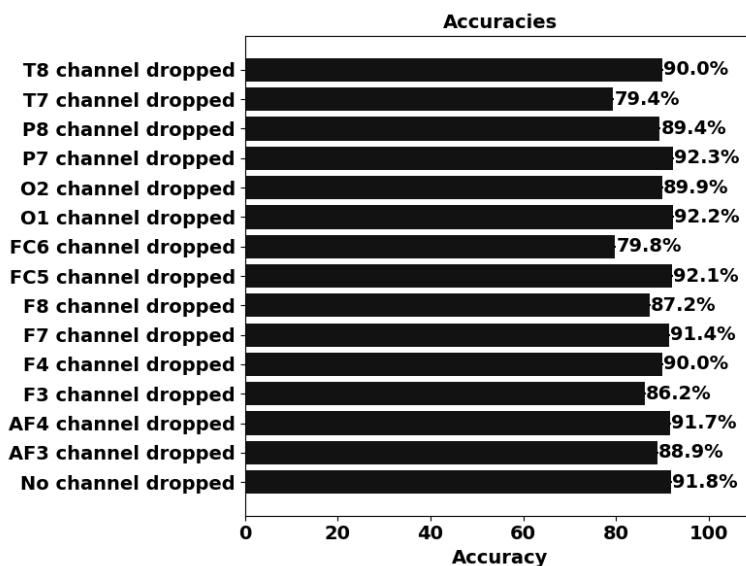


Figure 5.94: Accuracies for Nigeria dataset using CNN-LSTM-M

Figure 5.95 and Figure 5.96 display the confusion matrices for Guinea-Bissau dataset and Nigeria dataset respectively using CNN-LSTM-M.

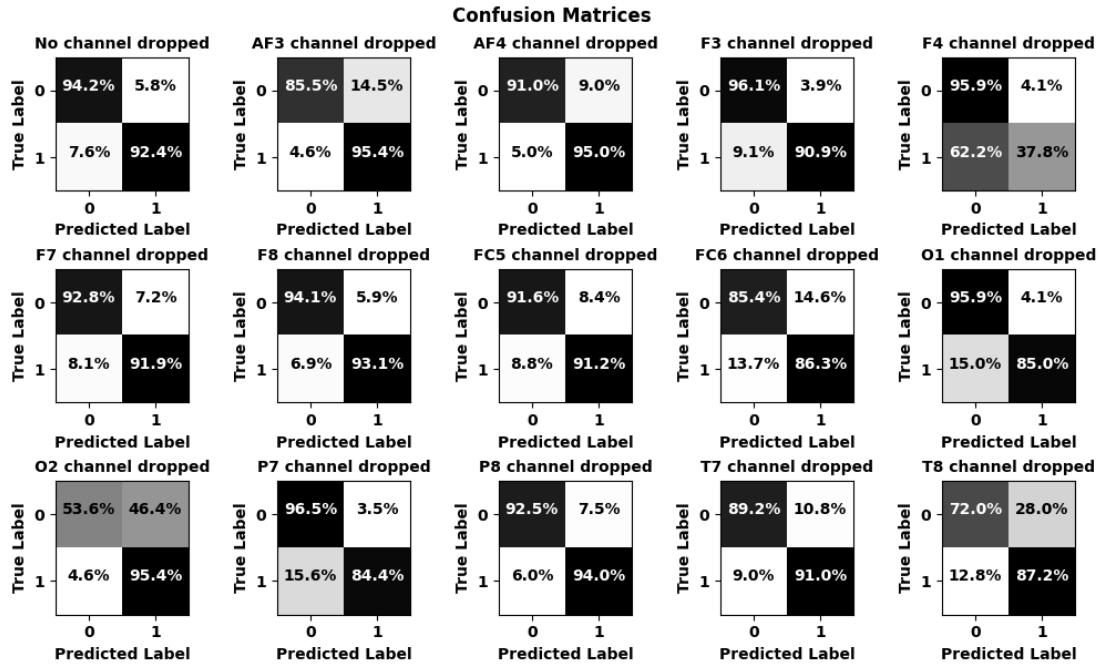


Figure 5.95: Confusion Matrices for Guinea-Bissau dataset using CNN-LSTM-M

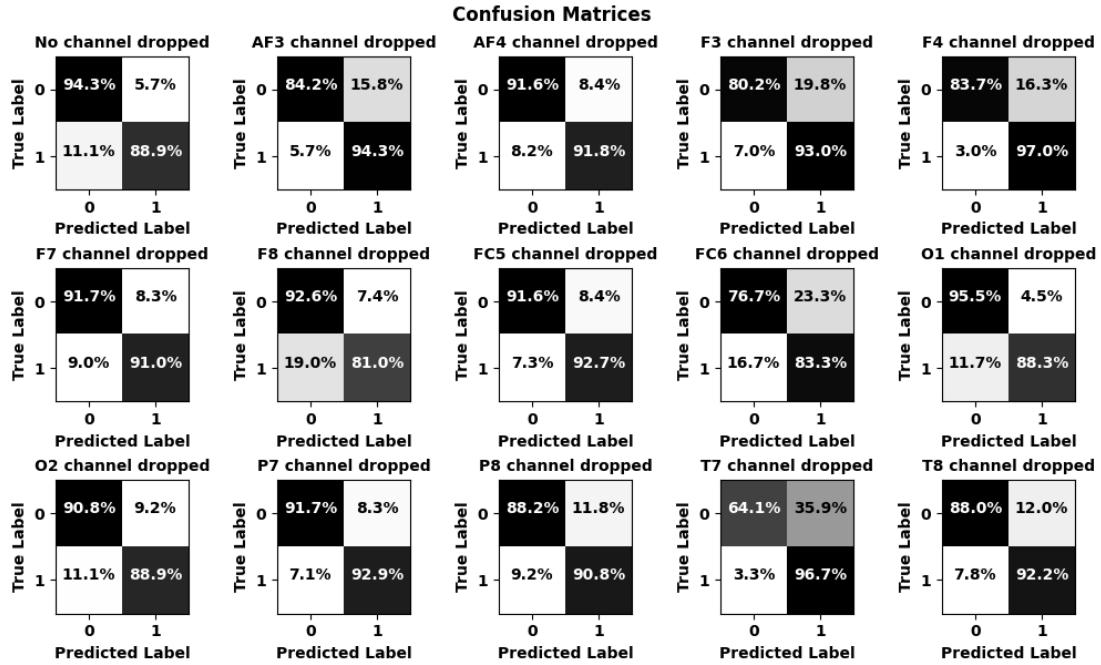


Figure 5.96: Confusion Matrices for Nigeria dataset using CNN-LSTM-M

The receiver output characteristic curves for Guinea-Bissau dataset and Nigeria dataset have been shown in Figure 5.97 and Figure 5.98 respectively using CNN-LSTM-M.

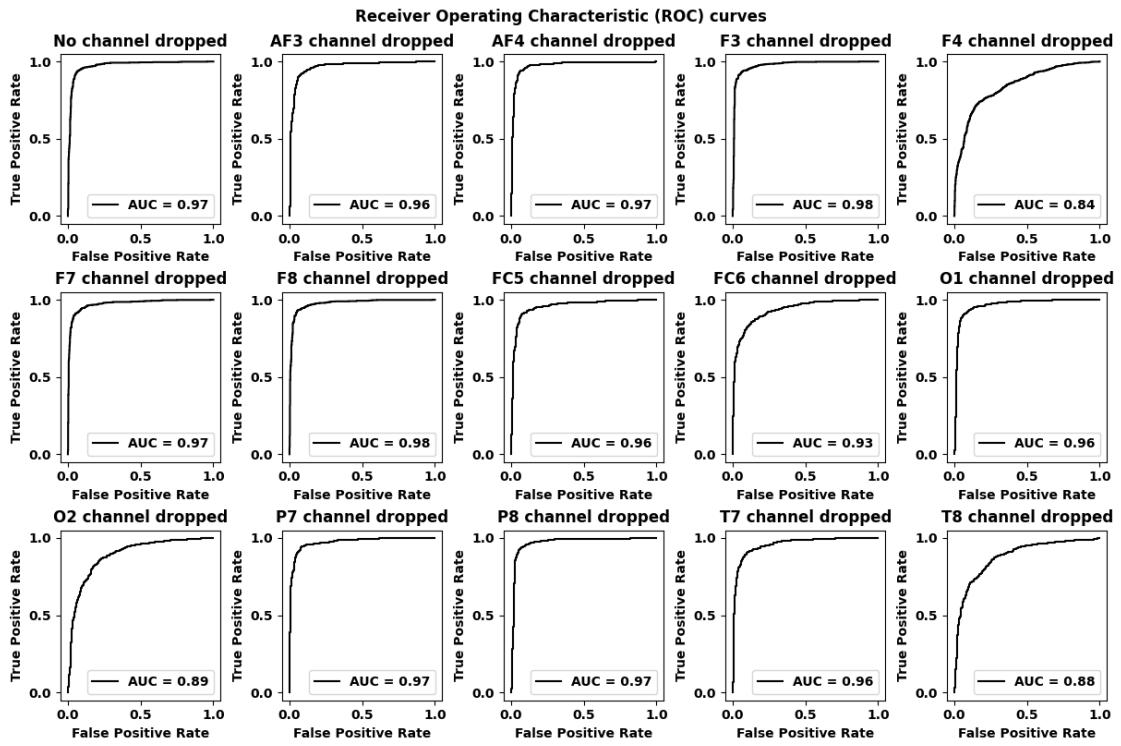


Figure 5.97: ROC curves for Guinea-Bissau dataset using CNN-LSTM-M

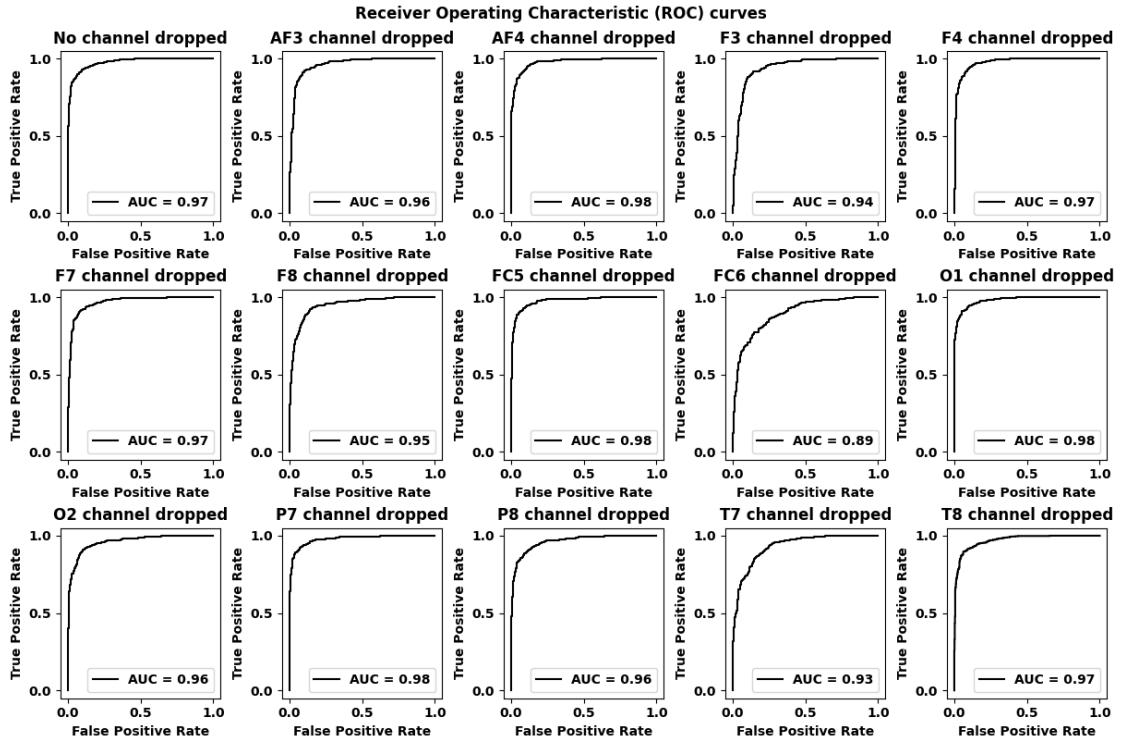


Figure 5.98: ROC curves for Nigeria dataset using CNN-LSTM-M

5.3 Effect of Multiple Channel Exclusion

According to the results in the previous section, ChronoNet has been found the best-performing model. In this section 2, 4, 6, 8, 10, and 12 channels have been dropped for the ChronoNet classifier. The objective was to identify how adversely the reduction of channel actually affects the accuracy of the classifiers.

5.3.1 Performance with Dropping Random Channels

5.3.1.1 2 Channels dropped

One pair of channels have been dropped from seven pairs at a time which make total seven combinations. The accuracies with Guinea-Bissau dataset for all seven combinations of 12 channels have been shown in Figure 5.99.

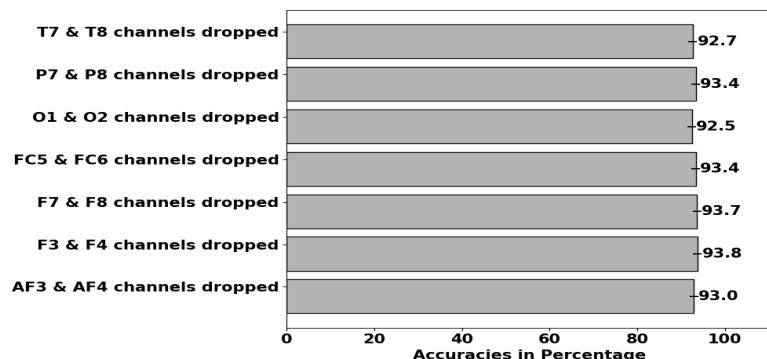


Figure 5.99: Accuracies for Guinea-Bissau dataset using ChronoNet for 12 channels

One pair of channels have been dropped from seven pairs at a time which make total seven combinations. The accuracies with Nigeria dataset for all seven combinations of 12 channels have been shown in Figure 5.100.

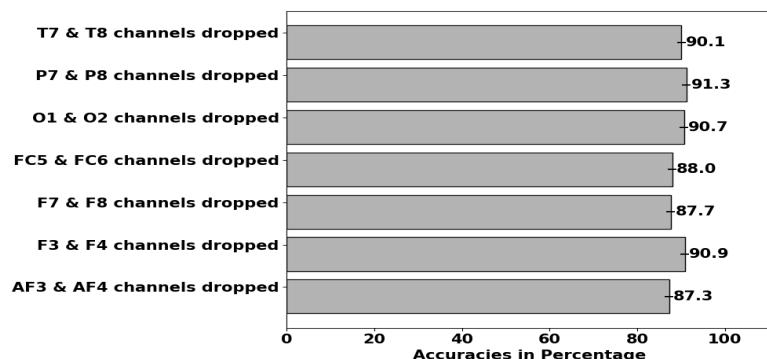


Figure 5.100: Accuracies for Nigeria dataset using ChronoNet for 12 channels

5.3.1.2 4 Channels dropped

Two pairs of channels have been dropped from seven pairs at a time which make total 21 combinations. The accuracies with Guinea-Bissau dataset for all 21 combinations of ten channels have been shown in Figure 5.101.

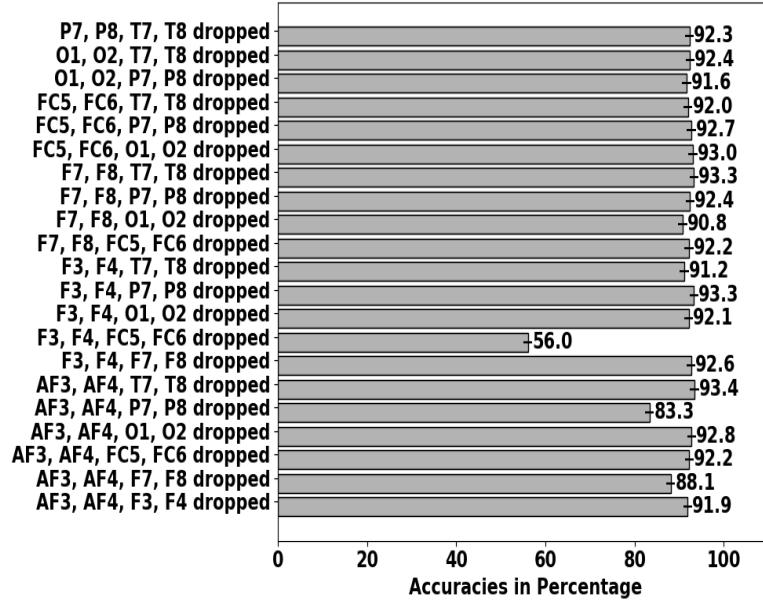


Figure 5.101: Accuracies for Guinea-Bissau dataset using ChronoNet for 10 channels

The accuracies with Nigeria dataset for all 21 combinations of ten channels have been shown in Figure 5.102.

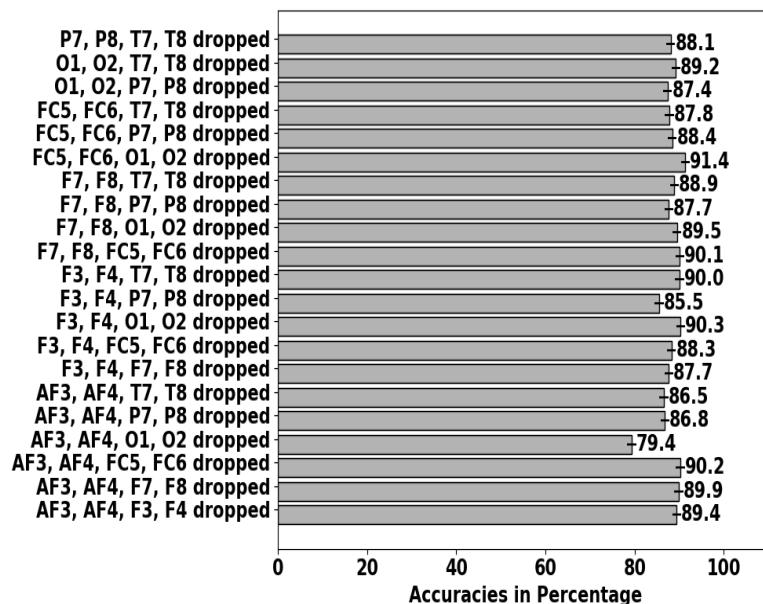


Figure 5.102: Accuracies for Nigeria dataset using ChronoNet for 10 channels

5.3.1.3 6 Channels dropped

Three pairs of channels have been dropped from seven pairs at a time which make total 35 combinations. The accuracies with Guinea-Bissau dataset for all 35 combinations of eight channels have been shown in Figure 5.103.

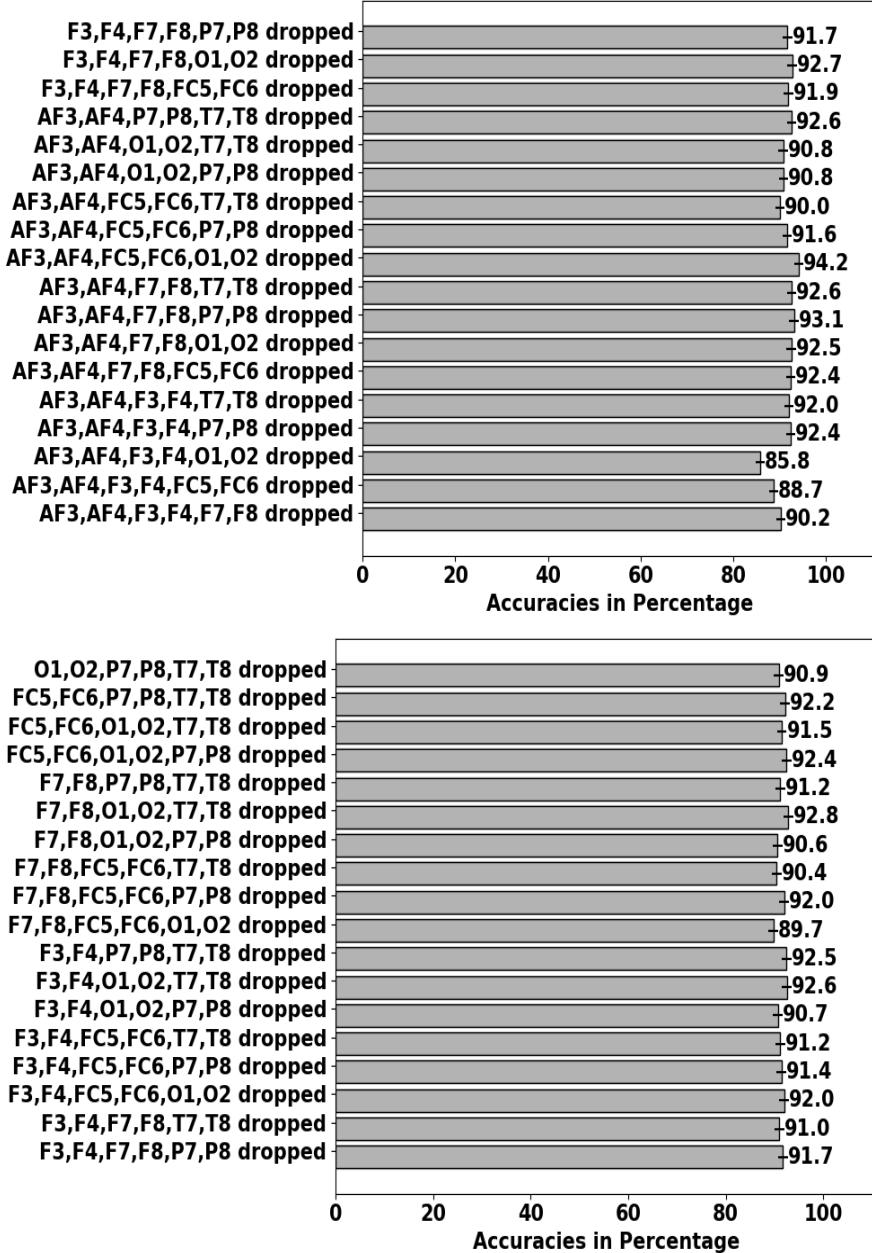


Figure 5.103: Accuracies for Guinea-Bissau dataset using ChronoNet for 8 channels

Three pairs of channels have been dropped from seven pairs at a time which make total 35 combinations. The accuracies with Nigeria dataset for all 35 combinations of eight channels have been shown in Figure 5.104.

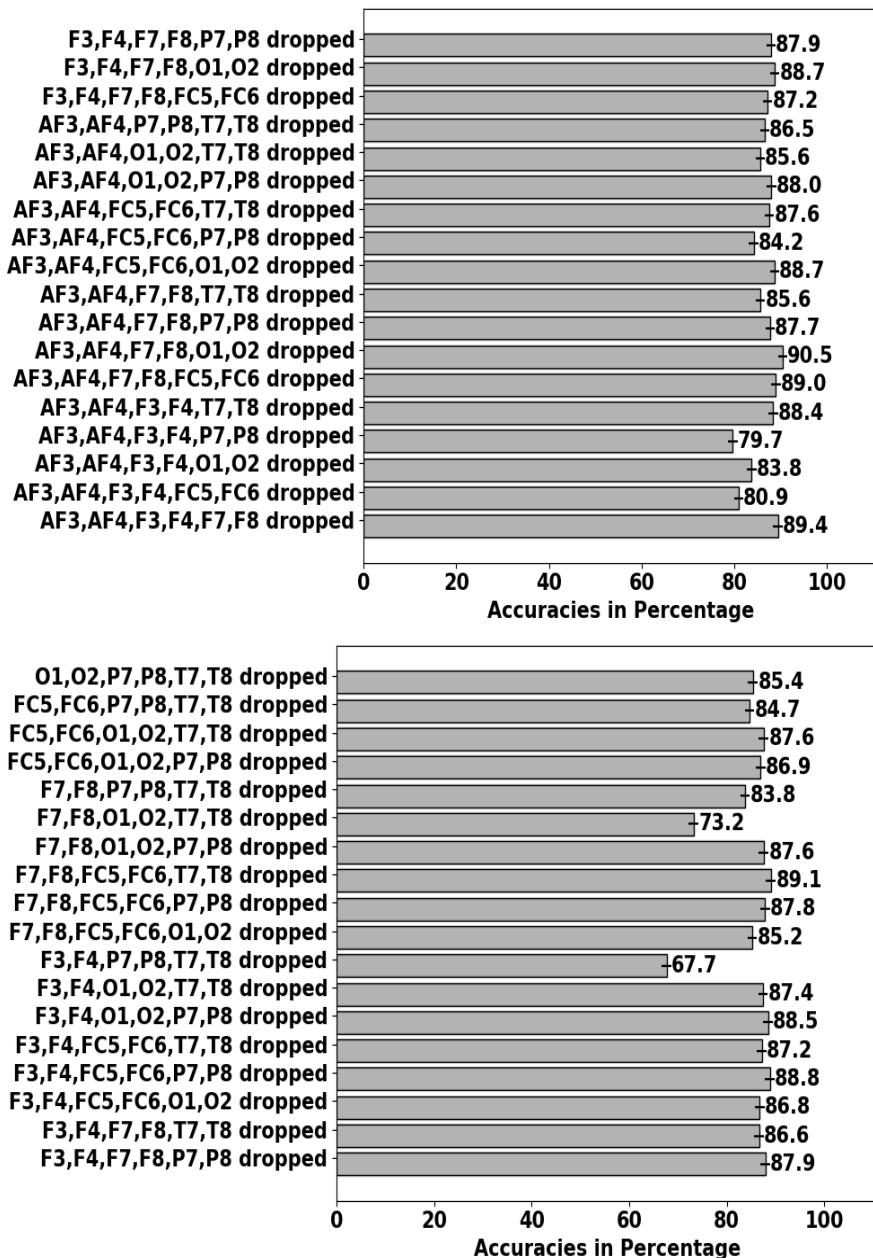


Figure 5.104: Accuracies for Nigeria dataset using ChronoNet for 8 channels

5.3.1.4 8 Channels dropped

Four pairs of channels have been dropped from seven pairs at a time which make total 35 combinations. The accuracies with Guinea-Bissau dataset for all 35 combinations of six channels have been shown in Figure 5.105.

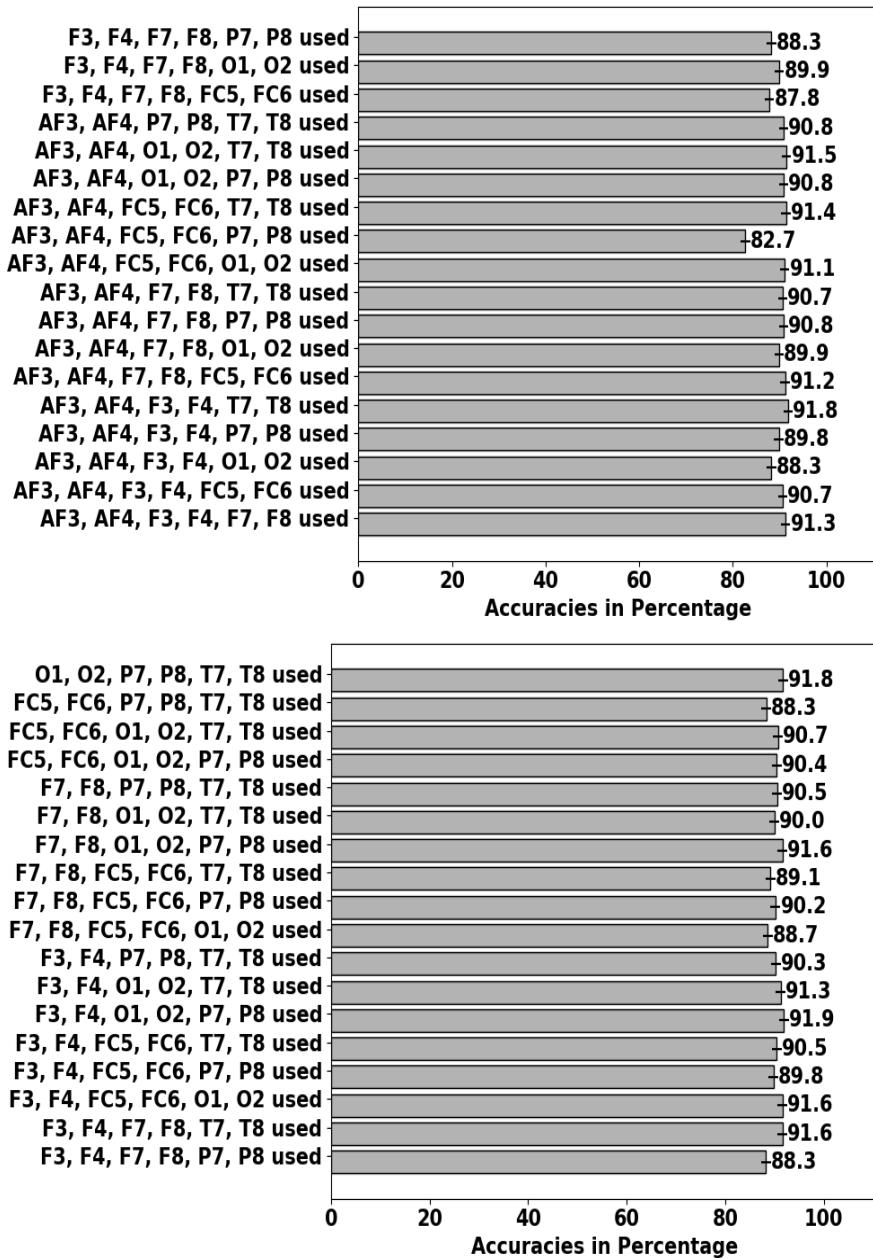


Figure 5.105: Accuracies for Guinea-Bissau dataset using ChronoNet for 6 channels

Four pairs of channels have been dropped from seven pairs at a time which make total 35 combinations. The accuracies with Nigeria dataset for all 35 combinations of six channels have been shown in Figure 5.106.

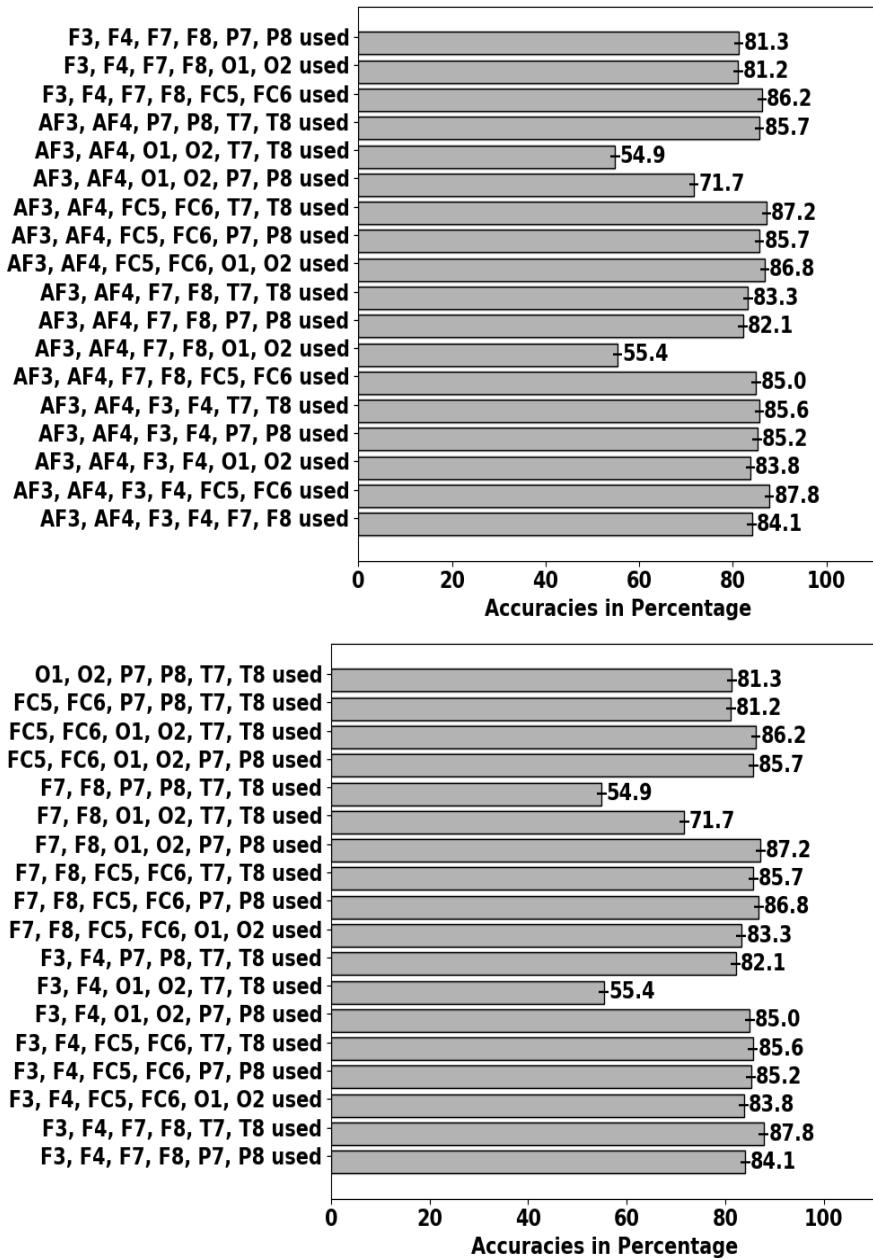


Figure 5.106: Accuracies for Nigeria dataset using ChronoNet for 6 channels

5.3.1.5 10 Channels dropped

Five pairs of channels have been dropped from seven pairs at a time which make total 21 combinations. The accuracies with Guinea-Bissau dataset for all 21 combinations of four channels have been shown in Figure 5.107.

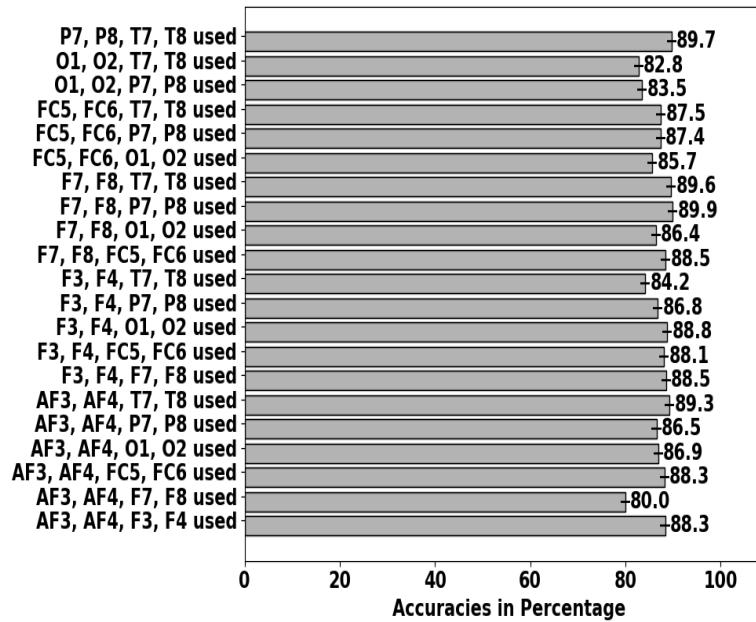


Figure 5.107: Accuracies for Guinea-Bissau dataset using ChronoNet for 4 channels

Five pairs of channels have been dropped from seven pairs at a time which make total 21 combinations. The accuracies with Nigeria dataset for all 21 combinations of four channels have been shown in Figure 5.108.

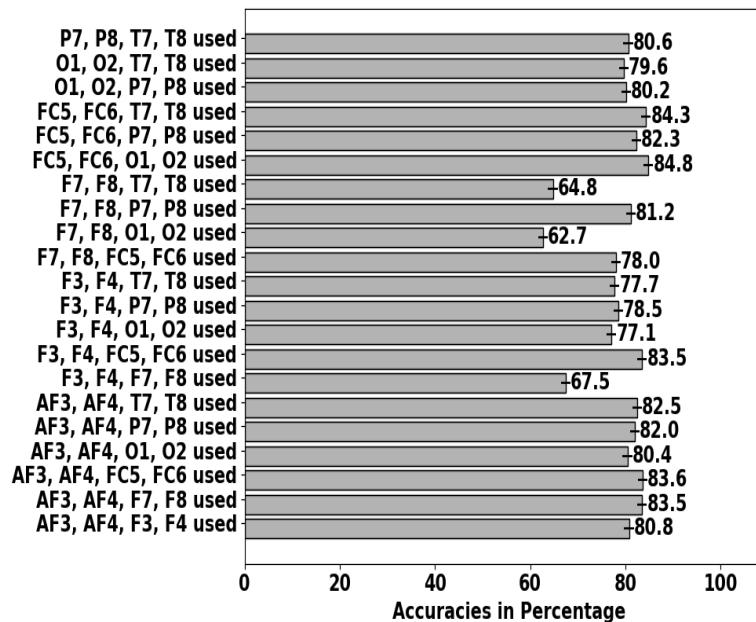


Figure 5.108: Accuracies for Nigeria dataset using ChronoNet for 4 channels

5.3.1.6 12 Channels dropped

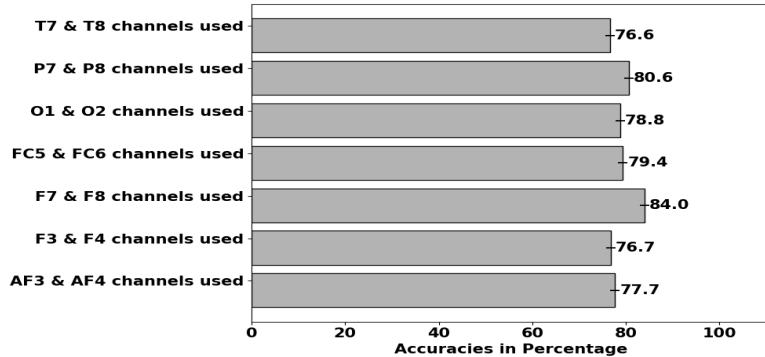


Figure 5.109: Accuracies for Guinea-Bissau dataset using ChronoNet for 2 channels

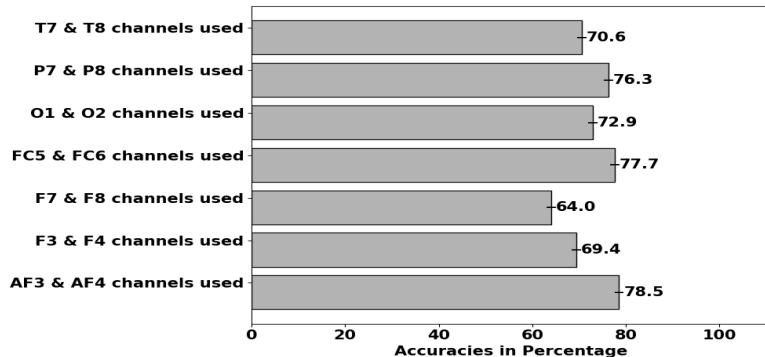


Figure 5.110: Accuracies for Nigeria dataset using ChronoNet for 2 channels

5.3.2 Performance with Selected 8 channels

Here the result of the five best performing classifiers have been investigated and recorded for selected 8 channel conditions shown in Figure 4.8. It is based on keeping only the most discriminative channels for seizure detection found in the literature [4]. The accuracies for all the classifiers are found sufficiently high.

Table 5.1: Accuracies for selected models with AF3, AF4, FC5, FC6, T7, T8, O1, and O2 channels

Model	Nigeria	Guinea-Bissau
C-DRNN	85.48	87.87
ChronoNet	87.9	91.7
mod ChronoNet	86.05	89.41
CNN-LSTM	84.01	86.13
mod CNN-LSTM	87.68	86.93

Chapter 6

Discussion

Epilepsy is a prevalent neurological disorder that affects individuals worldwide, including those residing in rural areas. Timely detection and management of epilepsy are crucial for improving patient outcomes. However, training individuals in rural areas to use complex EEG systems with 14-channel Emotiv EPOC+ headphones can be challenging and costly. In this article, we explore the importance of channel reduction in EEG systems for low-cost epilepsy detection in rural settings, focusing on its potential to enhance accessibility, affordability, and simplicity of initial epilepsy screening.

6.1 Performance Comparison with All Channels

From the experimental results, it can be stated that the implemented deep Learning models work fairly well than machine learning models. Guinea-Bissau dataset have higher accuracies than Nigeria dataset in almost all the models. Maximum accuracy achieved for Guinea-Bissau and Nigeria dataset are 94.6% and 92.1% respectively using ChronoNet and CNN-LSTM model. In case of machine learning models, the performance of SVM classifier performs best with maximum accuracy of 83.2% (Guinea-Bissau) and 77% (Nigeria) without excluding any channels.

6.2 Effect of Single Channel Exclusion

From Table 6.1, it can be noticed that dropping AF3, AF4, F7, F8, F3, P7, and P8 have resulted in significant drop in accuracies. So, these channels are the most significant and cannot be dropped. These channels should be placed accurately.

Table 6.1: Worst accuracies with best performing models

Models	Minimum Accuracy (GB)		Minimum Accuracy (N)	
	Acc %	Channel Excluded	Acc %	Channel Excluded
IC-RNN	88.9	AF3	84.1	AF4
C-DRNN	87.6	P7	86.1	F3
ChronoNet	92.1	P7	82.1	F8
CNN-LSTM	84.2	F8	46.9	F7

No significant performance degradation has been observed for single channel exclusion of this classifier. Among the deep learning models, the four best-performing models in terms of accuracy are ChronoNet (94.6%), CNN-LSTM (92.5%), IC-RNN (91.8%), and C-DRNN (88.6%).

After excluding one channel at a time and investigating their effect on the performance of the four DL models, it has been observed that the most significant and most sensitive channels lie within the frontal and parietal zone.

This finding will be very useful in practice as it indicates that the electrodes in the frontal and parietal zone should be placed precisely for accurate diagnosis of the diseases. In addition, this study also explores the effectiveness of the selected classifiers in detecting seizures in case of failure of any particular EEG signal channel.

6.3 Effect of Multiple Channel Exclusion

From Table 6.2 and Table 6.3, It is evident that the accuracy level remains sufficiently high for ChronoNet when 8 channels are used. The eight-channel combination that may not be used is {AF3, AF4, F7, F8, FC5, FC6, O1, and O2} as the accuracy level drastically falls to a very level of 67.7% for the Nigerian database.

6.4 Effect of Selected 8 Channels

We have conducted an investigation on the results of the top-performing four classifiers under selected 8 channel conditions. The selection process was based on the identification of the most discriminating channels for seizure detection, as previously reported in the literature [4]. Using the Common Spatial Pattern (CSP) technique, we carefully selected the following channels as in Figure ?? for our analysis. The achieved accuracy for all the classifiers was found to be remarkably high as shown in Table 5.1

Table 6.2: Minimum accuracies for ChronoNet model

No. of Channels	Guinea-Bissau	Channels	Nigeria	Channels
2	76.6%	T7 & T8	64%	F7 & F8
4	80%	AF3,AF4,F7,F8	62.7%	F7,F8,O1,O2
6	82.7%	AF3,AF4,FC5, FC6,P7,P8	54.9%	F7,F8,P7, P8,T7,T8
8	85.8%	F7,F8,FC5,FC6 P7,P8,T7,T8	67.7%	AF3,AF4,F7,F8 FC5,FC6,O1,O2
10	56%	AF3,AF4,F7,F8, O1,O2,P7, P8,T7,T8	79.4%	F3,F4,F7,F8, FC5,FC6,P7, P8,T7,T8
12	92.7%	AF3,AF4,F3,F4, F7,F8,FC5,FC6, O1,O2,P7,P8	87.3%	F3,F4,F7,F8, FC5,FC6,O1,O2, P7,P8,T7,T8
13	92.1%	AF3,AF4,F3,F4,F7, F8,FC5,FC6,O1, O2,P8,T7,T8	82.1%	AF3,AF4,F3,F4,F7, FC5,FC6,O1,O2, P7,P8,T7,T8

Table 6.3: Maximum accuracies for ChronoNet model

No. of Channels	Guinea-Bissau	Channels	Nigeria	Channels
2	84%	F7 & F8	78.5%	AF3 & AF4
4	89.9%	F7,F8,P7,P8	84.8%	FC5,FC6,O1,O2
6	91.9%	F3,F4,O1, O2,P7,P8	87.8%	F3,F4,F7 F8,T7,T8
8	94.2%	F3,F4,F7,F8 P7,P8,T7,T8	90.5%	F3,F4,FC5,FC6 P7,P8,T7,T8
10	93.4%	F3,F4,F7,F8, FC5,FC6,O1, O2,P7,P8	91.4%	AF3,AF4,F3,F4, F7,F8,P7, P8,T7,T8
12	93.8%	AF3,AF4,F7,F8, FC5,FC6,O1,O2, P7,P8,T7,T8	91.3%	AF3,AF4,F3,F4, F7,F8,FC5,FC6, O1,O2,T7,T8
13	94%	AF3,AF4,F3,F4,F7, F8,FC5,O1,O2, P7,P8,T7,T8	92.3%	AF3,AF4,F3,F7,F8 FC5,FC6,O1,O2, P7,P8,T7,T8

6.5 Performance of ChronoNet & ChronoNet-M

The performance result and comparison of ChronoNet with ChronoNet-M have been presented here.

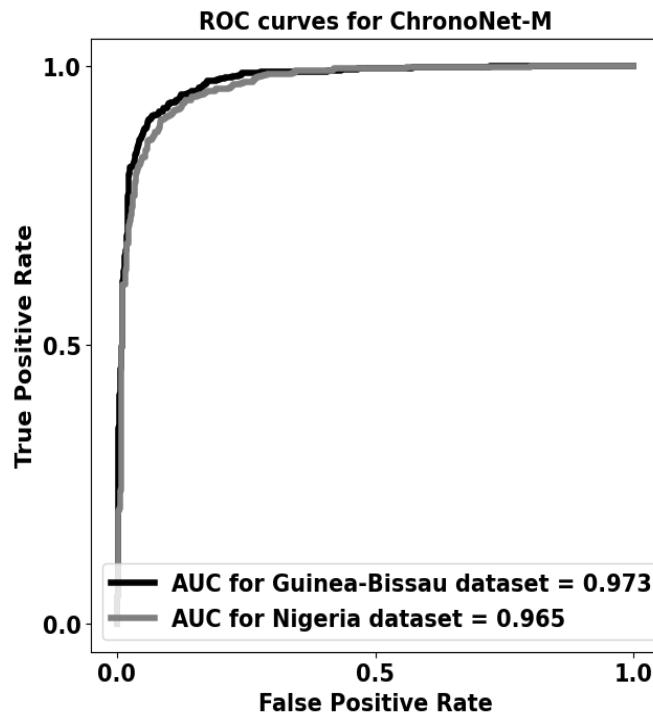


Figure 6.1: AUC of Proposed Model1: ChronoNet-M

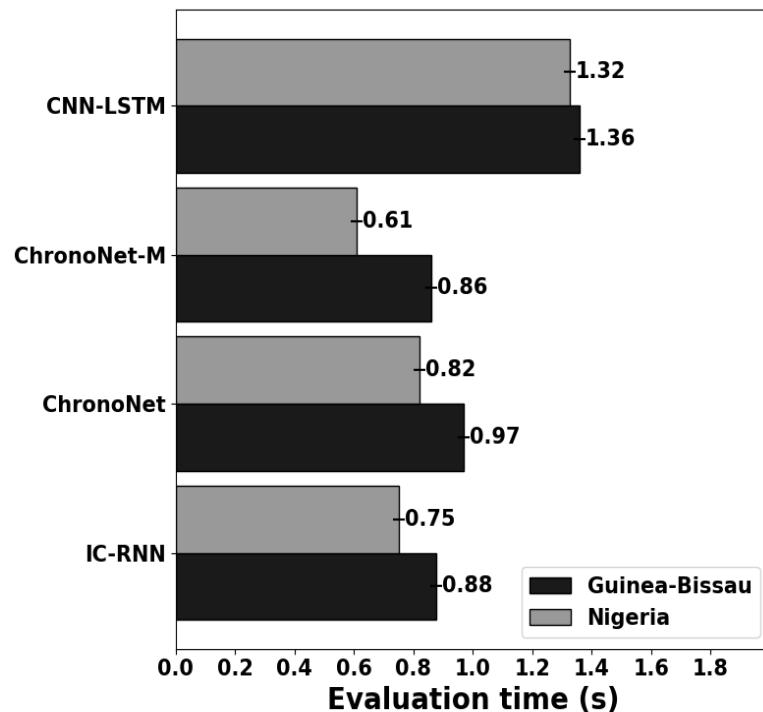


Figure 6.2: Comparison of models evaluation times

For the evaluation of models Tesla T4 (15.1 GB) GPU has been used. From Figure 6.2, it can be observed that model evaluation times for the ChronoNet-M model are

Table 6.4: Comparison of Trainable Parameters

Model	Number of Trainable Parameters
ChronoNet-M	65,121
IC-RNN	124,097
Chrononet	133,313
CNN-LSTM	2,233,921

considerably less than the evaluation time for ChronoNet model. From Table 6.4, it can be noticed that, for ChronoNet-M, the trainable parameters have become almost half compared to ChronoNet.

6.6 Performance of CNN-LSTM & CNN-LSTM-M

The performance result and comparison of CNN-LSTM with CNN-LSTM-M have been presented here.

From table 6.5, we can see that CNN-LSTM-M shows sufficiently high performance for the selected 8-channel electrode montage. It also shows robustness in performance in other combinations of different 8-channel montages.

Table 6.5: Accuracies for CNN-LSTM and CNN-LSTM-M with 8 channels

Channels used	Guinea-Bissau		Nigeria	
	CNN-LSTM	CNN-LSTM-M	CNN-LSTM	CNN-LSTM-M
AF3,AF4,FC5,FC6, T7,T8,O1,O2	86.13	90.55	84.01	85.73
AF3,AF4,F3,F4, F7,F8,P7,P8	91.15	89.68	73.41	81
AF3,AF4,T7,T8, P7,P8,O1,O2	87.87	89.54	85.32	72.51
F3,F4,T7,T8, P7,P8,O1,O2	87.33	87.06	86.62	81.89
F7,F8,T7,T8, P7,P8,O1,O2	88.87	90.15	84.75	82.71
FC5,FC6,T7,T8, P7,P8,O1,O2	86.46	83.65	86.7	89.72

Figure 6.3 shows that CNN-LSTM-M have accuracies greater than 90% for both datasets. It outperforms CNN-LSTM for Guinea-Bissau dataset. Figure 6.4 shows that the AUCs for CNN-LSTM-M for Guinea-Bissau and Nigera dataset are 0.973 and 0.974 respectively which are pretty close to 1.

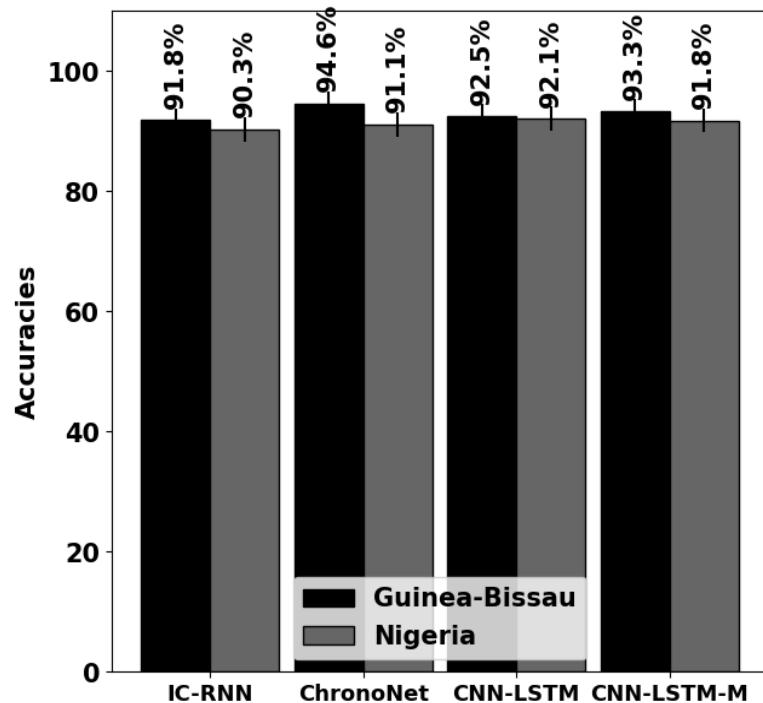


Figure 6.3: Accuracy comparison of CNN-LSTM-M with other models for 14 channels

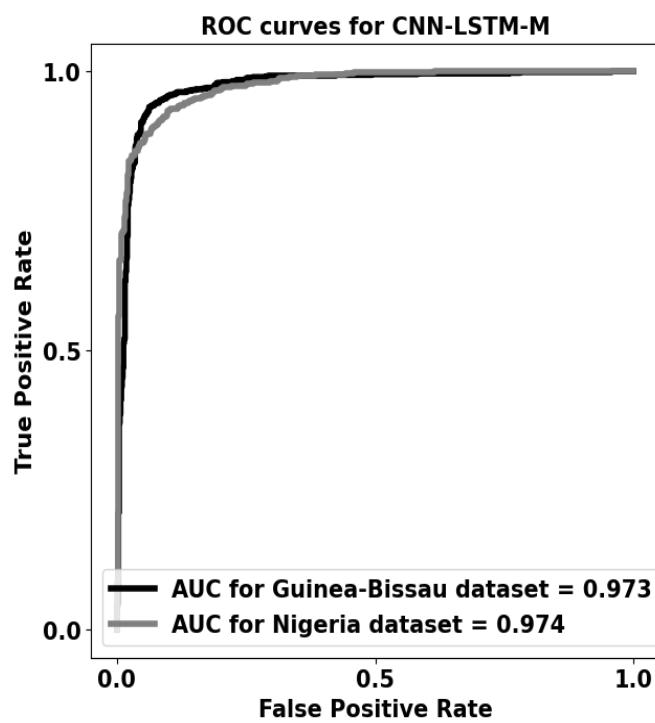


Figure 6.4: AUC of CNN-LSTM-M

In this study, we have employed 14 state-of-the-art classifiers, as reported in recent literature, to distinguish between individuals with epilepsy and healthy individuals. To assess the significance of each channel, we have observed the decline in accuracy across all classifiers when excluding a single channel at a time. Additionally, we have investigated the impact of excluding multiple channels on the performance of the best-performing models to determine the optimal electrode configuration.

In combination with the data-driven experimental approach, we have also incorporated prior knowledge-based methods from recent literature to identify the best electrode montage. Consequently, an 8-channel EEG montage have been selected as the most effective configuration for the classification task.

In this research, novel modified classifiers, referred to as ChronoNet-M and CNN-LSTM-M, have been introduced based on the two best-performing classifiers, ChronoNet and CNN-LSTM. The research has extensively investigated the influence of channel reduction on these newly proposed classifiers.

The results demonstrate that the innovative approach to classifier selection, coupled with the reduced EEG channels, not only enhances cost-effectiveness but also preserves high detection accuracy. Remarkably, the newly developed classifiers outperformed others utilizing the same dataset, even when operating with the reduced EEG montage.

Chapter 7

Conclusions

In this study, we have investigated how well channel reduction work in identifying epilepsy from EEG data. Starting with 14 pre-built models and 14 Emotiv Epoc+ channels, we have used a methodical strategy to gradually remove one channel at a time which make 14 combinations left. Finding the most important channels for epilepsy detection was our goal. In order to find the best configurations, we have then concentrated on using eight channels, carefully choosing combinations from the partial, frontal, temporal, and occipital zones.

We have discovered a number of important things through our extensive investigation. First, using channel reduction, we have been able to pinpoint important channels that showed a direct association with epileptic activity. We have learned more about the relative significance of each electrode and its contributions by gradually removing channels. Additionally, the outcomes of our investigation of the eight-channel combinations have been encouraging. We have discovered that the overall effectiveness of epilepsy detection has been enhanced by carefully chosen combinations from particular brain areas. According to these results, fewer channels do not always result in a decrease in the diagnostic precision of EEG-based seizure identification. Instead, it enables a more focused and effective approach, concentrating on the most educational channels.

Our research has important ramifications, especially in environments with little resources, like rural places. We have solved the practical issues of cost-effectiveness and training by lowering the number of needed electrodes. Our research paves the way for the deployment of streamlined EEG equipment in outlying healthcare institutions, enabling early epilepsy identification without compromising precision. This strategy may increase access to epilepsy screening and enhance patient outcomes in disadvantaged areas. It's crucial to recognize the limits of our study though. It is important to confirm the generalizability of our results using a bigger and more varied dataset. Further inves-

tigation is required to examine the potential effects of various channel combinations as well as their resilience across various populations and epilepsy subtypes. Our study's result emphasizes the need of channel reduction for EEG-based epilepsy identification. We have proved the viability of employing a smaller set of electrodes without affecting diagnostic accuracy by identifying critical channels and optimizing combinations. This research adds to the expanding corpus of work targeted at making EEG-based diagnostics for epilepsy easier to use and more accessible, especially in environments with limited resources. The area of epilepsy diagnosis may undergo a revolution, and those who suffer from this neurological illness may benefit from continued research in this regard.

7.1 Major contributions

- The performance of 14 state-of-the-art classifiers has been analyzed.
- The effect of input channel reduction has been investigated.
 - Most significant electrodes have been identified.
- Optimal number of input channels has been selected.
 - Input channel reduced to 8 with a slight loss of accuracy.
- 2 Novel less-complex classifiers have been proposed.

7.2 Challenges

There are a number of difficulties in epilepsy detection that we have faced. Here are a few issues with data availability, computing power, and algorithm complexity.

- Scarcity of GPU and RAM: It frequently takes a lot of computer power to train and evaluate machine learning or deep learning models for the diagnosis of epilepsy. The effective processing and interpretation of EEG data may be hampered by the lack of sufficient RAM and specialized GPUs. The training and testing procedures may become more labor-intensive and ineffective without access to these resources.
- Limited Data Availability: It might be difficult to gather a sizable and varied collection of EEG recordings for epilepsy detection. There may be restrictions

on data collecting because to things like resource shortages, patient access restrictions, and privacy issues. The construction and assessment of reliable and accurate models might be hampered by a lack of data.

- Interpretability and Understanding: Epilepsy detection models must be readable and understandable, especially in clinical settings. Gaining the confidence and acceptance of medical professionals and patients requires an understanding of the rationale behind model predictions and the recognition of the significant elements or patterns that influence a categorization choice.
- Algorithm Selection and Optimization: It might be difficult to select the best deep learning or machine learning algorithms for epilepsy detection. In terms of processing resources and data properties, different algorithms have distinct strengths, limits, and requirements. For optimal performance and accuracy, tweaking algorithm parameters and designs frequently necessitates a great deal of testing.
- Generalization and Variability: Individual differences in epileptic seizure patterns and features can be substantial. It is difficult to create models that can adequately generalize across different patient groups and adjust to inter-individual and intra-individual heterogeneity.
- Data Preprocessing and Feature Extraction: The preparation of EEG data and the feature extraction procedure can be difficult and time-consuming. EEG recordings can suffer from signal distortions, noise, and electrode movement, necessitating rigorous preprocessing procedures to improve signal reliability.

7.3 Future Scope

- The effects of electrode displacement are visible. The accuracy of the recorded EEG signals may be harmed if electrodes are not positioned correctly or move during data capture. It can be difficult to precisely detect epileptic occurrences as a result of noise, artifacts, or distorted signal patterns.
- Transfer learning and transformer models can be used for classification after transforming the signals to pictures. Convolutional neural networks (CNNs) and transformer-based models are examples of pre-trained models that may be used to accurately identify epilepsy. This would include tuning the models, particularly for the categorization of epileptic seizures after training them on large-scale datasets.

References

- [1] A. Miltiadous, K. D. Tzimourta, N. Giannakeas, M. G. Tsipouras, E. Glavas, K. Kalafatakis, and A. T. Tzallas, “Machine learning algorithms for epilepsy detection based on published EEG databases: A systematic review,” *IEEE Access*, 2022.
- [2] A. A. Ein Shoka, M. M. Dessouky, A. El-Sayed, and E. E.-D. Hemdan, “EEG seizure detection: concepts, techniques, challenges, and future trends,” *Multimedia Tools and Applications*, pp. 1–31, 2023.
- [3] S. Gong, K. Xing, A. Cichocki, and J. Li, “Deep learning in EEG: Advance of the last ten-year critical period,” *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 2, pp. 348–365, 2021.
- [4] M. Tacke, K. Janson, K. Vill, F. Heinen, L. Gerstl, K. Reiter, and I. Borggraefe, “Effects of a reduction of the number of electrodes in the EEG montage on the number of identified seizure patterns,” *Scientific Reports*, vol. 12, no. 1, pp. 1–7, 2022.
- [5] A. Apicella, P. Arpaia, F. Isgrò, G. Mastrati, and N. Moccaldi, “A survey on eeg-based solutions for emotion recognition with a low number of channels,” *IEEE Access*, vol. 10, pp. 117411–117428, 2022.
- [6] L. A. Moctezuma and M. Molinas, “EEG channel-selection method for epileptic-seizure classification based on multi-objective optimization,” *Frontiers in neuroscience*, vol. 14, p. 593, 2020.
- [7] P. Mathur, V. K. Chakka, and S. B. Shah, “Ramanujan periodic subspace based epileptic EEG signals classification,” *IEEE Sensors Letters*, vol. 5, no. 7, pp. 1–4, 2021.
- [8] M. Chakraborty, D. Mitra *et al.*, “Epilepsy seizure detection using kurtosis based vmd’s parameters selection and bandwidth features,” *Biomedical Signal Processing and Control*, vol. 64, p. 102255, 2021.

- [9] S. Rahman, M. M. Hasan, and A. K. Sarkar, “Methods for detecting epilepsy: A comparative analysis on improved deep learning algorithms,” in *2022 12th International Conference on Electrical and Computer Engineering (ICECE)*. IEEE, 2022, pp. 356–359.
- [10] J. F. Moreau, E. L. Fink, M. E. Hartman, D. C. Angus, M. J. Bell, W. T. Linde-Zwirble, and R. S. Watson, “Hospitalizations of children with neurological disorders in the united states,” *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*, vol. 14, no. 8, p. 801, 2013.
- [11] C. M. Horvat, H. Mtaweh, and M. J. Bell, “Management of the pediatric neuro-critical care patient,” in *Seminars in neurology*, vol. 36, no. 06. Thieme Medical Publishers, 2016, pp. 492–501.
- [12] B. Appavu, B. T. Burrows, T. Nickoles, V. Boerwinkle, A. Willyerd, V. Gunnala, T. Mangum, I. Marku, and P. Adelson, “Implementation of multimodality neurologic monitoring reporting in pediatric traumatic brain injury management,” *Neurocritical Care*, vol. 35, pp. 3–15, 2021.
- [13] S. T. Herman, N. S. Abend, T. P. Bleck, K. E. Chapman, F. W. Drislane, R. G. Emerson, E. E. Gerard, C. D. Hahn, A. M. Husain, P. W. Kaplan *et al.*, “Consensus statement on continuous eeg in critically ill adults and children, part i: indications,” *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, vol. 32, no. 2, p. 87, 2015.
- [14] T. Rowberry, H. K. Kanthimathinathan, F. George, L. Notghi, R. Gupta, P. Bill, E. Wassmer, H. P. Duncan, K. P. Morris, and B. R. Scholefield, “Implementation and early evaluation of a quantitative electroencephalography program for seizure detection in the picu,” *Pediatric Critical Care Medicine*, vol. 21, no. 6, pp. 543–549, 2020.
- [15] L. F. Haas, “Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, no. 1, pp. 9–9, 2003.
- [16] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, “A practical vep-based brain-computer interface,” *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 14, no. 2, pp. 234–240, 2006.
- [17] J. Li, Y. Wang, L. Zhang, and T.-P. Jung, “Combining erps and eeg spectral features for decoding intended movement direction,” in *2012 Annual International*

- Conference of the IEEE Engineering in Medicine and Biology Society.* IEEE, 2012, pp. 1769–1772.
- [18] Y. Zhang, G. Zhou, J. Jin, Q. Zhao, X. Wang, and A. Cichocki, “Sparse bayesian classification of eeg for brain–computer interface,” *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2256–2267, 2015.
 - [19] J. Li, C. Li, and A. Cichocki, “Canonical polyadic decomposition with auxiliary information for brain–computer interface,” *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 263–271, 2015.
 - [20] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, “Eeg-based emotion classification using deep belief networks,” in *2014 IEEE international conference on multimedia and expo (ICME)*. IEEE, 2014, pp. 1–6.
 - [21] H. Chao, L. Dong, Y. Liu, and B. Lu, “Emotion recognition from multiband eeg signals using capsnet,” *Sensors*, vol. 19, no. 9, p. 2212, 2019.
 - [22] F. Li, G. Zhang, W. Wang, R. Xu, T. Schnell, J. Wen, F. McKenzie, and J. Li, “Deep models for engagement assessment with scarce label information,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 598–605, 2016.
 - [23] L. Cao, W. Tao, S. An, J. Jin, Y. Yan, X. Liu, W. Ge, A. Sah, L. Battle, J. Sun *et al.*, “Smile: A system to support machine learning on eeg data at scale,” *Proceedings of the VLDB Endowment*, vol. 12, no. 12, pp. 2230–2241, 2019.
 - [24] I. B. Slimen, L. Boubchir, and H. Seddik, “Epileptic seizure prediction based on eeg spikes detection of ictal-preictal states,” *Journal of biomedical research*, vol. 34, no. 3, p. 162, 2020.
 - [25] E. Giourou, A. Stavropoulou-Deli, A. Giannakopoulou, G. K. Kostopoulos, and M. Koutroumanidis, “Introduction to epilepsy and related brain disorders,” *Cyberphysical Systems for Epilepsy and Related Brain Disorders: Multi-parametric Monitoring and Analysis for Diagnosis and Optimal Disease Management*, pp. 11–38, 2015.
 - [26] F. Ibrahim, S. Abd-Elateif El-Gindy, S. M. El-Dolil, A. S. El-Fishawy, E.-S. M. El-Rabaie, M. I. Dessouky, I. M. Eldokany, T. N. Alotaiby, S. A. Alshebeili, and F. E. Abd El-Samie, “A statistical framework for eeg channel selection and seizure prediction on mobile,” *International Journal of Speech Technology*, vol. 22, pp. 191–203, 2019.

- [27] F. Li, Y. Liang, L. Zhang, C. Yi, Y. Liao, Y. Jiang, Y. Si, Y. Zhang, D. Yao, L. Yu *et al.*, “Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp eeg network analysis,” *Cognitive Neurodynamics*, vol. 13, pp. 175–181, 2019.
- [28] A. Shoka, M. Dessouky, A. El-Sherbeny, and A. El-Sayed, “Literature review on eeg preprocessing, feature extraction, and classifications techniques,” *Menoufia J. Electron. Eng. Res*, vol. 28, no. 1, pp. 292–299, 2019.
- [29] M. I. Jordan and T. M. Mitchell, “Machine learning: Trends, perspectives, and prospects,” *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [30] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, “Machine learning, neural and statistical classification,” 1994.
- [31] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [32] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen *et al.*, “Deep speech 2: End-to-end speech recognition in english and mandarin,” in *International conference on machine learning*. PMLR, 2016, pp. 173–182.
- [33] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, “Object detection with deep learning: A review,” *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [34] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: A comprehensive review,” *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.
- [35] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, “Speech recognition using deep neural networks: A systematic review,” *IEEE access*, vol. 7, pp. 19 143–19 165, 2019.
- [36] S. Roy, I. Kiral-Kornek, and S. Harrer, “Chrononet: a deep recurrent neural network for abnormal eeg identification,” in *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17*. Springer, 2019, pp. 47–56.
- [37] G. Xu, T. Ren, Y. Chen, and W. Che, “A one-dimensional cnn-lstm model for epileptic seizure recognition using eeg signal analysis,” *Frontiers in neuroscience*, vol. 14, p. 578126, 2020.

- [38] X. Zou, Y. Hu, Z. Tian, and K. Shen, “Logistic regression model optimization and case analysis,” in *2019 IEEE 7th international conference on computer science and network technology (ICCSNT)*. IEEE, 2019, pp. 135–139.
- [39] J. Qin and Y. Lou, “l 1-2 regularized logistic regression,” in *2019 53rd Asilomar conference on signals, systems, and computers*. IEEE, 2019, pp. 779–783.
- [40] A. H. Jahromi and M. Taheri, “A non-parametric mixture of gaussian naive bayes classifiers based on local independent features,” in *2017 Artificial intelligence and signal processing conference (AISP)*. IEEE, 2017, pp. 209–212.
- [41] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu *et al.*, “Top 10 algorithms in data mining,” *Knowledge and information systems*, vol. 14, pp. 1–37, 2008.
- [42] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” *arXiv preprint arXiv:1302.4964*, 2013.
- [43] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
- [44] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “Knn model-based approach in classification,” in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*. Springer, 2003, pp. 986–996.
- [45] D. J. Hand, “Principles of data mining,” *Drug safety*, vol. 30, pp. 621–622, 2007.
- [46] W. S. Noble, “What is a support vector machine?” *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [47] M. Zhou, C. Tian, R. Cao, B. Wang, Y. Niu, T. Hu, H. Guo, and J. Xiang, “Epileptic seizure detection based on eeg signals and cnn,” *Frontiers in neuroinformatics*, vol. 12, p. 95, 2018.
- [48] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [49] A. Graves, “Generating sequences with recurrent neural networks,” *arXiv preprint arXiv:1308.0850*, 2013.
- [50] S. Roy, I. Kiral-Kornek, and S. Harrer, “Chrononet: a deep recurrent neural network for abnormal eeg identification,” in *Artificial Intelligence in Medicine: 17th*

- Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings 17.* Springer, 2019, pp. 47–56.
- [51] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv preprint arXiv:1409.1259*, 2014.
- [52] V. T. van Hees, E. van Diessen, M. R. Sinke, J. W. Buitenhuis, F. van der Maas, L. Ridder, and W. M. Otte, “Reliable and automatic epilepsy classification with affordable, consumer-grade electroencephalography in rural sub-saharan africa,” *BioRxiv*, p. 324954, 2018.