

TOWARDS AN EFFECTIVE APPROACH OF  
INSIDER ATTACKS DETECTION USING  
THE HUMAN PHYSIOLOGICAL SIGNALS

BY

**AZZAT AHMED ALI AL-SADI**

A Dissertation Presented to the  
DEANSHIP OF GRADUATE STUDIES

**KING FAHD UNIVERSITY OF PETROLEUM & MINERALS**

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the  
Requirements for the Degree of

**DOCTOR OF PHILOSOPHY**

In

**COMPUTER SCIENCE AND ENGINEERING**

NOVEMBER 2018

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **Azzat Ahmed Ali AL-Sadi** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **DOCTOR OF PHILOSOPHY IN COMPUTER SCIENCE AND ENGINEERING.**



Dr. Adel Ahmed  
Department Chairman



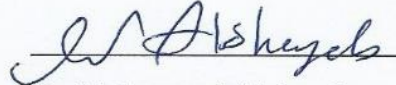
Dr. Salam A. Zummo  
Dean of Graduate Studies

1-04-2019

Date



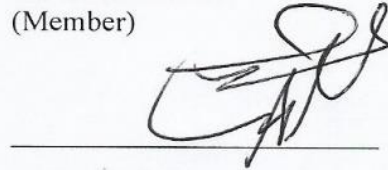
Dr. Mahmood Niazi  
(Advisor)




Dr. Mohammad Alshayeb  
(Co-Advisor)



Dr. Shokri Selim  
(Member)



Dr. Tarek Sheltami  
(Member)



Dr. Mohammad Elrabaa  
(Member)

© Azzat Ahmed Ali AL-Sadi

2018



*Dedication*

*To my father, whose journey through life has demonstrated the true meaning of hard work, courage, and perseverance; I dedicate this work to your valuable and imprinted words for higher academic achievements.*

*May your soul rest in eternal peace.*

## ACKNOWLEDGMENTS

*Acknowledgment is due to King Fahd University of Petroleum and Minerals and Hadhramout University for supporting this research. I would also like to acknowledge my sponsors, Hadhramout Establishment for Human Development, for granting me this outstanding opportunity to obtain my Ph.D. degree, and for their generous support.*

*Moreover, I would like to express my sincere appreciation to the dissertation committee for the encouragement and trust that they have extended to me. Many thanks due to Dr. Mahmood Niazi, Dr. Mohammad Alshayeb, Dr. Shokri Selim, Dr. Tareq Sheltami and Dr. Mohammad Elrabaa.*

*I would also like to express my deep gratitude for my main mentor in this research, Dr. Mahmood Niazi for his guidance and support. I am particularly grateful and fortunate to have worked with him, and with my thesis committee members, as I have gained valuable insights into this field.*

*I would also like to thank all my family and friends who cared to share this experience and provided mental support.*

# TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS</b> .....	<b>VI</b>
<b>TABLE OF CONTENTS</b> .....	<b>VII</b>
<b>LIST OF TABLES</b> .....	<b>XIII</b>
<b>LIST OF FIGURES</b> .....	<b>XIV</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>XVI</b>
<b>ABSTRACT</b> .....	<b>XVII</b>
<b>ملخص الرسالة</b> .....	<b>XIX</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>1</b>
<b>1.1 Overview</b> .....	<b>1</b>
<b>1.2 Problem Statement and Motivation</b> .....	<b>2</b>
<b>1.2.1 The Ability of Physiological Signals to Identify Human Moods and Intentions</b> .....	<b>4</b>
<b>1.2.2 Ease of Physiological Data Collection and Analyzing</b> .....	<b>5</b>
<b>1.3 Research Objectives</b> .....	<b>6</b>
<b>1.4 Research Questions</b> .....	<b>9</b>
<b>1.5 Summary of Research Contribution</b> .....	<b>10</b>
<b>1.6 Thesis Roadmap</b> .....	<b>12</b>
<b>1.7 Summary</b> .....	<b>13</b>
<b>CHAPTER 2 BACKGROUND</b> .....	<b>14</b>
<b>2.1 Cybersecurity</b> .....	<b>14</b>
<b>2.1.1 Access Control</b> .....	<b>15</b>
<b>2.1.2 Software Development Security</b> .....	<b>15</b>

2.1.3	Business Continuity and Disaster Recovery Planning .....	15
2.1.4	Cryptography.....	16
2.1.5	Information security and risk management .....	16
2.1.6	Law, Investigation, and Ethics.....	17
2.1.7	Operations Security .....	17
2.1.8	Physical and Environmental Security .....	17
2.1.9	Security Architecture and Design.....	18
2.1.10	Telecommunications and Network Security .....	18
2.2	Classics System Security Characteristics .....	19
2.3	Types of Insider Attacks.....	20
2.3.1	Fraud .....	20
2.3.2	IT Sabotage.....	21
2.3.3	Insider Theft of Intellectual Property.....	22
2.3.4	Espionage .....	22
2.4	Motivations behind Insider Attacks .....	24
2.5	Impact of Insider Attacks .....	27
2.6	Insider Threat Detection: Existing Solutions .....	28
2.7	Summary .....	29
<b>CHAPTER 3 LITERATURE REVIEW .....</b>		<b>31</b>
3.1	Overview .....	31
3.2	Awareness of Insider Attacks.....	31
3.3	Insider Attack Detection Methods .....	33
3.3.1	Anomaly Detection Methods.....	35
3.3.2	Honeytrap Traps .....	40
3.3.3	Graph-Based Methods.....	44



3.3.4	Game-Based Methods.....	48
3.3.5	Physiological Methods .....	50
3.4	Limitations in the Existing Studies.....	64
3.4.1	Limitations of Anomaly Detection Methods.....	64
3.4.2	Limitations of Honeypot Traps .....	65
3.4.3	Limitations of Game-Based Approaches.....	66
3.4.4	Limitations of Graph-Based Approaches .....	67
3.4.5	Limitations of physiological methods.....	68
3.5	Summary .....	71
<b>CHAPTER 4 RESEARCH METHODOLOGY .....</b>		<b>73</b>
4.1	Overview .....	73
4.2	Stage 1: The Literature Survey .....	73
4.3	Stage 2: Building the Bio-signals Data Set .....	75
4.4	Stage 3: Features Extraction.....	79
4.5	Stage 4: Proposed System .....	80
4.5.1	Sensors .....	81
4.5.2	Interfaces.....	83
4.5.3	Features Extractor.....	83
4.5.4	Attack Assessment.....	84
4.5.5	Comparative Signal Database .....	85
4.5.6	Attack Detector.....	86
4.6	Evaluating the Proposed System .....	86
4.7	Summary .....	87
<b>CHAPTER 5 EXPERIMENTAL WORK.....</b>		<b>89</b>

5.1	Overview .....	89
5.2	Experiment Setup .....	90
5.3	Experiment Environment.....	90
5.4	Experiment Scenarios .....	91
5.4.1	First Scenario: Normal Activities .....	92
5.4.2	Second Scenario: Malicious Activities .....	93
5.5	Bio-signals Data set .....	95
5.6	Ethical Considerations .....	95
5.7	Experiment Devices .....	96
5.7.1	The NeuroSky MindWave .....	96
5.7.2	Wild Divine.....	97
5.8	Brain Waves .....	98
5.8.1	An Electroencephalogram (EEG) .....	98
5.8.2	Brainwaves: Types and Functions .....	98
5.9	An Electrocardiogram (ECG).....	100
5.9.1	Types of ECG Wave Components .....	101
5.10	Feature Extraction .....	103
5.10.1	Signal Preprocessing .....	103
5.10.2	EEG Features .....	104
5.10.3	ECG Features .....	107
5.10.4	Feature Frame .....	108
5.11	Machine Learning .....	108
5.11.1	Random Forest .....	110
5.11.2	Support Vector Machine.....	111
5.11.3	Neural Network.....	113

5.12	Summary .....	114
<b>CHAPTER 6 RESULTS AND EVALUATION .....</b>		<b>116</b>
6.1	Overview .....	116
6.2	Hypotheses .....	116
6.3	Evaluation Metrics .....	118
6.3.1	Accuracy .....	120
6.3.2	Precision .....	120
6.3.3	Recall .....	121
6.3.4	F-score .....	121
6.3.5	Area Under the Curve AUC.....	121
6.3.6	Kappa.....	122
6.3.7	Matthews Correlation Coefficient .....	122
6.3.8	Percent Difference .....	123
6.3.9	Cross-Validation.....	123
6.3.10	Confidence Interval.....	123
6.4	Hypotheses Testing and Validity .....	124
6.5	Presenting Results .....	125
6.6	Results of EEG Features.....	127
6.6.1	Rejection of Null Hypothesis H0-1.....	135
6.7	Results of EEG+ECG Features .....	136
6.7.1	Rejection of Null Hypothesis H0-2.....	138
6.8	Classification Accuracy Assessments.....	138
6.9	Evaluation Using Three Classifiers.....	141
6.10	Evaluation Using a Group of Frames .....	146
6.11	Evaluation Using Varied Amount of Malicious Data .....	147

6.12	Evaluation for New Incoming Data .....	152
6.13	Evaluating proposed method with Suh and Yim approach.....	155
6.14	Summary .....	157
<b>CHAPTER 7 CONCLUSION, LIMITATIONS AND FUTURE RESEARCH.....</b>		<b>158</b>
7.1	Conclusion .....	158
7.2	Limitations.....	162
7.2.1	Sensors .....	162
7.2.2	Number of Used Devices .....	163
7.2.3	Hardware Limitations .....	163
7.2.4	Unused Bio-signals .....	163
7.2.5	Environment of collecting Data .....	164
7.2.6	Final Product Deployment .....	164
7.3	Future Research.....	165
<b>REFERENCES.....</b>		<b>166</b>
<b>APPENDIX A.....</b>		<b>180</b>
<b>APPENDIX B.....</b>		<b>189</b>
<b>VITAE.....</b>		<b>191</b>

## LIST OF TABLES

Table 1: The Description of XABA Scenarios .....	38
Table 2: The Proposed Scenarios by Young <i>et al.</i> [57].....	39
Table 3: The Categorization of Insider Threats Target The Nuclear Reactors .....	50
Table 4: The Main And Subclasses Of The Human Factors by Greitzer <i>et al.</i> [93] .....	52
Table 5: The Evaluation Indicators Greitzer <i>et al.</i> [9].....	53
Table 6: Extracted Features by Babu and Bhanu [83].....	55
Table 7: Summarizing The Surveyed Methods .....	62
Table 8: Features Extractor Output .....	84
Table 9: Frequency Ranges of EEG Bands .....	105
Table 10: Confusion Matrix .....	119
Table 11: PD of Accuracy Ranges .....	137
Table 12: Comparing Approaches using Confidence Interval, FPR and FNR.....	142
Table 13: Comparison of Proposed Approaches and the Raw Brainwaves .....	143
Table 14: Comparing EEG+ECG with the Extracted EEG Based on Gender .....	143
Table 15: PD of The Classifiers' Accuracy Using EEG Features.....	145
Table 16: PD of Classifiers's Accuracy Using EEG+ECG .....	146
Table 17: Confidence Intervale, FP and FN Rates of Proposed EEG+ECG.....	152
Table 18: Results of EEG+ECG Approach Per Participant .....	153
Table 19: Average Accuracy, FPR and FNR of The 84 Participants .....	153
Table 20: Comparing Approaches Based on Gender .....	155

## LIST OF FIGURES

Figure 1: The CERT Breakdown of Intentional Insider Crimes in United States .....	23
Figure 2: Vulnerable Assets For Insider Attacks .....	28
Figure 3: The Taxonomy of Insider Attacks Detection Methods .....	34
Figure 4: KFUPM Distributed Honeynet.....	42
Figure 5: PAS Graph For a Single User.....	46
Figure 6: The Main Classes of Organizational Factors by Greitzer <i>et al.</i> [93].....	53
Figure 7: Research Methodology .....	74
Figure 8: Data set Naming Process .....	78
Figure 9: Data set Structure .....	78
Figure 10: ECG Data Representation.....	79
Figure 11: Schematic Diagram of The Proposed System .....	81
Figure 12: Wearable ECG Sensors .....	82
Figure 13: Transforming EEG From Time Domain To Frequency Domain .....	83
Figure 14: The Experimental Lab Environment .....	91
Figure 15: NeuroSky MindWave Headset.....	96
Figure 16: The Wild Divine Device.....	97
Figure 17: Frequency Spectrum of Normal EEG.....	100
Figure 18: R Peak To R Peak Interval of ECG [147] .....	101
Figure 19: Heart Rate Variability (HRV) [149].....	102
Figure 20: The Main Diagram of The Features Extraction Process.....	104
Figure 21: EEG Feature Frame .....	108
Figure 22: Feature Frame.....	108
Figure 23: Random Forest Procedures.....	111
Figure 24: Selecting The Best Hyperplane .....	112
Figure 25: Neural Network Structure.....	114
Figure 26: One Tail t-Test.....	118
Figure 27: Testing The Validity of The Null Hypotheses .....	125
Figure 28: Results Presentation .....	126

Figure 29: Generating a Group of Frames.....	127
Figure 30: Features for Comparison.....	128
Figure 31: Classification Accuracy using RF.....	129
Figure 32: Accuracy of Proposed Approach in Detail .....	129
Figure 33: Impact of Features on the Results .....	130
Figure 34: Effect of AD Factor .....	131
Figure 35: Effect of AD Factor in Details .....	131
Figure 36: Accuracy Comparison of Feature-frames .....	132
Figure 37: Accuracy Comparison of Feature-frames using Scatter Chart.....	133
Figure 38: EEG Frequency Bands During Normal and Malicious Acts .....	134
Figure 39: Accuracy of The Proposed EEG And Raw EEG Data Using RF .....	134
Figure 40: Scatter Chart of Accuracy for The Proposed EEG and Raw EEG.....	135
Figure 41: Accuracy of The Proposed EEG+ECG and EEG Features Using RF .....	136
Figure 42: Scatter Chart of Participant' Accuracy using EEG+ECG .....	137
Figure 43: (a,b,c,d,e,f): Evaluating the Results using Several Metrics .....	141
Figure 44: Accuracy of proposed EEG+ECG method using three classifiers.....	144
Figure 45: Accuracy of proposed EEG method using three classifiers .....	145
Figure 46: Comparing EEG+ECG Accuracy of AVG, Median, and STD.....	147
Figure 47: Percent incorrect with Different Size of Data.....	148
Figure 48: Incorrect Data of the Proposed EEG+ECG and the EEG Only .....	149
Figure 49: ROC Curves and the Training Time per Second Using RF.....	150
Figure 50: ROC Curves and the Training Time per Second Using SVM .....	151
Figure 51: ROC Curves and the Training Time per Second Using NN .....	151
Figure 52: Accuracy per participant .....	154
Figure 53: Feature Frame of Suh ans Yim Approach.....	155
Figure 54: Comparing proposed approach with Suh's Method using Accuracy.....	156
Figure 55: Comparing proposed with Suh's Method using Training Time .....	156

## LIST OF ABBREVIATIONS

<b>EEG</b>	:	Electroencephalogram
<b>ECG</b>	:	Electrocardiogram
<b>EEG+ECG</b>	:	A Combination of EEG and ECG Features
<b>SVM</b>	:	Support Vector Machine Classifier
<b>RF</b>	:	Random Forest Classifier
<b>NN</b>	:	Backpropagation Neural Network Classifier
<b>TP</b>	:	True Positive (Detected Malicious Signals)
<b>TN</b>	:	True Negative (Detected Normal Signals)
<b>FP</b>	:	False Positive (Undetected Malicious Signals)
<b>FN</b>	:	False Negative (Undetected Normal Signals)



## ABSTRACT

Full Name : Azzat Ahmed Ali AL-Sadi  
Thesis Title : TOWARDS AN EFFECTIVE APPROACH OF INSIDER ATTACKS  
DETECTION USING THE HUMAN PHYSIOLOGICAL SIGNALS  
Major Field : Computer Science and Engineering  
Date of Degree : November 2018

**CONTEXT:** Insider threats are among the most serious security concerns for organizations because of their catastrophic consequences on the organization's revenues and reputation. Several approaches have been proposed to detect the insider threats and mitigate their risk. However, discovering such attacks is a very challenging task because of the difficulty of distinguishing between the normal and the malicious activities conducted by trusted users. Insider threats are committed by people with enhanced knowledge about the organization's security mechanisms, such as employees and trusted partners who have authorized access to the digital systems.

**OBJECTIVES:** The primary objective of this study is to detect the insider attacks before it causes catastrophic damage to the organization system. Furthermore, providing a data set of bio-signals that collected during real insider threats to develop and extend the research in this area.

**METHODS:** We propose an approach that utilizes a combination of the human brain activities and the electrocardiogram (ECG) to identify the malicious acts of a trusted employee. The approach compares the normal and malicious patterns of the human bio-signals to recognize the insider threats. The experimental scenarios for collecting the brain activities and ECG signals were carefully designed based on physiological considerations,

without affecting the participants' decision to commit the attacks. The wearable devices were utilized to collect the bio-signals because of their benefits such as being small, easy to connect to a computer, comfortable, and cheap.

**RESULTS:** Eighty-four participants were included in this study. The achieved results illustrate that the proposed approach can detect the malicious threats with an average accuracy up to 98.4%. The results were evaluated further using several metrics to achieve high credibility and great confidence. In addition, this research provides a data set of the bio-signals collected from a wide range of participants for further research in this area.

**CONCLUSION:** The proposed approach can accurately identify the malicious activities even if the amount of the incoming data is too small. Thus, it will help organizations to detect the insider attackers and to take the necessary actions to mitigate the risk of such devastating attacks.

## ملخص الرسالة

الاسم الكامل: عزت أحمد علي السعدي

عنوان الرسالة: نحو أسلوب فعال للكشف عن الانتهاكات الداخلية باستخدام الإشارات الحيوية للإنسان

التخصص: علوم وهندسة الحاسب الآلي

تاريخ الدرجة العلمية: نوفمبر 2018

تعتبر التهديدات الداخلية من بين أكثر المخاوف الأمنية خطورة للمنظمات بسبب عواقبها الكارثية على إيرادات المنظمة وسمعتها. وقد أقرت العديد من الأساليب للكشف عن التهديدات الداخلية والتخفيف من خطورتها. ومع ذلك، فإن اكتشاف هذه الهجمات يعتبر مهمة معقدة للغاية بسبب صعوبة التمييز بين الأنشطة العادية والنشاطات الضارة التي يقوم بها المستخدمون الموثوقون. حيث يتم تنفيذ التهديدات الداخلية من قبل أشخاص لديهم معرفة جيدة حول آليات أمان المؤسسة، مثل الموظفين والشركاء الموثوق بهم الذين لديهم حق الوصول إلى الأنظمة الرقمية للمؤسسة. تهدف هذه الدراسة للكشف عن الهجمات الداخلية قبل أن تتسبب في أضرار كارثية للمؤسسات. وعلاوة على ذلك، إثراء وتطوير نطاق البحث في هذا المجال من خلال توفير قاعدة بيانات من الإشارات الحيوية التي تم جمعها خلال تهديدات داخلية حقيقية. ولتحقيق هذه الأهداف تقترح هذه الدراسة نهجًا يستخدم مزيجًا من أنشطة الدماغ البشري (EEG) وتخطيط القلب (ECG) لتحديد الأعمال الخبيثة لموظف موثوق به. يقارن النهج المقترح بين الأنماط الطبيعية والخبيثة للإشارات الحيوية البشرية للتعرف على التهديدات الداخلية. للحصول على الإشارات الحيوية المستخدمة في هذه الدراسة تم تصميم السيناريوهات التجريبية لتجميع أنشطة الدماغ وإشارات تخطيط القلب بعناية، مع مراعاة الاعتبارات الفيزيولوجية، ودون التأثير على قرار المشاركين بارتكاب الهجمات الداخلية. تم استخدام الأجهزة القابلة للارتداء لجمع الإشارات الحيوية بسبب فوائدها مثل كونها صغيرة وسهلة التوصيل بجهاز كمبيوتر، مريحة، ورخيصة. تشمل هذه الدراسة مشاركة أربعة وثمانون متطوعًا. وتوضح النتائج التي تم تحقيقها أن الطريقة المقترحة يمكنها اكتشاف التهديدات الخبيثة بمتوسط دقة يصل إلى 98.4%. تم اختبار النهج المقترح وتقييم النتائج بشكل أكبر باستخدام عدة مقاييس لتحقيق مصداقية عالية وثقة كبيرة. يمكن للنهج المقترح تحديد الأنشطة الضارة بدقة عالية حتى إذا كان حجم البيانات الواردة صغيرًا جدًا. مما يساعد المنظمات على الكشف عن المهاجمين واتخاذ الإجراءات اللازمة للتخفيف من مخاطر هذه الهجمات المدمرة.



# CHAPTER 1

## INTRODUCTION

### 1.1 Overview

Currently, information is the new world currency: money has value and so too does information. Multinational corporations, financial institutions, military organizations, and even small companies go to great lengths to protect their privacy and security from attacks. When a single computer is compromised in the organization, all other computers are vulnerable to attack [1]. Although traditional security methods are commonly used to protect computers and networks from attacks or unauthorized intrusions, these standard methods cannot prevent modern sophisticated insider attacks or initiate alerts to malicious insider activity. Several solutions have been developed for protecting organizations from outsider attacks. Among these solutions are data encryption, intrusion detection and prevention systems (IDS/IPS), and firewalls [2, 3]. However, attacks do not just come from outside; most of the harmful attacks occur from inside the organization, where the trusted employees can compromise the organization's security.

Insider attacks or insider threats become a significant security concern to organizations since differentiating between these crimes and non-malicious activities is difficult. Insider threats are committed by insiders who have more knowledge than outsiders do about the organization's system and its security mechanisms [4]. There are two categories of insider

attacks: the first category occurs when insiders accidentally or without malicious intent cause harm to the organization's security, whereas second-category insider attacks are malicious and happen when insiders use their privileges intentionally to attack the organization's security [5].

The scope of this research is to protect organizations from the second category of insider threats. This research aims to reduce the risk of intentional insider attacks by utilizing bio-signals associated with human behavior in addition to machine-learning classifiers. This research includes conducting experimental scenarios on a segment of volunteers to discriminate between malicious and normal activities by distinguishing their bio-signals.

## **1.2 Problem Statement and Motivation**

When most people think of attacks, they picture criminals and hackers trying to break into a network from outside the organization. However, they do not realize that some of the biggest threats are already inside. Insiders are employees and trusted partners with authorized access to digital systems and information. A recent survey on insider threats conducted by SpectorSoft on 355 IT professionals reveals that almost all organizations have experienced at least one insider attack; moreover, around 75% of all insider crimes are undetected. The survey goes further, describing the state of internal attacks as unlikely to improve soon, and pointed out that the total losses of organizations amounted to around \$2.9 trillion globally per year as a result of employee fraud, while the US losses amounted to around \$40 billion because of such insider fraud [6].

Insider attacks have become a significant security concern to financial institutions, organizations, and small companies. Several incidents of insider attacks targeting

governments, organizations, and even universities [7, 8] have been reported. The trusted nature of the insider access means that data breaches are largely undetectable by standard cybersecurity measures such as antivirus programs, firewall filtering, blocking systems, and disk encryption.

Insider attacks cause extensive damage to organizations. Thinking about the added cost of the data breach is even more disconcerting and brings a bigger financial burden to an organization. The added costs come from a variety of sources, not just the financial loss of that information but also responding to that incident, fixing all damages, and installing preventative systems. Moreover, there are numerous tangible costs such as the loss of customer loyalty. Therefore, governments and organizations invest money and enact laws to reduce the impact of such attacks.

Currently, several solutions have been used by organizations to mitigate the risk of insider attacks, such as monitoring employees' behaviors or applying signature-based solutions. Monitoring employees' behaviors (i.e., anger or revenge) is based on human experience to distinguish such behaviors [9]. Thus, this solution does not provide accurate results, and the organization could be deceived since human behaviors, which are bound to emotions, are not very clear and are difficult to reveal. In Greitzer's model [9] or verbal behavior [10] to recognize a potential risk could be deceived by individuals. On the other hand, signature-based solutions allow organizations to act against the insider attack—but after the incident [11]. When the incident occurs, organizations attempt to identify the signatures of insider attacks and develop a mechanism or policy to prevent the reoccurrence of this incident. This solution can protect organizations from the same attacks only and does not provide impregnable protection against different mechanisms of insider attacks. A new insider

attack could seriously damage the organization's data and cost them substantial financial losses as the rapid development of technology supports attackers to develop new and sophisticated methods of insider attacks.

Therefore, different and new solutions to insider attacks are urgently needed. The ongoing optimal solution for addressing the insider threat issue requires new tools and new thinking. Being able to detect an insider attack before it causes catastrophic losses is far better than being attacked and then building a defense system using policies and signatures to prevent similar incidents from happening again. So, the insider attack detection system must be based on fixed measures present in each attacker. No matter how different the techniques used by the attacker, these measurements remain constant as the attacker cannot control them. In this research, we target human physiological involuntary signals to distinguish between the employee's normal and suspicious activities, thus detecting insider attacks. Physiological signals are spontaneous patterns that give indications of emotions and bad intentions [12]. Therefore, the difference between the physiological patterns of normal and suspicious acts would well indicate insider threats without human intervention or even knowledge of the attack mechanism. The reasons that prompted us to utilize human physiology in detecting insider attacks are as follows:

### **1.2.1 The Ability of Physiological Signals to Identify Human Moods and Intentions**

Human physiological signals are spontaneous signals—i.e., signals done without will or self-control. Therefore, imitating these signals is extremely difficult. Physiological signals occur inside the human body and nervous system, where these processes are measured to assess bodily functions. Physiological signals are constantly changing in response to



changes in the outer environment of the human body and also in response to human emotions and thoughts. Physiological signals such as blood pressure, heart rate, temperature, and brain signals change themselves to find the optimal balance to human physiological states based on feedback from the human body's built-in sensors [12, 13].

Human behaviors are usually accompanied by feelings and psychological changes such as anger, stress, and fear. Many researchers are keenly interested in studying the relationship between human behaviors and nervous system feedback. The process of monitoring bio-signals and neuro-signals is called biofeedback or neurofeedback, which enables us to associate physiological signals with human feelings and behaviors.

Moreover, biofeedback and neurofeedback have been used by researchers in several systems such as emotion recognition, intention detection [14], cryptographic systems (cryptographic key) [15], detecting read book genres [16], control systems, and crime detection systems such as a polygraph or lie detector [17].

### **1.2.2 Ease of Physiological Data Collection and Analyzing**

Human bodies radiate data loudly, continuously, and individually. Typically, clinics and hospitals have the appropriate equipment to listen to that data, and these devices are expensive, large, and difficult to use. However, with the advent of wearable sensors, we now deal with gadgets and trackers that can collect physiological data and allow us to analyze these data in real time.

Wearable sensors are small devices of varying shapes that fit on different locations on the body. These sensors can collect several types of human data, such as brain waves through electroencephalography (EEG), heart rate through electrocardiography (ECG), body

temperature, and skin conductance. These devices can be easily connected to the computer or to the smartphone to transfer the collected data. Moreover, the ease of use and low costs led to the widespread use of these devices [18, 19].

Physiological readings resulting from wearable sensors have become very important for individual health care. Several people have these devices. Recently, some third-party companies have increased the importance of collecting data via wearable sensors to the employers because these data could eventually affect the health insurance payments of their employees. Wearable sensors could be utilized by companies that want to lower their health-care bills. Some companies like Fitbit have begun selling their devices in bulk to employers; for instance, Autodesk sells discounted Fitbit devices to their employees, with the idea to encourage healthier behavior and a happier workforce [20].

We are motivated by all these features of modern techniques in measuring and collecting human physiological signals, which would give quick and accurate measurements of biofeedback and neurofeedback of the human body. Moreover, these devices have already been used in some companies [20], which increases the acceptance of using these devices to determine insider attacks.

### **1.3 Research Objectives**

To mitigate the catastrophic influences of insider attacks by providing a solution to detect such threats, we have developed the following main research objective and a set of objectives that seek to achieve it.

**Main Objective:** To explore and study the potential of using physiological signals to detect insider attacks.

The potential of using physiological signals to detect insider attacks is investigated by determining the ability of these signals to differentiate between normal and suspicious activities. To address the main objective challenge, we support the research with the following objectives:

**Objective 1:** To conduct a survey of the existing insider threat detection approaches.

This objective promotes research through knowledge regarding previous approaches, which will give a clear idea of the previous systems that were used to detect insider threats.

The objective supports the further verification of the quality of the proposed system.

**Objective 2:** To create a data set of physiological signals and make this data set available for further research.

To the best of our knowledge, there is no available data set containing human physiological signals that were collected during suspicious and normal acts. Therefore, objective 2 is essential for achieving the main objective because the existence of such a data set supports the ability to conduct research experiments on the proposed system. In addition, the availability of the data set for research use would provide an opportunity to develop and extend the research in this area. The data set would contain human physiological signals that were collected during normal and malicious activities.

**Objective 3:** To characterize changes in some physiological activities that might lead to the detection of insider malicious activities.

If the physiological signals provide certified results in the determination of insider attacks, certainly not all physiological signals are equal in the determination of such attacks. Therefore, highlighting signals that play a significant role in identifying internal violations as well as clarifying the changes in these signals will be directed at improving the proposed system's mechanism for detecting attacks.

**Objective 4:** To propose and design a continuous monitoring system that might help in the detection of insider threats.

This objective provides the design of the proposed system for combatting insider attacks. The main research objective will be taken into consideration as the design of the proposed system will depend on the human physiological signals collected but not on the monitoring of user actions in the network (for instance, the monitoring of the user's log files).

**Objective 5:** To evaluate the potential of using the proposed system for detecting insider attacks using physiological activities.

Objective 5 explores the potential of using the proposed approach to reduce the damage of internal attacks by identifying these attacks using human physiological signals. This objective will assess the effectiveness of the proposed approach in distinguishing between the physiological signals of humans during normal and suspicious activities. Furthermore, this objective will clarify the ease and possibility of using the proposed system by organizations to act as a line of defense against insider attacks.

## 1.4 Research Questions

To fulfill the above research objectives and to provide a feasible solution for detecting insider attacks, we have addressed the following research questions:

- 1) What is the potential of using physiological signals to detect insider attacks?

The first research question investigates the ability to use human involuntary signals to detect insider threats. This research question raises other inquiries such as what physiological signals should be available to conduct this research and, in case no data set that contains such signals is available, how experiments should be conducted to collect such signals.

- 2) What are the experimental scenarios that will be followed for collecting data and building the signal database?

To the best of our knowledge, no available data set contains physiological signals collected during normal and suspicious activities. So, the signals should be collected from volunteers during different experiments. RQ2 explores how to collect signals and what possible experimental scenarios will be used to assemble these signals.

- 3) Which physiological signals are most important in detecting insider attacks from the selected signals in the experiments?

RQ3 investigates whether all signals have an equal ability to identify internal attacks—in other words, whether all the signals collected during the experiments would be used in the proposed approach. Moreover, this research question determines the format of the signals that would help increase the accuracy of detecting insider threats and inspects the

relationship between changes in the collected signals and the identification of malicious activities.

- 4) How can a practically robust monitoring system for the identification of insider attacks be designed?

This research question investigates a critical point in the proposed system for detecting the insider threats: how the system should be designed. The design of the proposed system should determine the format of input signals and the output results that will clearly indicate insider attacks.

- 5) Is the monitoring system practically robust in terms of identifying insider attacks?

RQ5 investigates the quality of the proposed system for detecting insider attacks and discusses how the proposed system will be evaluated as well as what criteria are used for evaluation and how accurate the system is.

## 1.5 Summary of Research Contribution

Conducting an insider attack is risky and, in many cases, has catastrophic consequences on an organization's security as well as its financial resources. To address the vulnerability of malicious insider threats, this research tackles the potential of using human physiological signals to detect such insider attacks. In particular, this work reports the design of a new approach that based on extracting **a new set of features** from biofeedback and neurofeedback. Instead of monitoring human-controlled behaviors, we emphasize on monitoring hard-to-imitate human involuntary signals. In the proposed approach, we utilize electroencephalography (EEG) and electrocardiography (ECG) data of the human body to

distinguish between the malicious and normal activities conducted by employees. The proposed approach converts the collected EEG and ECG signals into features to increase the efficiency of these signals to detect insider attacks, and the extracted features showed promising results.

Additionally, the proposed approach was built on the identity theory, which states that trusting a user's identity is one of the main weaknesses of the system and could leave a system vulnerable to insider attacks [21]. The proposed system does not rely on the user's identity, but it acts as a line of defense against the authorized user's attacks. Thus, the authorized user may be prevented from using the system if he or she is suspected of performing malicious activities.

The proposed set of features is automatic and does not rely on human intervention or human experience to detect insider threats and utilizes machine learning to classify the malicious activities. Machine learning is a field of artificial intelligence that seeks to create predictive models and algorithms, giving computers the ability to build enough experience to carry out tasks without being explicitly programmed [22, 23].

The overall evaluation of results shows that the extracted features from EEG and ECG signals can correctly distinguish between a user's malicious and benign activities. The results also ensured that the proposed approach would produce accurate results despite the period difference between normal and suspicious activities. In other words, the proposed approach detects an insider attack using a small period of malicious activities conducted by the attacker; thus, the proposed approach can detect the insider attack before any serious damage is done to the organization.

This research provides a data set that contains the physiological signals for 84 volunteers. Experiments were conducted to collect the physiological signals of the volunteers during their normal and suspicious processes. The proposed experimental scenarios for collecting data simulate as much as possible real-life insider attacks.

Our research can help mitigate the risk of insider attacks targeting organizations. Regardless of the insider attack mechanisms, the proposed system helps address such insider threats. This will help organizations protect their privacy and provide security against attacks. In addition, the human physiological data set collected during normal and malicious activities for research use would provide an opportunity to develop and extend the research in this area.

## **1.6 Thesis Roadmap**

The remaining chapters of this thesis are organized as follows. The glimpse of the research background is presented in chapter 2. Chapter 3 surveys the related work of insider attacks. Chapter 4 describes the research methodology, where the research challenges are addressed, and the proposed approach is discussed. Chapter 5 describes the experimental scenarios for collecting data and presents the experimental devices. It also discusses the data set creation and characterizes the process of extracted features from the collected signals. Chapter 6 introduces the evaluation criteria and presents and discusses the experimental results. It also highlights the evaluation of the proposed approach using several comparisons. Finally, chapter 7 concludes this thesis and presents the future work.



## **1.7 Summary**

Information has substantial value for large organizations or even small companies. Chapter one presents an overview of the importance of information security for organizations and illustrates the risk of insider attacks over those that occur from outside the organization. Moreover, this chapter describes the two categories of insider attacks, namely, accidental and malicious. It also defines the scope of this research, which tackles the second category of insider threats.

The research problem, which is detecting insider threats by utilizing human bio-signals, is presented. To detect insider attackers, normal and malicious activities must be distinguished first. To address this problem, the main objective of this research is divided into five supporting objectives. These five objectives are integrated to solve the research problem. To fulfill the research objectives, five research questions are developed and discussed in this chapter.

Furthermore, the research motivation is presented and divided into two main parts, namely, utilizing the advantage of human bio-signals' ability to identify the human mood that may help to distinguish malicious activities and utilizing the advantage of modern technologies in capturing and analyzing human bio-signals. Moreover, the research contribution, which aimed at mitigating the risk of insider attacks that target organizations, is discussed. Finally, the roadmap of this research is demonstrated.

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Cybersecurity**

Cybersecurity has been known by several names, such as data security, IT security, and computer security. Regardless of the definition, the information stored on computers is almost always worth more than the computers themselves. Cybersecurity is the protection of computer systems, programs, networks, and data from undesirable behaviors of attackers under different circumstances and is an important issue for organizations because damaged security systems may cost significant financial losses. The organization's security requires constant development to provide protection from several modern attacks [24].

The scope of cybersecurity is wide, growing, and constantly changing as a result of the development of new attack methods. Generally, cybersecurity aims at protecting organizations' invaluable data, such as assets where the organizations' information security efforts should be focused. These assets can be divided into three types: data, software, and hardware assets. Data assets have the greatest value over other assets and include but are not limited to databases, the organization's files, and the information that the organization generates daily, whereas software assets include mobile applications, programs, and operating systems. On the other hand, hardware assets include computers, communication channels, networks, and mobile devices that belong to the organization's employees [25]. Cybersecurity can be divided into the following domains:

### **2.1.1 Access Control**

Controlling the access to the organization's assets or protecting the organization's resources from unauthorized disclosure is one of the important cybersecurity domains. Access control is the process of rejecting or accepting a specific request to enter physical facilities and/or using information services. Sometimes, there is a confusion between the term access with the authorization and authentication. Access is the ability to reach, read and modify the resources. On the other hand, the authorization is the rights and permissions to use the resources, whereas the authentication is the process of identifying and proving the user who he claims to be, using different methods such as username and password [26].

### **2.1.2 Software Development Security**

Software Development Security is the process of embedding the principles of software security through the software development life-cycle. As the best practices of software development, embedding the security in the early stages of software development leads to ensuring the software quality [27].

### **2.1.3 Business Continuity and Disaster Recovery Planning**

This domain mainly focuses on business; it does not concentrate on the problem of data violation or unauthorized access. This domain aims at providing emergency plans to ensure the business continuity in the event of a disaster. It consists of two sub-domains, which are the Business Continuity Planning (BCP) and the Disaster Recovery Planning (DRP). For organizations, the BCP requires a comprehensive approach to ensure the continuity of the organization's business even after the occurrence of disasters, such as a natural disaster and even power outages. On the other hand, the DRP deals with the procedures of how the organization can resume its business after major disruptions. However, both sub-domains

have several common considerations include the development, testing, updating of the essential actions for protecting the critical processes of the organization's business from being influenced by a disruptive event such as a network failure [28].

#### **2.1.4 Cryptography**

Cryptography is protecting the stored and transmitted data from being understood or read by unauthorized parties. Encryption is the process of transforming the plaintext into ciphertext using several methods such as transposition and substitutions. On the other hand, the process of returning the plaintext is called the decryption. The strength of encryption depends on the algorithm and the key. Based on the key, there are two main types of cryptography, which are symmetric and asymmetric cryptography. The symmetric cryptography system utilizes the same key for encrypting and decrypting the information such as the Data Encryption Standard (DES) algorithm, whereas the asymmetric cryptography system utilizes a pair of keys which are called public and private keys. The public key is used for enciphering the plaintext, and the private key is used for deciphering the ciphertext. A common example of asymmetric cryptography system is the RSA algorithm [29].

#### **2.1.5 Information security and risk management**

Information security and risk management is an important cybersecurity domain that focuses on identifying data assets, risk management, and mitigation. Also, this domain includes the organizational structures, the importance of security awareness training, as well as the development of standards, procedures and guidelines to address the confidentiality, integrity, and availability of information system [30].

### **2.1.6 Law, Investigation, and Ethics**

Cybersecurity crimes have become very dangerous to organizations as well as individuals. We often hear about internal and external attacks that cost organizations billions of dollars. Statistics show that one of the main motivations of cybercrimes is the financial gain, as we will discuss later in this Chapter. The stolen information can be resold in a black market for a financial gain. Therefore, the public and private sectors have worked together to establish rules and regulations about cybersecurity crimes, and how to deal with the perpetrators. This domain addresses ethics and compliance with various regulatory frameworks as well as the understanding of the laws that associated with the cybercrimes and the liabilities to these laws. Also, this domain focuses on the basics of conducting investigations to determine if a crime has occurred, as well as the evidence gathering techniques [31].

### **2.1.7 Operations Security**

Security operations domain focuses on identifying the critical data and the execution of specific measures that eliminate or mitigate the risk of the adversary attacks on the information system. This domain describes the controls over the resources (such as hardware and media) that needs to ensure the security, as well as the definition of the operators with access privileges to any of these resources [32].

### **2.1.8 Physical and Environmental Security**

This domain addresses the problem of securing the physical environment that affects the confidentiality, integrity, and availability of the organization's information system. This domain examines the infrastructure and physical environment around the information

system against physical threats such as natural disasters, emergencies, sabotage, and even the electric power issues. The physical security includes alarms, guards, and locks [33].

### **2.1.9 Security Architecture and Design**

This domain essentially concentrates on securing the hardware, software, and operating system of the organization. This domain covers several topics such as the evaluation criteria, the distributed environment security issues, as well as the security models that provide the framework for ensuring the organization's security policies. Examples of security models that assist in designing a system to protect the organization's assets are the Role-Based Access Control (RBAC) and Mandatory Access Control (MAC) [34]. Furthermore, an example of evaluation criteria is the Software Engineering Institute Capability Maturity Model Integration (SEI-CMMI) [35].

### **2.1.10 Telecommunications and Network Security**

This domain tackles the problem of securing the transmitted information over the private and public communication networks. It includes the structures, transmission formats and transport methods of the communication networks. This domain involves protecting transmitted data, detection and correction of transmission errors, intrusion detection and response, network attacks and abuses as well as different network protocols such as connection-oriented and connectionless protocols. It is considered as the largest and most comprehensive domain of cybersecurity. The Open Systems Interconnect (OSI) model is an important area of this domain which was developed to assist the vendors in developing interoperable network devices. The OSI model consists of seven layers that describe the procedures of how the applications can communicate over the network [1].

## 2.2 Classics System Security Characteristics

Every secure information system should satisfy three classic security characteristics; breaches to these characteristics are considered undesirable behavior. These characteristics are confidentiality, integrity, and availability. Confidentiality ensures that only authorized parties who have sufficient privileges may edit or view the information. The most common tool used to achieve confidentiality is encryption. Integrity, probably more critical than either confidentiality or availability, ensures that the stored data on organization resources is correct and unaltered by unauthorized parties or malicious applications. Measures to protect integrity include error-checking methods such as checksums and file hashing. Availability means that network resources are readily available to authorized users. Although a secure computer must prevent access attempts by unauthorized users, it still must allow immediate access to authorized users; for instance, a banking customer should be able to check their balance or withdraw their funds effortlessly [36].

Violations of the security properties can occur during insider and outsider attacks [37]. Despite numerous reports of destructive outsider attacks, both accidental and malicious insider attacks put a lot of corporate data at risk. Predominantly, organizations do not know how much data they have at risk. The inside attacker has one or more of the following advantages: authorized system access, knowledge about the organization's system, the ability to reveal the organization system vulnerabilities to outsiders, and trust by the organization [21, 38]. The Insider Threat Center of Computer Emergency Response Team (CERT) defines the malicious insider threat as:

*"A malicious insider threat is a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems" [39].*

## **2.3 Types of Insider Attacks**

According to the CERT Insider Threat Center, there are four primary types of insider attacks, classified based on the similarity of attack patterns analyzed among more than seven hundred real insider attack cases: fraud, IT sabotage, insider theft of intellectual property (IP), and espionage [39]. Figure 1 shows the CERT breakdown of intentional insider attacks targeting the United States.

### **2.3.1 Fraud**

Fraud usually targets financial services and is one of the crime types where the attacker maliciously accesses information, stealing credit card data or changing the data for financial gain. This type of malicious crime could be committed even by employees who have low-level access to the organization data or even customers. Fraud may continue for a considerable period before it is identified [39].

An example of fraud includes the fraud case that occurred at the military contractor's office where a member of the computer help desk team took advantage of his position for creating a fake e-mail address on the military system. He fraudulently requested replacement parts of equipment from the vendor using the fake e-mail address. The vendor expected that the original parts of equipment would be restored later after the replacement parts were sent.



The fraudulent member put his home address for receiving the shipments. The fraud was successful, with shipments being sent to the employee for more than twenty months. The employee received five hundred products at a cost of around \$8 million and sold around ninety products on the Internet for more than \$0.5 million. The fraudulent employee was convicted and sentenced to four years imprisonment and ordered to pay the amount of \$8 million to the vendor [39].

### **2.3.2 IT Sabotage**

Usually, the main purpose behind IT sabotage crime is revenge. For instance, disgruntled workers cause damage to IT systems or destroy data. Unlike fraud, IT sabotage is committed by insiders who have high-level access to the organization's system, such as database administrators, network managers, and system programmers. The preparation of IT sabotage is typically done while the attacker is still working at the organization, but the execution of the IT sabotage may occur after the employment termination [39].

A case of insider IT sabotage is presented in [39]. Around one thousand files related to employee compensation were deleted from an organization by a disgruntled former employee of a human resources department. After the employee had been dismissed from the organization, he broke into the organization's systems remotely using his previous privileges. To implicate another person in this crime, the attacker modified the payroll records of his former coworker. He increased the salary and added substantial bonus to her payroll records. The attacker also used the last name of this female coworker to send an e-mail to senior managers containing some parts from the deleted files. He was angry at this coworker because she rejected his previous romantic interest. The attacker was

convicted and sentenced to eighteen months of imprisonment and was ordered to pay more than \$90,000 as compensation.

### **2.3.3 Insider Theft of Intellectual Property**

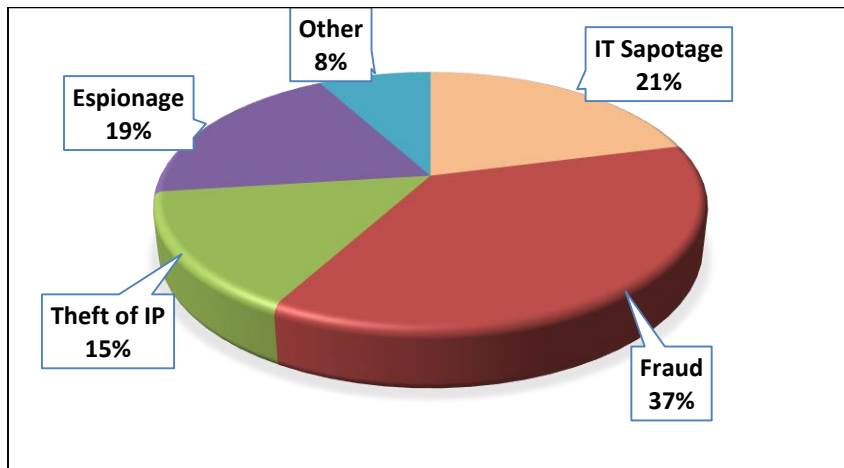
The aim of insider theft of intellectual property (IP) is to steal the IP using IT resources. IP theft includes stealing business plans and source codes. These crimes are usually committed by insiders who are aware of the IP value, such as scientists, programmers, and engineers. The main purpose behind IP theft is personal gain, where the attacker sells the stolen IP to a competitor company or utilizes the stolen IP for his own company [39]. For example, a government organization responsible for maintaining a reliable medical database was contracted formally with a programmer to help create their programs. Prior to the termination of the contract, the contractor was informed that his privileges to access the system had been disabled and his responsibilities were reduced. After these actions taken by the organization against him, the programmer resigned from the organization. However, before he quit the organization, he installed a back door into the system with administrator privileges. The attacker used the installed back door to attack the organization three times over two weeks to steal source codes and password files from the system. The organization was alerted by the large amount of remotely downloaded files. The downloaded files were traced. Then the attacker was convicted and sentenced to five months in jail in addition to paying around \$10,000 as compensation [39].

### **2.3.4 Espionage**

Espionage is the process of obtaining secret or confidential information without the permission of the information holder. Espionage may target governments, organizations,

and even individuals. But the term “espionage” is mostly related to spying on veritable enemies and is usually committed for military purposes [39, 40].

An example of espionage is the Robert Philip Hanssen case, known as possibly the worst intelligence disaster in US history. Hanssen was an FBI counterintelligence agent who began spying for the KGB in 1985. Using his privileges, he voluntarily passed highly classified national security and counterintelligence documents to Soviet intelligence officers in return for diamonds and large quantities of cash. As a counterintelligence agent, he could monitor the FBI’s surveillance of the KGB and lead investigators down false trails, which allowed him to continue leaking classified information for an extended period. He was discovered after a Soviet spy had switched over to the CIA. Hanssen was eventually caught, found guilty on fifteen charges of espionage, and sentenced to life without the possibility of parole [41, 42].



**Figure 1: The CERT Breakdown of Intentional Insider Crimes in United States**

## 2.4 Motivations behind Insider Attacks

Conducting an insider attack is risky and, in many cases, highly dangerous, so what would make an individual decide to commit such an action? Motivations for conducting insider attacks are highly diverse, but they can be classified into four basic categories, forming the acronym MICE: money, ideology, coercion, and ego [43–45]. An attacker may seek monetary payment if they face a large debt or simply have expensive tastes. Many have betrayed their organization or country for personal financial gain.

Ideology as a motivation can take various forms. One is political in nature, such as procommunism. In 2002, Ana Montes, an analyst with the Defense Intelligence Agency, was found guilty of spying for the Cuban government. She was recruited by the Cubans as a result of her disagreement with US policies toward Latin America and motivated by political ideology [46]. Another politically motivated insider attack penetrated the Greek cell phone provider Vodafone. A malicious software was injected into a phone switch to control the incoming and outgoing calls for specific numbers. The primary goal of this attack was to eavesdrop on the prime minister and prominent legislators. When exactly the malicious software was injected and what information was leaked are unknown. The attack was accidentally discovered in 2005, when the malicious software was incorrectly updated after the beginning of tapping. The incorrect update conflicted with other system processes and initiated an alarm. The attack was reported as an insider threat and was attributed to an employee with sufficient experience on the operating system of the cell phone switches [47].

Ideology can also take the form of fanatical convictions, such as extreme religious or anti-establishment beliefs or the idea that their actions are somehow helping people or an oppressed portion of society. For instance, the WikiLeaks leakage is a famous insider attack that targeted sensitive classified documents. WikiLeaks is a well-known journalistic website that leaked the confidential information of governments. Private Bradley Manning, who was the US Army intelligence analyst, violated US information security and had legitimate access to a secure network belonging to the US Department of Defense. Over 250,000 secret US embassy cables were leaked and passed to WikiLeaks. To conduct this attack, Manning utilized his authority to access a computer with a writable drive. He smuggled the data out on a rewritable CD (music CD) [48]. Manning was depressed by US counterterrorism operations in Iraq. “Manning said he’d sought to make the world a better place” [49].

Coercion, commonly used as blackmail, is the forced participation of an insider in an attack against their will. They may not always be aware of their participation, making them an unwitting passive insider. An example of an insider attack through coercion is the Northern Bank headquarters robbery in Belfast, Ireland. Late in the evening of Sunday, December 19, 2004, armed and masked gang members arrived at the homes of two Northern Bank executives. Pretending to be police officers, the gang members entered the homes, taking both families as hostage at gunpoint. The following day, both executives were instructed to go to work as usual or risk the deaths of their families. That night, the armed men entered the underground vaults of the bank, where they took over twenty-six million in pounds sterling and smaller quantities of various other currencies, including US dollars and euros. Who committed the crime is still unclear, but all hostages survived the incident [50, 51].

Ego encompasses both personal and psychological motivations. Personal motivations are highly diverse and wildly unpredictable. They include anger, revenge, problems at work or home, and divided loyalties. Psychological motivations—including mental instability and sociopathic behavior, such as finding thrill or adventure in malicious acts—can be detected in some cases by preemployment testing and evaluations. However, this is not foolproof, and many cases are not detected. In 2001, a series of letters containing anthrax arrived at locations at several places in the United States. It took the US government years to figure out how this had happened [52]. Bruce Ivins was a scientist at Fort Detrick, the army’s biological defense labs, and he was deeply troubled; he had signed documents authorizing the organization to look at his medical records, but no one took care. Long before the anthrax attacks, Bruce’s psychiatrist thought that he was the most dangerous patient she had ever seen in her entire career. When the anthrax attacks occurred, even though she did not know he was working with biological agents, she immediately thought that Bruce was behind these attacks. Bruce sent so many red flags; he e-mailed his own staff, complaining about his increasing paranoia. Even though the staffers were afraid he was going to hurt them, nobody reported his case. Bruce’s case ended up in the newspaper with the headline “Paranoid man in charge of deadly anthrax” [53, 54].

As seen in many cases, usually, more than one underlying motivation is behind a malicious act. John Walker, a US Navy warrant officer, was found guilty in 1985 of passing one million classified messages to the Soviet Union over a seventeen-year period [37, 55]. John was known to have a major ego, driving him to take risks in life to prove his superior abilities, displaying sociopathic behavior. He also took great joy in engaging in dangerous acts, committing his first crime as a young boy and never looking back. Walker was also

paid large sums of cash, which he initially used to pay off debts but eventually used to supplement his insatiable spending habits [56]. He embodies the money and ego prongs of MICE; these dual categories make the detection of insiders that much more complex as no two insider motivations are the same.

Some insiders carefully plan out their intentions and deliberately take steps to put themselves in the best position to carry out these attacks, as seen with the Walker case. However, not all individuals fall into the role of malicious insider through long-term scheming. A major life change such as a divorce, unexpected debt, or the loss of a job may trigger an insider to commit an attack on a whim. When an insider's opportunities and their motivations align, the environment for an insider attempt is created. Because of the expansive nature of motivations behind insider attacks, combined with the multitude of insiders and sensitive positions, prescribing a formula to identify and stop all insider threats before they occur is difficult. However, steps such as data encryption, access management, and log monitoring can be taken to reduce the ability of insiders to commit these attacks and mitigate the damage they're capable of inflicting.

## **2.5 Impact of Insider Attacks**

According to the 2018 insider threat report from Securonix Security Analytics, based on an online survey of 472 cybersecurity professionals about the insider threats that faced their organizations, 66% of the organizations consider insider attacks more likely to happen than external attacks. Moreover, 44% perceive that outsider and insider attacks have the same impact, whereas 42% believe that insider attacks are more damaging than outsider attacks [57]. Figure 2, shows the vulnerable assets targeted by insider attacks.

Furthermore, the true cost of a successful insider attack is hard to determine; however, 27% of the organizations estimate the cost of each successful insider attack in range of \$100,000 to \$500,000, whereas 21% estimate the cost ranges between \$500,000 and \$2 million, and 9% believe that the cost exceeds \$2 million [57]. The cost is not limited to financial losses and may include that of the incident response, loss of reputation, and loss of customer loyalty—a catastrophic impact on organizations.

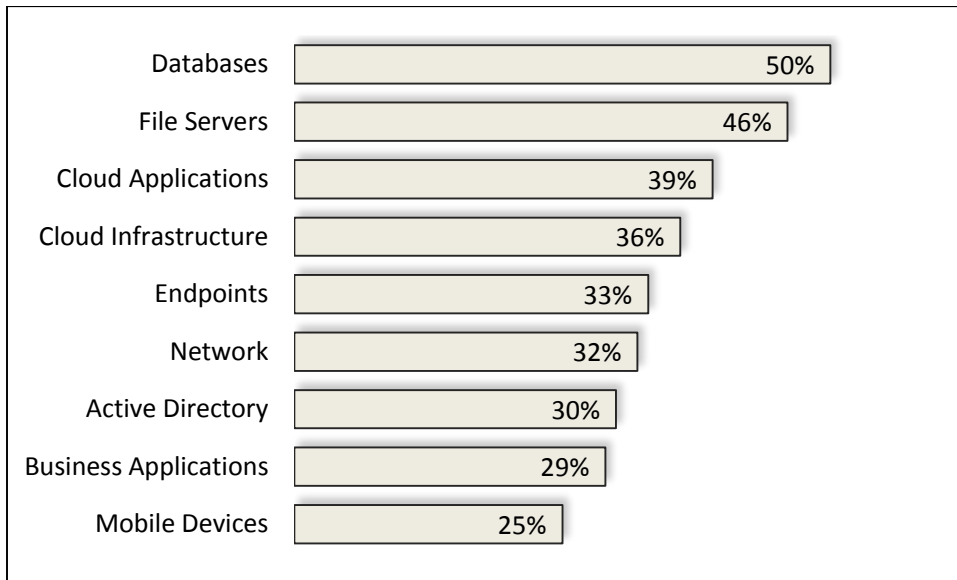


Figure 2: Vulnerable Assets For Insider Attacks

## 2.6 Insider Threat Detection: Existing Solutions

Although there are several security tools for detecting insider threats, as per the insider threat report, only 36% of the organizations have formal insider threat detection programs. Some existing solutions to insider attacks are data loss prevention, data encryption, identity and access management, monitoring users’ behavior, and endpoint security. Moreover, Securonix’s report revealed that 63% of the organizations use IDS or IPS to detect the insider attacks, and 62% utilize user log management, whereas 51% depend on the



available security information [57]. More details about the existing techniques for detecting insider attacks are presented in the literature survey.

Recently, a new research direction was established to distinguish between the benign and malicious activities of employees [18, 26]. The aim of the new research direction is to detect the malicious insider activities beforehand. Therefore, it can serve as an early detection system from the insider attacks. This direction of research is based on human physiological signals (i.e., human bio-signals) such as EEG and ECG signals. Physiological signals provide clear ideas about human emotions. EEG signals are brain waves—delta, alpha, beta, gamma, etc.—whereas ECG signals pertain to heart rate.

## **2.7 Summary**

Chapter two provides the research's background by discussing the cybersecurity and presenting the classic system security characteristics which are the confidentiality, the integrity, and the availability of the system. According to CERT, the formal definition for the insider threats has been presented. Also, the four types of insider threats have been discussed which are fraud, IT sabotage, IP theft, espionage. Several examples have been discussed to explain the four types insider threats.

Four motivations which are money, ideology coercion and ego behind each insider attack. Where the insider attack may be conducted for a single motive, which facilitates the identification of this attack by conventional methods. on the other hand, the detection of insiders who aim at more than one motivation is more complex. The four motivations of insider attacks have been discussed. The devastating impact of internal attacks on

companies has been explained, which include not only the financial losses, but also the cost to respond to the incident, loss of reputation, and the loss of customer loyalty.

The existing solutions of insider threats utilized by organizations have been presented. Presenting these solutions were based on a recent insider threat report that provided by Securonix. This report surveyed 472 cybersecurity professionals. The report revealed that most of organizations use intrusion detection and prevention, and the user log management techniques to detect the insider attacks.

## **CHAPTER 3**

### **LITERATURE REVIEW**

#### **3.1 Overview**

This chapter discusses the importance of increasing awareness of insider attacks and presents some research that aims to increase such awareness. This chapter additionally provides a categorization for existing strategies to mitigate the risk of insider threats. Each category will be discussed separately, and several methods from each category will be presented. The research gap in the field of insider attack detection will be discussed by presenting the limitations of each category, with emphasis on methods that utilize human bio-signals as they are in the area of this research.

#### **3.2 Awareness of Insider Attacks**

When you hear the word “hacker” or “attacker,” often, an external attacker comes to mind. Numerous officers in the cybersecurity field are more concerned about outsider threats because they perceive the enemy as outside the organization. Therefore, almost all the military, educational, and financial organizations have several mechanisms and defensive actions to defeat the outside attacker. However, the insider threat is considered as one of the most complicated situations to deal with in cybersecurity. According to an Association of Certified Fraud Examiners report, organizations and companies in the United States lost around 5% of their revenue because of fraudulent insider attacks. In addition, about two-

thirds of the malicious fraud cases faced by US organizations are conducted by insider employees [58]. An inside attacker has more privileges, facilities, and advantages than an outside attacker. Moreover, according to a report by the Intelligence and National Security Alliance (INSA), no training and qualification programs exist that can be used as a reliable framework to tackle insider attacks [59, 60]. For these reasons, raising the awareness of organizations about the insider threat risk is important. Consequently, some researchers and institutes tackle the problem of increasing awareness by developing strategies as well as providing training and courses about insider attacks.

For example, Ortiz *et al.* [61] described the necessary processes to develop a training environment for insider threat situations. Their study aims at ensuring that cognitive processing and some insiders' behaviors should be included in the training environment. They encouraged to utilize serious gaming, an area of game development for the purpose of training. Moreover, they encouraged the training designers to use three-dimensional game development tools, such as Unreal Engine, to simulate the insider threat training environment. Ortiz *et al.* [61] suggested that the development process of the training environment include some essentials, such as scenario narratives, artistic components, training components, and programmatic components. Furthermore, they proposed an insider threat training scenario that consists of two levels of conditions: the control condition and the insider threat condition.

Moreover, Chi *et al.* [62] from Florida University describe the guidelines for implementing an educational virtual lab that would enhance the knowledge and increase the security skills of trainers regarding insider threats. They implemented training modules utilizing the CyberCIEGE scenario development kit. CyberCIEGE is an educational security game

supported by some US governmental organizations, among them the US Navy and the Office of the Secretary of Defense. This video game allows students to design the environment of the corporate's network and customize several attack situations that could target this corporation, including physical and logical attacks [63]. The training module proposed by Chi *et al.* [62] includes three types of insider attacks: fraud, IT sabotage, and IP theft.

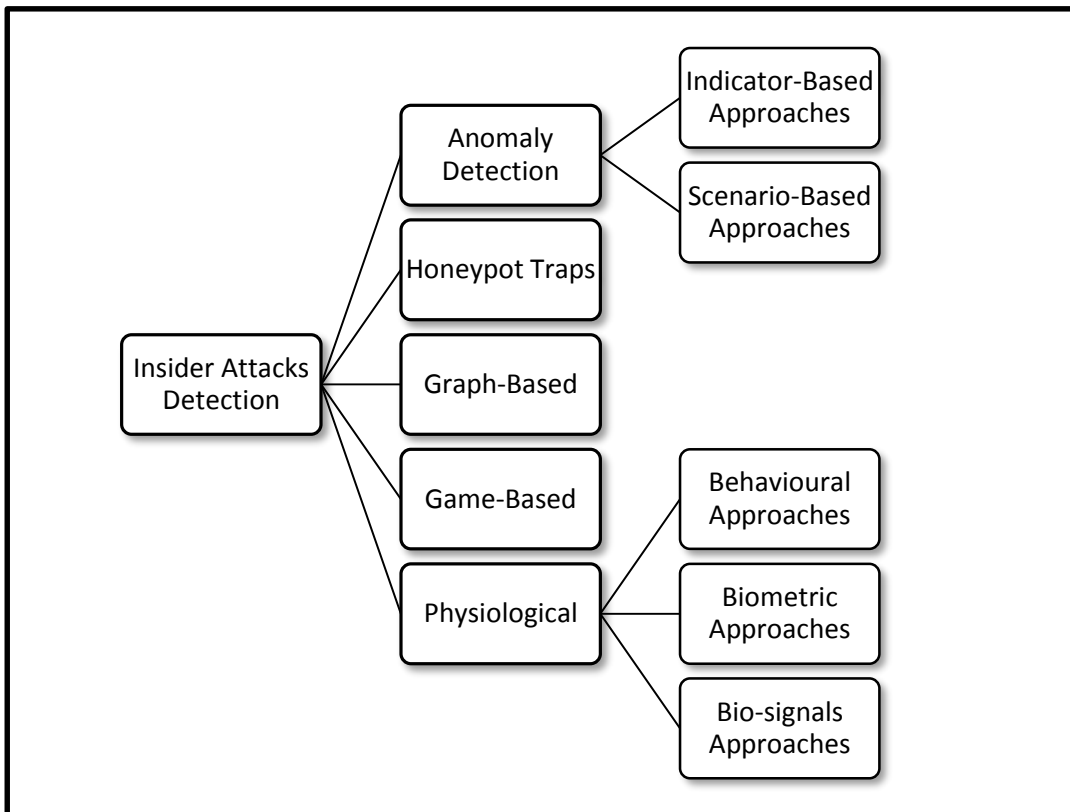
On the other hand, the Rochester Institute of Technology (RIT) offers a new security course for software developers. The course focuses on insider threats and their devastating consequences on security. This course was proposed as a practical activity for students. Students are divided into groups; each group consists of four to six students and is responsible for designing and implementing a reasonable-sized security system. However, one student in each group will play the role of an inside attacker and will try to design and implement some vulnerabilities in the system, such as leaving back doors in the source code to be utilized maliciously later. This student is informed secretly by e-mail to play the role of the inside attacker. The course gives a chance for software development students to be aware of insider threats [64].

### **3.3 Insider Attack Detection Methods**

Because of the major damage caused by insider attacks, detection and protection from these attacks became a necessity. To achieve this goal, researchers have proposed several methods to detect and relieve the risk of these attacks. Although researchers have utilized different sciences in their methods, the main goal of these methods is to detect and protect from insider threats. The conducted research so far aims to answer the following questions:

1. How the insider threat will be identified?
2. In case of insider attack, how to enhance defenses against such an attack?
3. What is the amount of risk on the organization when a specific user conducts an insider attack?

In the following parts of this chapter, a literature survey about insider attack detection methods has been conducted. The aim of this survey is to investigate the previous approaches to detect insider attacks as well as discuss their shortcomings, in addition to providing a categorization of these methods. In our classification, we relied on the technique used in each method of detecting insider threats. Based on the threat-detection technique, insider attack detection methods can be classified into five categories. Figure 3, illustrates the taxonomy of insider attacks detection methods.



**Figure 3: The Taxonomy of Insider Attacks Detection Methods**

### **3.3.1 Anomaly Detection Methods**

These methods usually come as software or hardware that monitors several devices, such as computers, communication networks, and logging information, to identify normal activities and attempt to build a model containing the characteristics of these activities. In addition, these methods aim to identify malicious activities that violate and deviate from this model [65]. Anomaly detection methods are often known as intrusion detection methods, which are approaches widely used for detecting external attacks. Although detecting insider attacks is more complicated than detecting external attacks since the inside attacker is a legitimate user with authorized access to the system and can commit attacks using his privileges, external threats can be detected more easily via sensing infiltrations to the system or any unauthorized and unusual behavior. However, some of the intrusion detection methods are adapted to identify the insider attacks based on the differentiation between the normal and malicious activities of insiders [66, 67]. Detecting insider threats using anomaly detection methods can be classified into indicator-based methods and scenario-based methods.

#### **3.3.1.1 Indicator-Based Methods**

In [68], the authors proposed an insider attack detection system based on the indication of the employee's suspicious behavior. This approach works as part of intrusion detection systems. It assesses the employee's behaviors based on several types of log information collected during the employee's activities. The logging information stored in a dedicated database are the device ID, user ID, activity name, time stamp, and attributes of each activity. The system updates the employee's profile through analyzing the stored logging information that has been observed from the employee's activities hourly. The analyzing

process divides the employee's profile into three parts: current, previously suspicious, and normal observations. Utilizing the previously suspicious and normal observations, the system can assess the deviation of current activities from the previously observed activities. The system has three levels of alert: the first alert takes place when the organization's policies are violated; the second alert happens when the specific employee's act exceeds a threshold level of anomalies; and the third alert occurs during a deviation between the employee's activities and his profile records.

Ambre and Shekokar [11] utilized the log management technique to build a log monitoring system for detecting insider threats. The log management technique consists of two related aspects, which are log analysis and event correlation. Log analysis deals with collecting, analyzing, and filtering log files (i.e., computer files that record several events for computer operating systems and users, such as system information, keystrokes, and data manipulation). On the other hand, event correlation is the process of finding mutual relationships among several events. Unlike offline monitoring approaches, which suffer from the inability of detecting attacks in real time, Ambre and Shekokar proposed a continuous monitoring system for the log files. However, given the huge number of log files in the network and types of malicious activities, analyzing and correlating different events would be extremely difficult. As a result, Ambre and Shekokar considered only three activities that indicate malicious events: Internet control message protocol (ICMP) requests, unsuccessful log-ins, and rebooting the server.

Moreover, detecting insider threats is not enough because detection always takes place after the attack. Predicting insider threats reduces the consequences and impact of these threats. Schultz [10] proposes a framework to predict and detect insider threats that uses



multiple indicators based on the best practice. Each indicator's contribution is represented mathematically by weight. The indicator's weight depends on the number of incident results from that indicator. Equation 3.1 shows the mathematical representation of Schultz's approach:

$$Xe = V_1X1_i + V_2X2_i + V_3X3_i \dots \dots + C \quad (3.1)$$

where  $Xe$  is the predictive value,  $X1_i$  is the first indicator,  $V_1$  is the first weight, and  $C$  is the constant.

The indicators suggested by Schultz [10] are deliberate markers, meaningful errors, preparatory behavior, correlated usage patterns, verbal behavior, and personality traits. The previous indicators are listed from the highest to the lowest weight based on Schultz's framework, whereas the weight for each one can be calculated from the indicator's density after carefully analyzing the number of attacks accrued.

### **3.3.1.2 Scenario-Based Methods**

Zargaret *et al.* [66] utilized the raw logs of user network sessions to detect insider attacks. They proposed a method called XABA, which analyzes the raw logs of each network session to detect abnormal activities. Based on the user behavior profile, which represents the exclusive user behavior and potential access, XABA analyzes the network traffic to detect the activities and patterns that violate the user profile meaningfully. The researchers in [66] classified XABA as a zero-knowledge approach because it is independent of any log syntax or any data entry about the user. XABA has been designed to detect five diverse scenarios of insider attacks that have a common feature, which is the misuse of the exclusive user behavior. The five scenarios are (1) betrayer admin, (2) third-party back

door, (3) credential sniffer, (4) e-mail spoofing, and (5) foothold hosting. Table 1 illustrates the description of these scenarios. Moreover, XABA consists of five units:

1. Session gathering utilizes the user IP address to collect the user sessions by analyzing the network traffic.
2. Session PBI (potential behavior indicator) making utilizes text mining to analyze the user session and extract useful information such as time stamps and the IP address.
3. Anomaly detection checks for malicious user sessions.
4. Insider detection matches malicious behavior with user-exclusive behavior.
5. Alert prioritization utilizes the suspicious degree to prioritize alerts.

**Table 1: The Description of XABA Scenarios**

<b>Scenario Name</b>	<b>Scenario Descriptions</b>
Betrayer Admin	Detecting the Traitor admin who misapply his privileges.
Third Party Backdoor	Detecting the web applications that stealthy provides server's control to the backdoor developer.
Credential Sniffer	Detecting the insider who sniffs the user's credentials
E-Mail Spoofing	Detecting the insider who spoof the user's e-mails
Foothold Hosting	Detecting the insider who Intentionally download a malicious software from the e-mail attachment.

Another approach that aims to mitigate the risk of insider threats by detecting inside attacker scenarios was proposed in [69]. Young *et al.* [69] use an ensemble with an unsupervised learning technique to detect insider threat scenarios without any prior knowledge about the kind of scenario and when it would take place. They conducted their experiments on a well-known database collected by the Advanced Research Projects Anomaly Detection at Multiple Scales (ADAMS) program, which consists of the data of

around 5,500 users [70]. The database contains several user actions, e-mails, log-in/log-off time stamps, printer URLs, and instant messages. Furthermore, Young *et al.* developed insider attack scenarios based on those created by the red team, an independent evaluator group who utilizes their experience in developing a description about several insider attack scenarios [39]. Additionally, the proposed scenario-based detector by Young *et al.* consists of a combination of three components, which are the classifier and the indicator-based and anomaly-based detectors. Table 2 illustrates the proposed scenarios by Young *et al.* [69] and the corresponding Red-Team scenarios.

**Table 2: The Proposed Scenarios by Young *et al.* [57]**

<b>Young <i>et al.</i> Scenarios</b>	<b>Descriptions</b>	<b>Corresponding Red-Team Scenarios</b>
<b>IP Thief</b>	The insider uses the organization IT resource to steal IP address. This scenario usually conducted by salespeople, engineers and scientists to get advantage of IP addresses in their work.	<ul style="list-style-type: none"> <li>•Anomalous encryption</li> <li>•Bona Fides</li> <li>•Manning Up (Redux)</li> </ul>
<b>Saboteur</b>	The technical insider such as a system administrator uses the IT resources to harm the organization. This scenario should be planned before the attacker leaving the organization.	<ul style="list-style-type: none"> <li>•Circumventing SureView</li> <li>•Layoff Logic Bombs</li> <li>•Survivor's Burden</li> </ul>
<b>Fraudster</b>	This scenario is conducted for a financial gain by low-level employees. The insider aims at destroying or denying the organization's data. The insider in this attack is usually recurred by outsider attacker.	<ul style="list-style-type: none"> <li>•Hiding undue Affluence</li> <li>•Indecent RFP</li> <li>•Masquerading 2</li> </ul>
<b>Ambitious Leader</b>	This insider is usually an IP thief motivated by the ambition to steal as much as possible before leaving the work.	<ul style="list-style-type: none"> <li>•Insider Startup</li> <li>•Selling Login Credentials</li> </ul>
<b>Rager</b>	Insider uses threatening, vociferous language in his mails to other employees or against the organization.	None
<b>Careless User</b>	This scenario is usually unintentionally conducted by an insider employee who expose the organization to considerable risk.	None

### **3.3.2 Honeytrap Traps**

Information is power; the more you know about your enemies and how they operate, the more power you have against them. One of the tools that assist in gathering information from attackers is a honeynet, a network-connected computer software or even a device that appears to be attractive and vulnerable. Honeynets were designed expressly to be attacked and to attract attackers just as honey attracts bears [71, 72].

A honeynet contains multiple honeypots. A honeypot has no production value and no authorized activity; it sits on the organization's system within its containing honeynet to be attacked because a honeypot can be accessed directly without any authorization [73]. All access to a honeypot is considered inimical. Any connection started from a honeypot to the external network is considered as an indication that the honeynet has been compromised. This connection is usually initiated by an attacker to download some malicious tool kit that will be used to commit his attack and to hide his trace from being tracked [74]. Without the knowledge of the attacker, the honeypot monitors every action of the attacker, and it strives to capture as much data as possible.

Data control is an important concept in deploying the honeypots. Data control means that the attacker must be prevented from using the compromised honeypot to attack other computers. Thus, the attacker will be locked into a cyber jail and be unable to utilize the compromised honeypot to commit an effective attack. A honeynet monitors and captures attackers' logs, actions, and methods [75].

Honeynets can be classified into three major categories based on the interaction level: high interaction, medium interaction, and low interaction. However, the higher the honeynet

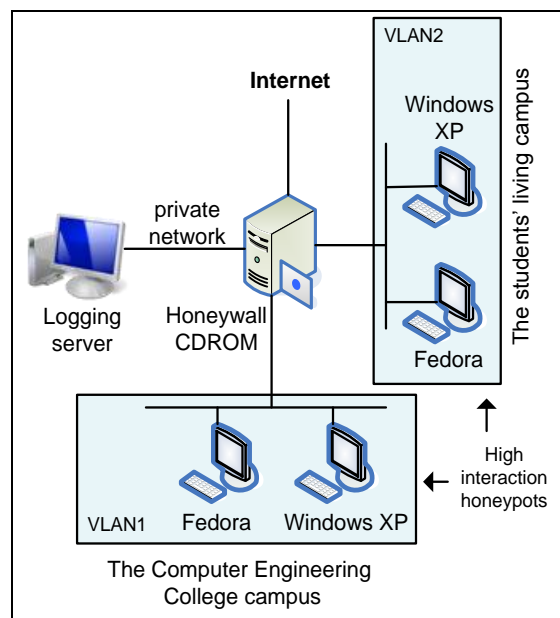
interaction level, the higher the risk on the operational system since the intruder would deal with real systems. Unlike in the low-interaction honeynet, where the intruder deals with the systems that emulate vulnerable services, the intruder in the high-interaction honeynet deals directly with real vulnerable systems. Moreover, the honeynet interaction level is directly proportional to the amount of data that can be collected from the intruders [75].

Honeynets can detect intentional as well as unintentional insider attacks, where the honeynet traps may contain several phantom assets for the organization, such as fake files, databases, and servers. These phantom assets play a vital role in the detection of the intentional inside attacker, where the honeynet's traps use unknown IP addresses for employees of the organization or are programmed to change their IP addresses automatically to deceive the insiders [76, 77]. On the other hand, the computer of an organization's employee can be infected by malware or an Internet worm. This computer becomes a source of penetration without the knowledge of the employee. A honeynet can detect and protect against such unintentional attacks.

We did previous work for deploying a distributed high-interaction honeynet at King Fahd University of Petroleum and Minerals (KFUPM) [78, 79]. The university's data traffic was captured for twenty different intervals, each of which was around ninety minutes, so the total interval for collecting data is thirty hours. More than thirty thousand activities were collected during all the experiment's intervals. Then the data was replayed on the proposed honeynet system. Figure 4, illustrates the implemented system. We utilized the following tools for implementing the system:

1. Honeywall CDROM, which is a high-level interaction honeynet that acts as a centralized logging server for the distributed honeypots.
2. Snort, which is a network intrusion detection/prevention system that was used for a real-time traffic analysis.
3. Sebek, which was used to intercept the attacker's data after decryption in the honeypot.
4. Wireshark, which was used to capture the university network traffic.
5. Tcpreplay, which was used to simulate and replay the university network traffic.

The results show that around 35% of the traffic is considered as low risk, which contains traffic such as DHCP requests and NetBIOS datagram services. On the other hand, around 65% of the traffic is considered as medium risk; most of this traffic is BitTorrent traffic. In addition, the proposed system successfully detects an insider attack on Internet Information Service (IIS), which was previously installed on the Windows XP honeypot [78, 79].



**Figure 4: KFUPM Distributed Honeynet**

Additionally, multiple deception techniques were proposed by Virvilis *et al.* [80] to address the advanced persistent threats (APTs) that may be utilized by the inside attackers. APTs are highly sophisticated attacks targeting organizations and looking to steal personal identifiable information or IP. In APTs, the attacker wants to gain access to the network and stay there as long as they can to understand what is happening in this network, searching for valuable data for exfiltration, i.e., transfer these data illegally from the organization. Virvilis *et al.* divide the APT life cycle into two phases. The first phase is the attack preparation (information gathering), and the second phase is the exploitation and data exfiltration. Virvilis *et al.* proposed a specific deception technique for each phase. The deception techniques for the first phase are as follows:

1. Using DNS honeypot: A honeypot or honeypot is a system resource that can track and analyze malware traffic [73]. DNS honeypots are fake records in the DNS server. Requesting these records accounts as brute-force network scanning, when the attacker tries to gather information about the network IPs. DNS is configured to initiate an alert when fake records are requested.
2. Using Web server honeypots, such as invisible links, fake entries in the Web server, and fake HTML comments.
3. Creating fake social network avatars in major social networks: To appear more realistic, these fake avatars must have connections to people from inside and outside the organization. In addition, these fake avatars should have real but monitored e-mails in the organization. Attackers may target these e-mails with malicious attachments. Therefore, these avatars can help detect attackers from inside or outside the organization.

The deception techniques for the second phase are as follows:

1. Monitoring the darknets, which are unallocated IP addresses: Connections to these addresses are done by scanning the darknets' range of IP addresses either by malicious activity or by user mistyping. Therefore, multiple connections are counted as suspicious activity.
2. Deploying a honeynet, which is a security system that attracts hackers to attack: These systems appear to be open and vulnerable to attackers, but they have deliberate vulnerabilities that are monitored and controlled [69].
3. Using the honeypot or honeypot in the database servers.
4. Generating and spreading honey-files in the organization's network: Honey-files are files containing fake interesting information to attackers, such as usernames, passwords, and credit card numbers [74].
5. Creating honey accounts or bait accounts with simple passwords for detecting attackers.

### **3.3.3 Graph-Based Methods**

Detecting insider threats using graph-based approaches is the process of finding the inside attacker's activities from the data represented as a graph. Graph approaches use several techniques to extract useful information about the insider attacks from graph data. The graph approach is one of the most powerful insider attack detection methods because of the following reasons:



## **1. Interdependent graph data objects**

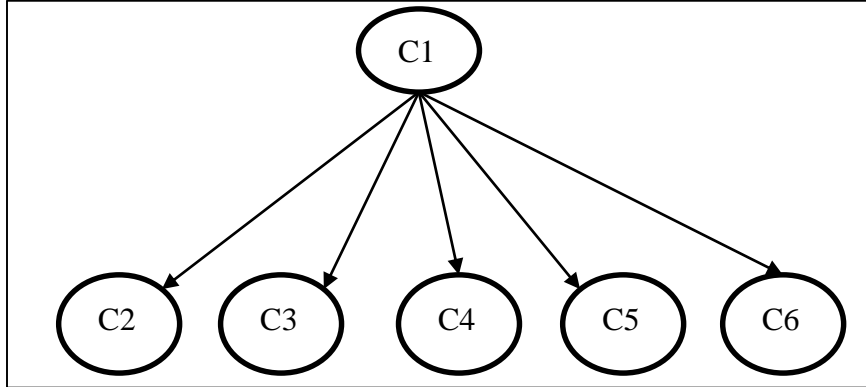
Graph data objects are dependent on one another, and these objects can strongly represent relational data, which makes graph data objects effective in presenting an organization network's information [81]. This can help detect insider attacks because using graph objects provides abundant information about the organization network, such as the users' activities on the network computers and, facilitates the representation of the organization network and the attacker activities on that network.

## **2. Robust representation**

Graph objects such as nodes, edges, and attributes can efficiently represent several data sets such as computer networks, cell phone networks, social networks, and biological data. For example, in computer networks, the computers are represented by the graph nodes, whereas the events and activities among these computers are represented by the graph edges. The powerful representation of an organization network facilitates that of the inside attacker's activities on that network, which improves the detection of such attacks.

The area of detecting insider threats using graph-based approaches is popular and promising. Several methods were proposed to mitigate the risk of insider attacks [81]. Kent *et al.* [82] use bipartite authentication graphs to mitigate insider attacks and assess the authentication of an enterprise network. To represent the activity of each user, they proposed a Parsons authentication subgraph (PAS), which is a directed subgraph representing the authentication user activity over a period of the data set. The period depends on the type of user. They select a period of one year for the administrators and

general users, whereas the period of compromised users is one month. Figure 5, shows an example of PAS.



**Figure 5: PAS Graph For a Single User**

Each vertex in a PAS graph represents a computer that the user accessed during the specified period, whereas each directed edge represents one or more authentication events on the specified computer. From Figure 5, we can notice that the user accessed six computers, including the user’s computer (i.e., node C1, normally his/her desktop). From each PAS, authors extract some features such as time features, i.e.,  $T_{\text{first}}$  and  $T_{\text{last}}$ , where  $T_{\text{first}}$  represents the first time the authentication event of the user on a specific computer was observed and  $T_{\text{last}}$  is the last time the event was observed. Authors use logistic regression to classify the general and administrative users as well as differentiate between malicious and benign users [82]. Moreover, to conduct their experiments, authors utilize the logs of the authentication data set from the LANL enterprise, which contains around 33.9 billion event logs collected during one year from around twenty-four thousand computers.

On the other hand, because the inside attackers are authorized persons, they do not perform abnormal activities all the time, so they try to hide malicious activities over time during

daily work to avoid being caught. Therefore, discovering rare activities, which are the minority compared to daily work, may help in detecting insider threats. A graph approach for detecting inside attackers by detecting the rare categories was presented by Zhou *et al.* in [83]. They proposed two incremental algorithms called SIRD and BIRD to solve the problem of detecting the rare categories on time-evolving graphs. The algorithms dynamically update the detecting models to fit the changes in data during the period. To increase the efficiency, these algorithms update only the local changes in the models instead of rebuilding the whole models. The SIRD algorithm deals with a single-edge update, when there is a change in only one edge. On the other hand, when there are changes in a batch of edges in the same period, BIRD facilitates the updating process of the batch of edges. In addition, the priors of the minority classes in some applications are hard to obtain; therefore, Zhou *et al.* proposed BIRD-L1, which is a modified version of the BIRD algorithm. Instead of detecting the exact boundaries, BIRD-L1 needs only the upper bound of all minority classes. Moreover, Zhou *et al.* tackle the problem of reducing the number of updated queries, and they introduce five categories to detect the rare activities as soon as possible with the minimum cost during multiple time steps.

Moreover, Mongiovi *et al.* [84] use a window-based events technique to detect the attacks in the network. This technique utilizes a time window on each network graph to detect suspicious behaviors and patterns. Normal behavior is modeled using a number of known previous instances of the graph, whereas the incoming graph that represents unknown behavior is compared with the normal behavior model to characterize it as malicious or benign behavior. Mongiovi *et al.* proposed a method to detect the significant anomalous, contiguous region in the network graph over time by detecting the heaviest dynamic

subgraph (HDS). For each weighted subgraph, authors converted the problem of finding the HDS to the NP-hard problem. Utilizing the empirical distribution of each edge weight, the authors represented the suspicious degree of each edge as a statistical probability measure (p-value). The higher positive p-value corresponds to a lower suspicious edge (i.e., network event). To approximate the NP-hard problem of HDS, the authors utilized a large-scale neighborhood search approach to propose an iterative algorithm that can detect the suspicious regions in the graph.

Another approach that traces user behaviors and converts these behaviors into a graph to detect malicious insider activities was proposed by Lamba *et al.* [85]. The graph represents the system architecture, where each node in the graph is a system resource. The path between the nodes represents the user behavior, which is a sequence of user events, so when the user accesses resource B from resource A, an edge between nodes A and B is generated. To trace the user behavior, this approach depends on three attributes collected from the user's log files. The attributes are the time stamp, user ID, and resource ID. Moreover, normal users' behaviors fall under different clusters based on the user role or task. Each user behavior belongs to only one cluster. Therefore, when the user behavior differs from its cluster, this behavior is suspected as an anomaly. Then the anomalous behavior needs to be investigated and determined if it is a malicious insider attack. However, the number of log files in a network is large compared to the number of threats. This challenge affects the results of similar approaches.

### **3.3.4 Game-Based Methods**

Feng *et al.* [86] used game theory to effectively protect against two types of complex threats, namely, APTs and insider threats. An APT is considered one of the most dangerous

and cost-effective attacks that targets the organization's security systems. APTs are usually conducted by stealthy outside attackers to do considerable harm to the organization. One of the reasons that complicate these attacks and make them more harmful is the presence of an insider who facilitates these attacks. Researchers in [86] used game theory to understand the strategies of these complex attacks, build a security system, and reduce the risk of such attacks. They proposed a game that consisted of three characters—the attacker, the insider, and the defender—to simulate the interactions among these characters during the attack. The game is an extension of a two-player game called FlipIt, proposed in 2013 by RSA Labs [87]. The role of the insider is to utilize his/her privileges to leak information about the organization to the attacker for money, whereas the role of the attacker is to utilize this leaked information to create sophisticated attacks against the organization's security system. On the other hand, the defender plays the role of the development of cost-effective protection methods, where the balance between an organization's data loss and the cost of defense against the attack is taken into consideration.

On the other hand, the traditional methods of protecting critical constructions such as nuclear reactors rely on protection from specific and predefined attacks. These traditional methods neglect the attacker's intentions, which can create a variety of intelligent and unexpected attacks. As a solution to this problem, Kim *et al.* [88] relied on game-theoretic modeling to analyze and develop the capabilities of physical protection systems for nuclear reactors. The researchers used the advantages of modeling games to create an intelligent inside attacker with full knowledge of the nuclear reactor parts, the potential to access the reactor, and the intention to harm it. The intelligent virtual attacker works on analyzing the physical protection system of the nuclear reactor and develops methods and strategies to

penetrate this system by conducting certain actions that will weaken the protection system. In addition, the researchers developed an intelligent defender who would protect the system by taking actions to protect the reactor and mitigate the risk of inside attacker. The authors aim at promoting the protection of nuclear reactors by changing and adopting new physical protection systems as well as changing the reactor policies. Furthermore, based on the evaluation of the physical protection and proliferation resistance, the researchers categorize the insider threats that target the nuclear reactors as in Table 3.

**Table 3: The Categorization of Insider Threats Target The Nuclear Reactors**

<b>Category</b>	<b>Types</b>
Type	Authorized individual.
Capabilities	<ol style="list-style-type: none"> <li>1. Knowledge.</li> <li>2. Number of attackers.</li> <li>3. Skills.</li> <li>4. Dedication.</li> </ol>
Objective	<ol style="list-style-type: none"> <li>1. Malicious attack on reactor facility.</li> <li>2. Sabotage.</li> </ol>
Strategy	<ol style="list-style-type: none"> <li>1. Exceeding security measures.</li> <li>2. Neutralize protection systems</li> </ol>

### **3.3.5 Physiological Methods**

Physiology is a part of biology that attempts to explain the activities and functions of living beings as well as the chemical and physical phenomena. Moreover, it deals with living matter such as cells, tissues, and organs in addition to attributes such as feeling, sensing, and emotions. Human physiology is the part of physiology that endeavors to clarify specific human body mechanisms and characteristics that make human life possible under vastly varying conditions [89, 90].

Because human physiology explains and deals with many of the functions, characteristics, and vital signals that characterize a person and determine his behavior, this part of science

has become a fertile environment for studying suspicious human behavior associated with insider attacks. Several methods utilizing human physiology have been proposed to detect insider threats in organizations. We have categorized these methods as behavioral, biometric, and bio-signals.

### **3.3.5.1 Behavioral Methods**

Human behavior refers to the observable emotions and physical actions that can be performed by an individual. There are two operations by which human behavior can be distinguished: It is observable and countable. So, human behavior is an action or emotion that can be observed several times. Human behavior may be influenced by several factors, such as ethics, attitudes, culture, and persuasion [91, 92].

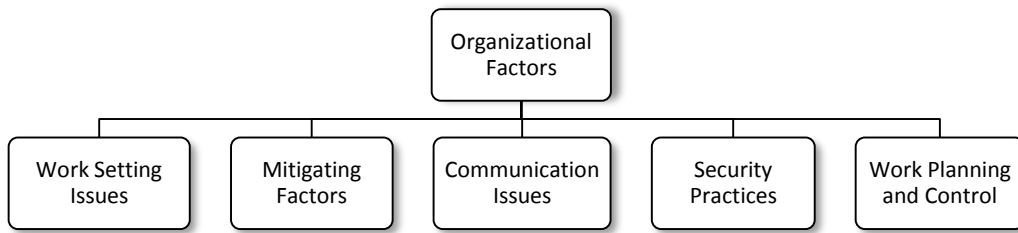
Several studies have been conducted to identify and classify the factors with which an insider attacker can be detected. Greitzer *et al.* [93] classified the human behavioral factors that help to understand the insider attacker and mitigate the risk of such attacks into individual human factors and organizational factors. The individual human factors concentrated on the human personality, temperament, ideology, attitude, and behavior issues, whereas the organizational factors concentrated on the work features that affect the attitude, satisfaction, protection, and safety of the employee as well as the organization's policies and practices. Greitzer *et al.* divided the human behavior into classes and subclasses up to 7 levels deep. The overall factors are 223 human factors and 39 organizational factors. Table 4 shows the main classes and subclasses of the human factors, whereas Figure 6, illustrates the main classes of organizational factors.

**Table 4: The Main And Subclasses Of The Human Factors by Greitzer *et al.* [93]**

<b>Human Factors</b>	<b>Concerning Behaviors</b>	Job Performance
		Boundary Violation
		Cyber security Violation
	<b>Capability</b>	
	<b>Dynamic State</b>	Attitude
		Affect
	<b>Static Trait</b>	Personality Dimensions
		Other Personality Traits
		Temperament
	<b>Life Narrative</b>	Financial Concern
		Criminal Record
		Personal History
	<b>Ideology</b>	Disloyalty
		Radical Belief
		Unusual Foreign Contact

Furthermore, Greitzer *et al.* developed a psychosocial model for predicting insider attacks [9]. This model utilized employee behavior to predict insider malicious activities through proactive means. The model evaluated employees' behavior to recognize potential risks. This evaluation was based on the automation of different indicators that result from the experience and best practices of the Human resources (HR) managers in recognizing employees' psychosocial behavior. Table 5 illustrates an example of the evaluation indicators based on the experience of two HR managers. Furthermore, the authors recommended three prediction algorithms for analyzing the model's result: nonlinear feedforward neural network, Bayesian model, and a linear regression model, which give the best results for predicting attacks.





**Figure 6: The Main Classes of Organizational Factors by Greitzer *et al.* [93]**

**Table 5: The Evaluation Indicators Greitzer *et al.* [9]**

<b>Indicator</b>	<b>Description</b>
Disgruntlement	Ex. negative feelings about being disregarded.
Not Accepting Feedback	Taking criticism personally, or does not confess to mistakes
Anger Management Issues	Cannot manage anger range or emotional feeling.
Disengagement	The employee does not cooperate with individuals and groups.
Disregard for Authority	Neglecting polices and authority feeling above the rules.
Performance	Receiving a worming because of poor performance.
Stress	Physical or mental tension.
Confrontational Behavior	Aggressive behavior such as intimidation.
Personal Issues	Personal issues interfere with work issues.
Self-Centeredness	Concerned mainly with own welfares
Lack of Dependability	Undeserving of trust and cannot keep promise
Absenteeism	Continuous absenteeism without reasonable excuse.

### **3.3.5.2 Biometric Methods**

Human biometrics are measurable and distinctive characteristics that can identify a person. Some biometric identifiers such DNA cannot be changed or revoked. Examples of human biometrics are iris, fingerprint, and face recognition, as well as human behavioral patterns such as the way a person walks or writes [94, 95]. The fingerprint is one of the earliest

human biometrics used in crime investigations. Following are some studies that used human biometrics for the purpose of detecting insider threats.

The control of eye movement is one of the physiological biometrics that may give an indication of voluntary and involuntary actions. Matthews *et al.* [96] investigated the potential for using eye tracking to detect probable insider attacks. Also, they proposed some design principles that may be used in real environments to detect such attacks. They utilized active indicators, which are stimuli that would evoke from the insider attacker distinctive responses, which can be distinguished by tracking eye movements. Matthews *et al.* utilized some eye tracking metrics such as the fixation duration and the saccadic frequency of eye movements to detect insider attackers. Such metrics were used to identify some characteristics of the insider attacker such as general concealment of interest, implicit responses during the malicious act and intentional strategic concealment. Participants in the experiments were divided into two groups, the control group which conducted (normal - normal) role, and the insider group which changed its role from normal to malicious activities and conducted (normal - insider attacker) role.

Babu and Bhanu proposed a biometric authentication system that utilized keystroke log files to mitigate the impact of insider attacks in cloud computing [97]. This approach consisted of a combination of trust, risk, access control, and user typing behavior. The log files of keystrokes represented user typing behavior. Risk was counted as any activity that caused damage to the cloud. Furthermore, they classified insider attackers into three types: Two were from the cloud service provider company (a rogue administrator and an employee who had access to the clients' sensitive data), while the third type of insider attackers was defined as a rogue administrator from the client company. This system

recognized the users' behavior by converting the dynamics of the users' keystrokes into numerical features. Then, these features are fed to the Support Vector Machine (SVM) classifier to identify the malicious activities, based on the threshold of risk analyzer. To provide more flexibility, scalability, and different levels of security for the authentication system, the threshold level of the risk was proposed to be dynamic. The extracted features from the keystrokes are illustrated in Table 6 where: P is pressing, and R is releasing.

**Table 6: Extracted Features by Babu and Bhanu [83]**

Feature Name	Feature value
X	1 or -1 to identify the data belong to the user or not
F1	Keywords represented by most frequently 2, 3 or 4 letters
F2	Difference between P(last key) and P(first key) of the word
F3	Difference between P(last key) and R(first key) of the word
F4	Difference between R(last key) and P(first key) of the word
F5	Difference between R(last key) and R(first key) of the word

Although the user's typing behavior and keystroke log files can help identify some malicious attacks, analyzing only the log files cannot add great impact on preventing insider attacks in cloud-computing systems because inside attackers have control over such authentication systems.

Rudrapal *et al.* [98] propose an algorithm to increase the immunity of the organizations' systems from insider attacks. Their algorithm works as the second line of defense to protect the user's privacy besides traditional authentication, which are the username and password. The algorithm utilizes the dynamics of keystrokes to identify the user's identity. It tends to protect the organization from the insider attack when the attacker tries to impersonate another person's identity by using legitimate credentials for a different user. The keystroke features for every user are extracted and stored in a database. Then these features are

compared with the user's keystrokes, when the user types a specific text or credentials, before authenticating the user. The extracted features rely on the duration between pressing and releasing keyboard keys for digraphs and trigraphs, a digraph defined as two keys typed one after the other and trigraph as three consecutive keystrokes. The duration between pressing the first key and pressing the second key represents the digraph duration, whereas the trigraph duration is represented by the duration between pressing the first key and pressing the third key or the duration between releasing the first key and releasing the third key. Although this approach can prove the authentication and can help verify the user's identity, it does not guarantee protection from insider attacks. The attacker can use his legitimate credentials to steal valuable and precious information. Moreover, the furious employee can install harmful software such as logic bombs that may cause catastrophic damages to the organization.

### **3.3.5.3 Bio-signals Methods**

Bio-signals are physiological signals emitted by biological beings, which can be measured through electrical probes. In other words, for human beings, bio-signals are tiny electrical signals emitted from the human body when the muscles flex or eyes move or even during thinking or sleeping. The human body produces thousands of bio-signals every second.

The most common bio-signals are as follows:

- Electroencephalogram (EEG) signals: brain signals
- Electrocardiogram (ECG) signals: heart signals
- Electromyogram (EMG) signals: muscle movement signals

The fact that vital signals are not voluntary is what distinguishes them and makes them difficult to imitate. Furthermore, because vital signals reflect the psychological state of a human being, these signals can be a distinctive tool to identify internal threats.

Recently, a new research direction was established to distinguish between the benign and malicious activities of employees [99]. The aim of this new research direction is to detect malicious insider activities beforehand. Therefore, it can be served as an early detection system for insider attacks. This direction of research based on the human bio-signals such as EEG ECG or EKG signals.

Suh and Yim [18] investigated the applicability of using bio-data to detect insider attacks in nuclear power plants. They collected the bio-data from only two males, both of whose ages were around 25, and one of whom was left-handed. The experiments were based entirely on two tests: an emotion test and an intentional wrong decision test. The emotion test was based on showing some photos randomly to participants. The authors selected about 30 photos related to emotions. These photos could stimulate bio-signals in participants. The photos contained natural stimuli as well as immoral stimuli. For example, one photo showed a crime scene, whereas another photo showed a smiling child. On the other hand, the intentional wrong decision test was based on 140 scenarios that were programmed using MATLAB. The participants clicked the YES/NO button to make a specific decision. The two participants played different roles in this experiment, where each role represented a particular position or job in the organization. Some of these positions were truck driver, director, nuclear power plant operator, IT worker, and official in the National Intelligence Service. The 140 procedures of this test included crimes such as

giving an identity card to strangers and planting a bomb in the organization. Although the authors claimed they received promising results, their work had some drawbacks:

1. The size of the experimental sample is too small, where conducting the experiments on two participants surely will provide a low level of confidence in the results. To achieve a high level of confidence, the experimental sample size must be more than twenty participants [94].
2. Some aspects of the experiment are unrealistic. For example, in the intentional wrong-decision test, the two participants played different roles, and they occupied various unreal jobs, such as IT worker, truck driver, and director, so the participants will not be completely affected by making the wrong decisions in these unreal jobs. Furthermore, in the emotion test, some images have been presented to the participants, such as crime images. This test has two drawbacks. First, such images may not represent any real internal violations in the institutions. Second, the reactions of the participants from such images are unrealistic and do not represent real reactions of the inside attackers.
3. The participants realize they are conducting experiments related to inside attackers, so their emotions may not reflect the attackers'. Such experiments must be conducted in realistic situations and include real procedures.

Moreover, Hashem *et al.* [99] investigated the usability of detecting insider malicious activities by analyzing human electroencephalography (EEG) signals. They conducted their experiments on ten participants, five males and five females, whose ages varied from 18 to 33 years old. Each participant performed three scenarios. The duration of each scenario was ten minutes. Two scenarios were malicious, whereas the third was benign.

Malicious scenarios were an unauthorized login to a remote computer and the theft of source code files from a folder in the lab network, whereas the non-malicious activity was a regular office job such as surfing the internet, reading e-mails or using computer applications. The brain EEG signals were collected by using Emotiv EPOC device [100]. Then, the wavelet transformation was used to decompose the EEG frequencies into different sub-bands, which was used as frequency domain features. by recording signals from each electrode of the device in the same period. Then, the features are reduced using the principal component analysis (PCA) [101, 102]. Hashem *et al.* used the Support Vector Machine (SVM) to differentiate between the malicious and benign signals. Error rate, Accuracy, Precision, F-measure, and recall were utilized as measurements in the experiments. However, the sample of 10 participants is not enough to prove the results accurately [103]. Furthermore, SVM classifier tends to ignore the minority class while building the model, especially when dealing with unbalanced data [104, 105]. Ignoring the skewness in the data set leads to provide less accurate results. Therefore, in such cases, the receiver operating characteristics (ROC) curve provides results that are more accurate than using accuracy.

Almehmadi and El-Khatib [106] proposed an access control system (IBAC) based on the bio-signals. They used the EEG reactions toward visual stimuli to grant or deny users from access the protected resources based on the calculated risk. IBAC was based on the idea that users know their intentions of access. So, the system measured the user's intention using the amplitude of P300 waveform after stimulating the brain with the question "What is your intention to access the resource?". P300 is a positive EEG peak takes place 300 milliseconds after the visual stimuli. However, the procedure of calculating P300 ignores

Gamma brainwaves and relies only on Theta, Alpha and Beta bands. The first experiment aimed at detecting the lousy intention of a specific resource. The participants were asked to think of burning a lab while looking at random images that represent intentions such as burning lab, studying in the lab, and organizing labs. The second experiment aims to detect the internal threat using the IBAC system. To this end, participants were informed not to access some private folders unless they can do without getting caught and if they were caught the experiment failed.

Furthermore, Almeahmadi and El-Khatib [107] used the amplitude of ECG signals, the galvanic skin response (GSR), and the skin temperature to detect the insider threats. The signals were collected from fifteen students during normal and malicious scenarios. The malicious scenario was similar to the second experiment in [106]. The system did not utilize different features in the ECG signals; it used the ratio of the average of data that each sensor records per second to the standard deviation of the collected bio-signals as illustrated in the following equations:

$$SensorDataAv = \frac{\sum_{k=1}^{sec} (biosensordata_k)}{Second} \quad 3.2$$

$$SensorData SD = \sqrt{\frac{1}{n} \sum_{k=1}^N (SensorData Av - \mu)^2} \quad 3.3$$

$$Deviation Rate = \frac{SensorDataAv_{i-1}}{SensorData SD_i} \quad 3.4$$

*BioSensorData*: the recorded sensor data for each sensor.

$\mu$ : the mean of the SensorDataAv.

$N$ : the number of periods since the deviation started to increase.

$i$ : represents the current period.



Hashem *et al.* [108] studied the possibility of detecting insider threats using EEG signals and tracking eye movements. Twenty-five participants conducted five experimental scenarios, three normal and two malicious. The normal tasks aimed to study the reactions of the participant while performing benign daily activities, daily activities under stress, and high mental workload activities. On the other hand, the malicious scenarios were the SQL injections attack and unauthorized remote access. Participants were instructed to conduct the malicious tasks and told that no extra rewards if they leave any trace while performing the tasks. Hashem *et al.* extracted two types of features: the EEG features using the wavelet transform algorithm and the eye tracking features (i.e., the movements and pupil features).

Suh and Yim [109] discussed the feasibility of using the EEG signals to predict the insider threats in the nuclear environment. Their study aimed to investigate the feasibility of developing specific EEG signals to identify an insider before committing the attack. They analyzed the difference in the EEG data during the normal brain state and malicious actions. They utilized a wearable device to collect the signals from eleven participants. During the experiment, the participants had to read some problems about the nuclear reactor and click on YES or NO button. Clicking YES button represents a bad action whereas NO button represents a good action. To predict the insider threat, they used a quantitative EEG analysis to develop two types of indicators which are the relative power of (alpha, beta, theta, the sum of alpha and theta and gamma waves) and the ratio of brainwave-to-brainwave (i.e., gamma/alpha and beta/alpha).

**Table 7: Summarizing The Surveyed Methods**

Category		Reference	Components	Description
Increasing awareness about insider threats	Training	Chi <i>et al.</i> [62]	Educational Virtual Lab	Training lab for insider attacks awareness
	Educational Course	Rochester Institute of Technology[64]	Security course to the software developers	Educational course for insider attacks awareness
Anomaly detection	Indicator Based	Legg <i>et al.</i> [68]	Log files, device ID, user ID, activity name, timestamp	Network monitor to detect the user's activities. Using the current user's profile as an indicator
		Ambre and Shekokar [11]	Log analysis and event correlation.	Network monitor to detect malicious (ICMP) request, rebooting server, and unsuccessful login as an indicator of attack.
		Schultz [10]	Deliberate marker, Meaningful errors, Correlated usage patterns.	Using multiple weighted indicators to detect the insider threat.
	Scenario Based	Zargaret <i>et al.</i> [66]	Betrayer Admin, Third Party Backdoor, Credential Sniffer, E-Mail Spoofing	Monitoring each network session, analyzing raw logs to detect the suspicious scenarios
		Young <i>et al.</i> [69]	IP Thief, Saboteur, Fraudster, Ambitious Leader	Using an ensemble with unsupervised learning technique detect the suspicious scenarios
	Honeypot Traps	High interaction Honeynet	Sqalli <i>et al.</i> [70]	Honey-wall CDROM, Snort Sebek
Multiple deception techniques		Virvilis <i>et al.</i> [80]	Ex. honeynets, social network avatars and dark-nets monitoring.	Monitoring the network traps to detect the insider threats
Graph Based		Kent <i>et al.</i> [82]	Parsons's Authentication Subgraphs (PAS)	Using bipartite authentication graphs to represent the activities of each user's
		Zhou <i>et al.</i> in [83]	SIRD, BIRD and BIRD-L1	Using the time-evolving graphs to detect the user's suspicious actions
		Mongioui <i>et al.</i> [84]	Heaviest Dynamic Subgraph	Utilizing time window on each network graph to detect the suspicious behaviors and patterns

		Lamba <i>et al.</i> [85]	Network architecture graph, timestamp, user and resource IDs	Monitoring the user behavior in network based on graphs.
Game Based		Feng <i>et al.</i> [86]	Three-player game attacker, insider, and defender	Using game theory to protect against two types of threats, namely the advanced persistent threats (APT) and the insider threats
		Kim <i>et al.</i> [88]	Intelligent attacker and defender	Using game theoretic modeling to analyze and develop the capabilities of physical protection systems for nuclear reactors
Physiological	Behavioral	Greitzer <i>et al.</i> [93]	Criminal Record, Financial Concern, Disloyalty and Security practices	Using individual human factors and organizational factors to detect the insider threats
		Greitzer <i>et al.</i> [9]	Disengagement, Disgruntlement and Stress	Using the best practice and experience of HR managers to detect the suspicious behavior
	Biometric	Matthews <i>et al.</i> [96]	Fixation duration and saccadic frequency of eye movements	Utilizing the eye tracking for detecting the insider attackers
		Babu and Bhanu [97]	Keystroke and Typing behavior	Using biometric authentication system to detect the insider attacks in the cloud computing
	Bio-signals	Suh and Yim [18]	EEG, ECG and GSR	Detecting insider attacks in nuclear power plants
		Hashem <i>et al.</i> [99]	Brain EEG signals	Monitoring brain EEG signals to detect threats
		Almehmadi and El-Khatib [106]	EEG signals	Access control system (IBAC) based on the bio-signals
		Almehmadi and El-Khatib [107]	ECG, GSR, and Temp	Detecting malicious insider threats
		Hashem <i>et al.</i> [108]	EEG and Eye tracking	Detecting insider attacks
	Suh and Yim [109]	EEG signals	Monitoring brain EEG signals to detect threats	

### **3.4 Limitations in the Existing Studies**

Despite the great variety of strategies to identify insider attacks, as we have already mentioned in the literature of this research, the physiological strategies remain efficient in detecting insider attacks and overcoming the shortcomings of most other strategies because they depend on fixed metrics related to the human body, where the individual cannot change or control these metrics easily. In this section, the limitations of the strategies mentioned in the literature will be presented; moreover, the gap in research of detecting insider threats will be discussed.

#### **3.4.1 Limitations of Anomaly Detection Methods**

Although anomaly detection methods are common and used to identify internal attacks as they are also used to protect against outsider attacks, these methods have some shortcomings. These shortcomings can be summarized as follows:

1. Huge number of log files and network packets

Many anomaly detection methods rely on log files and network packets to extract the data and behaviors of the users; however, the number of log files in large organizations is extremely high. For example, a single user can have thousands of log files per workday, so the process of analyzing these log files lengthens the detection period of the suspicious behavior of inside attackers.

2. Detection of predefined attacks

These methods provide protection against a limited number of insider attacks, where the idea behind these methods is to identify specific scenarios or indicators of insider attacks as a blacklist [110], where new unknown scenarios are compared

with previously known suspicious scenarios in the blacklist to distinguish the behaviors of inside attackers. However, in the case of a new insider attack scenario that is not listed in the blacklist, this attack will not be recognized by these methods before causing catastrophic losses to the organizations. Research has proved that even the best application for malware detection can only detect, at most, 87% of the latest attacks [111]. Furthermore, the principle of these methods depends on the assumption that the number of insider attacks the organizations may face is low. On the contrary, it has been proven that the number of internal attacks grows because of the rapid development of technology, where these attacks pose a great threat to organizations.

### 3. Information security staff capabilities

Angry information security managers and staff who participate in the installation and management of anomaly detection devices have full knowledge of the capabilities of these devices and can control them, making them suspected as inside attackers because they can utilize their knowledge and their facilities to bypass these security systems.

## **3.4.2 Limitations of Honeypot Traps**

Honeypots and network traps, sophisticated anomaly detection methods, have some advantages over traditional malware detection methods. The honeypot approaches can detect and protect from unknown malware that have been caught by these traps. Moreover, these approaches can provide the opportunity to enhance the security system of the organization by providing further investigation and analyzing for anomalies. However, honeypots and network traps have substantial shortcomings:

1. Complexity of honeynet

The design and control of the honeynet, especially the high-interaction honeynet, is complex because building an emulating system for several network services and controlling the attackers are not easy processes; moreover, this emulating system must be immune to malware [112].

2. Honeynets have the highest risk

If the honeynet is compromised and controlled by the attacker, the attacker would have an opportunity to attack the organization's resources and use the honeynet as a platform to attack different institutions using its IP address, which would expose the institution to organization accountability [113, 114].

3. Honeynets increase the cost and complexity of the network

Building a honeynet requires the organization to dedicate resources, systems, and IP addresses to the honeynet. This may be expensive and complicate the organization network [113].

### **3.4.3 Limitations of Game-Based Approaches**

Although using game theory in designing security games has several advantages, such as using intelligent systems for discovering vulnerabilities in security systems and simulating characters, game-based approaches have considerable shortcomings:

1. Lack of scalability

Most of the security-game models are not scalable where the game consists of two or three players. Players represent the attacker, the defender, and the insider (betrayor). These models neglect some real-life cases of multiple attackers dealing

with multiple defenders because these models treat the whole number of defenders as only one, as with the attackers [115].

## 2. Lack of motivation

One main problem in modeling network security games is the lack of motivation from the difficulty of quantifying the value of protecting the organization's network as a result of modeling such games by non-specialist persons in the security systems because of confusion between how to assess and how to quantify the security of the organization's network [116].

## 3. Unrealistic Models

Security games treat the security system as finite, with no errors, which is false in real-life situations, where intrusion detection systems are erroneous [115].

### **3.4.4 Limitations of Graph-Based Approaches**

Despite the factors that give graph-based methods advantages and robustness in simulating the organization's network topology to detect network threats, these methods have several drawbacks:

#### 1. Approximate values from NP-complete problems

Although graph-based methods can represent the structure of the organization's network effectively, detecting insider attacks using these methods could lead to approximate results. The insider attack is a sequence of events in the network; each event can be represented as an edge. Traversing these edges in the organization's network, which consists of many computers, printers and other resources, leads to nondeterministic polynomial time problems, such as NP-complete or NP-hard

problems. These problems have approximation solutions that may affect the detection results [84, 117].

## 2. Unacceptable computing time

Graph-based methods consume relatively little time when run on small graphs (around fifteen to twenty minutes with a graph of three hundred nodes). However, in the real world, an organization's network consists of thousands of computers. Detecting insider threats using graph-based approach leads track the insider attack events in the graph; the computing time for traversing graph edges to detect such threats may lead to NP problems, which cannot be solved in real time [118].

### **3.4.5 Limitations of physiological methods**

Physiological strategies, especially bio-signals, are among the latest strategies to detect insider threats. Several factors have contributed to the emergence of such methods, including but not limited to the rapid technical development, low cost, and ease of use of small bio-signal measurement devices. Detecting insider threats with bio-signals is one promising research area. Although these approaches use bio-signals such as EEG and ECG signals for mitigating the risk of insider attacks, the existing approaches have some considerable shortcomings:

#### 1. Size of the experimental sample

The experimental sample size of existing methods is small. Experiments with fewer than 20 participants may achieve low credibility and poor confidence scores [103].

#### 2. Measurements and classification

Detecting suspicious behavior using bio-signals is the process of searching for minority-class activities compared with normal activities. Existing bio-signal



methods did not use the right measurements or more than machine-learning classifiers to prove their results. Almost all of these methods relied on accuracy alone, which may not be a good indicator. Moreover, they used classifiers that tend to ignore the minority class while building their models, especially when dealing with the unbalanced data of insider attacks.

### 3. Unrealistic experimental scenarios

In the existing bio-signals methods to detect insider attackers, the experiments for collecting bio-signals from participants relied on unrealistic scenarios in which the participants were instructed to act suspiciously. This significantly affected participants' decisions and certainly had a negative effect on their locus of control. The locus of control, a concept developed in 1954 by Julian B. Rotter, is defined as “the degree to which people believe that they have control over the outcome of events in their lives, as opposed to external forces beyond their control” [119]. When individuals believe that events in their lives derive mainly from their own actions and that they are responsible for their actions, they are known as individuals with an internal locus of control. On the other hand, individuals with an external locus of control believe that events in their lives are affected by external forces beyond their control. Giving instructions to the participants affected their locus of control and led to a lack of a sense of responsibility for the consequences of their actions, thus affecting their brains' responses [120, 121]. Furthermore, Edward and Ryan [122] published a model of human motivation called self-determination theory. In the context of self-determination theory, autonomy represents your ability to choose not only the things you do but the way in which you do that.

Autonomy has a significant impact on human's motivation. The motives of individuals who simply conduct an objective that they're responsible for are entirely different from the way that they do the same objective when instructions are given. So, the truly human acts are associated with the freedom to conduct that act. So, in the existing bio-signals approaches, the participants were not truly acting suspiciously like real insider attackers, which completely affected the shape of their brainwaves as well as the rest of their bio-signals.

#### 4. Lack of responsibility

A sense of responsibility arises from the awareness that an individual has control over his actions. This awareness stems from linking an action with its consequence and from counterfactual reasoning when the individual has a choice to do a different action [121]. Responsibility plays a vital role in individual decision-making and in intentional antisocial behavior. It has been proven that the feeling of responsibility affects brain signals because it is associated directly with regret, which is also associated with processes occurring in the brain's prefrontal cortex [123]. Therefore, instructing the participants to act maliciously affected the brain signals used in decision-making and feeling. Thus, the participants' bio-signals do not represent real insider attackers' bio-signals. Moreover, Frith [121] proved that volition and responsibility are also influenced by instructions.

Despite the bio-signal methods in the literature survey, little research has been carried out on the area of detecting insider threats using human bio-signals. As we have explained, many shortcomings in these methods affect the credibility of the results. One major shortcoming was that all the scenarios used in the experiments did not represent real

internal attacks to the best of our knowledge. The participants received instructions to act maliciously and did not feel responsible about their actions, which affects their bio-signals. Therefore, conducting research in this area is promising and may provide useful insight into the implementation.

In this research, the shortcomings of existing bio-signal approaches were taken into consideration. We reported the results of 84 participants to achieve high credibility. Also, we developed a real insider attack scenario based on physiological considerations, where we did not instruct the participants to carry out malicious activities, so they had real motivations. The decision-making was left to the participants to perform the attack, so not all the participants acted maliciously. Additionally, participants were fully aware that they were responsible for their activities as we will discuss in the next Chapter 5. Furthermore, to provide more accurate results without relying solely on the accuracy measurement, several measurements have been used to ensure the correctness of the results, such as the receiver operating characteristics (ROC) curve. These measurements would overcome the problem of searching for the minority classes as discussed in Chapter 6.

### **3.5 Summary**

The difficulty in distinguishing crimes of internal attacks from non-malicious activities is one of the reasons internal attacks are a concern for organizations. Increasing awareness among the organizations' employees about insider attacks is one of the solutions to prevent such attacks. Several studies that aim to increase the awareness about insider attacks have been presented in this study. Some of these studies provide technical solutions, such as the

virtual laboratory, to increase the security skills of trainers, while others have developed security courses about insider threats.

Moreover, based on the technique used in the threat detection, the existing detection methods of insider attacks have been categorized into five categories which are anomaly detection, honeypot traps, graph-based methods, game-based methods, and physiological methods. Each of these categories has been discussed, and several examples about each category has been surveyed in this chapter.

Additionally, the shortcomings of the insider attack categories have been discussed, where the focus was on the disadvantages of the detection methods that were related more to the problem of this research. One of the disadvantages of the existing insider threat identification methods is using unrealistic scenarios to simulate the attack from within and affecting the nature of the bio-signals being collected. However, the existing bio-signal methods to detect insider attacks neglected some of the physiological theories, such as the locus of control and the sense of responsibility. In addition, the size of the experimental sample for these methods is very small, which makes the results unreliable.

## CHAPTER 4

### RESEARCH METHODOLOGY

#### 4.1 Overview

To investigate the research problem, which is the utilization of human bio-signals to detect insider attacks, we divided the research methodology into five stages, where each stage aims to address a separate objective from this research. Moreover, each stage has its own results and deliverables, which are vital to the other methodology stages. Figure 7 illustrates the stages of the research methodology and the relationship between these stages.

#### 4.2 Stage 1: The Literature Survey

The first stage of this research methodology is to survey the literature as it provides full and up-to-date understanding about the research problem in addition to delivering the shortcomings of existing methods. So, to investigate the problem of insider attacks and to explore the existing approaches that aim to protect from such attacks, we conducted a literature survey about the methods that tackle this problem.

We discovered diverse methods of protection from internal attacks and classified these methods according to the technique used to identify the attack into five categories. Despite the number and variety of existing methods, we have found that only a few researches utilize human bio-signals to protect against insider threats [107]. An effective insider attack protection system should overcome the disadvantages of previous protection methods.

Therefore, in our literature review (Chapter 3, section 3.4.5), we have provided and discussed the disadvantages of existing methods that used human bio-signals. Moreover, as shown in Figure 7, this stage of research methodology (i.e., literature survey) has played an important role in the design of the protection system as it discovers the weaknesses of the previous approaches.

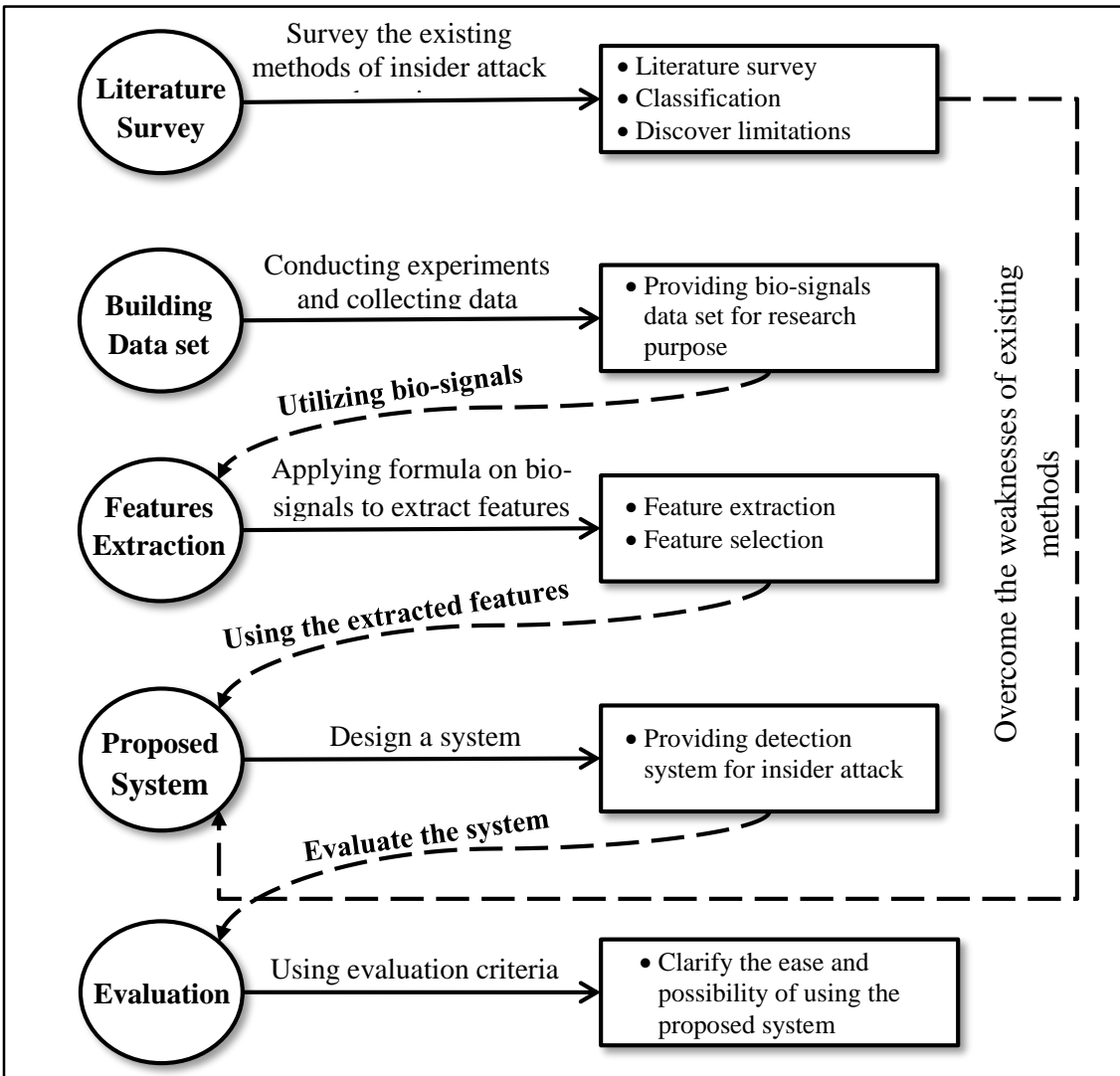


Figure 7: Research Methodology

### **4.3 Stage 2: Building the Bio-signals Data Set**

Protecting the organizations' security by detecting insider threats is a hot research area, despite that only few methods have been proposed recently using human bio-signals. Regardless of the number of these solutions, to the best of our knowledge, no data set exists that contains such bio-signals for research purposes. The data set for developing this research area—and as a reference of comparison among insider attack detection methods that use bio-signals—must meet the following conditions:

1. The data set should contain a sufficient number of bio-signal samples, from more than twenty persons [103] .
2. The volunteers' bio-signals should be collected through real scenarios that simulate the reality of the problem of the research (i.e., insider threats).
3. Psychological factors should be considered during the scenarios of collecting the bio-signals. These factors include the locus of control, decision making, and the feeling of responsibility about the consequences of the decision.
4. The data set should contain real human bio-signals collected during experiments and must not contain any autogenerated bio-signals using a simulator.

Consequently, to build a bio-signals data set that will be used also to design our system, which aims at accurately identifying internal violations using human physiological signals, the experimental scenarios for collecting the bio-signals from volunteers must be designed to simulate real cases. To this end, we have proposed two scenarios. The first scenario simulates the presence of an employee in a normal work environment, where the employee performs daily routine tasks using the computer. The second scenario simulates the

presence of the employee in a normal work environment, but the employee acts suspiciously and violates the organization's laws. In the second scenario, the volunteer has the opportunity and decision to carry out internal attacks without his knowledge that such violations are part of the experiment. The experiment scenarios are discussed in Chapter 5.

On the other hand, to build a data set of human physiological signals for detecting inside attackers requires identifying which bio-signals will be collected during experiments, where these signals are key to solving the main research problem. So through the literature review and exploring human physiological signals in some research fields—such as emotion recognition, intention detection [14], cryptographic systems (cryptographic key) [15], detecting read book genres [16], control systems, and crime detection systems, such as polygraphs or lie detectors [17]—several signals (biofeedback and neurofeedback) have been identified in these researches. The identified signals are vital in determining human moods, behaviors, and emotions. These signals are also important in distinguishing between a person's conditions when performing normal and suspicious activities. In this research, the experiments were conducted to collect these influential signals for the design of the proposed system.

This data set has been collected at Hadhramout University in the Republic of Yemen. The process of collecting bio-signals and conducting experimental scenarios continued for about seven months, where the experiments were conducted in two stages (i.e., suspicious and normal activities) according to the availability of the students. Each stage has been conducted separately. The data set contains bio-signal files for eighty-four volunteers, forty-three of which are males. These bio-signals were collected when volunteers performed two types of scenarios: normal work and suspicious activities. As a result of



these two types of scenarios, the data set contains four files for each volunteer's bio-signals. Each scenario produces two bio-signal files for each volunteer, the first file containing EEG signals and the second one containing ECG signals.

To protect the privacy of the volunteers, all personal information, such as the volunteers' names and departments, will be removed from the files. To differentiate the files in the data set, we have divided the file names into four parts. Figure 8 illustrates the naming process. Moreover, we have followed the subsequent rules for naming the files:

1. Part one of the file name contains a two-digit number that identifies the volunteer's number.
2. The second part of the file name contains three letters that identify the type of bio-signals, either EEG or ECG.
3. The third part of the file name contains only one letter that represents the scenario or the behavior of the volunteer; this will be either S for suspicious behavior or N for normal behavior.
4. Part four of the file name consists of one letter that identifies the volunteer's gender; this will be either F for females or M for males.

The files in the data set are grouped into two separate groups; the first group contains the files of the EEG signals for both suspicious and normal scenarios, whereas the second group contains the files of ECG signals for both scenarios. Moreover, one of the most well-known and popular file formats has been used to store the files in the data set: the extensible markup language (XML), which can be transformed easily into other files and can be used by any programming language. Furthermore, the XML format can store more than one type

of signals, such as heart rate, the standard deviation of the peaks and the peaks value, etc. where each type has a different style of representation, so storing data such as ECG in XML format is easy. Figure 9 shows the data set structure.

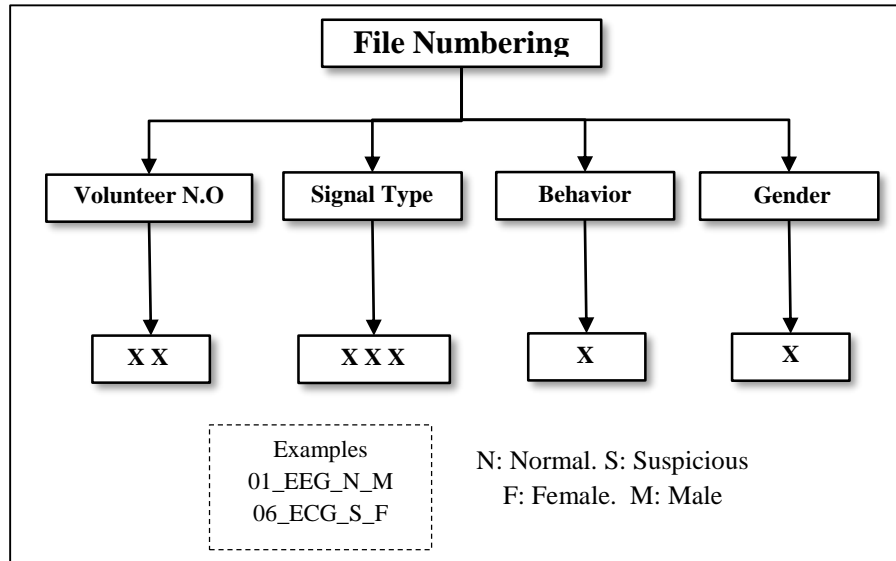


Figure 8: Data set Naming Process

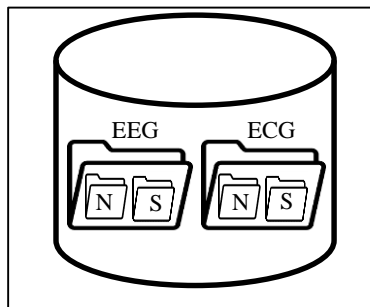


Figure 9: Data set Structure

The EEG signals were collected at a sample rate of 512 samples per second. Each EEG signal is represented in the file using ten fields; the first field represents the sample's number, the second one represents the sample's time, and the rest of the fields represent the frequency components of the EEG signal (i.e., delta, theta, alpha1, alpha2, beta1, beta2, gamma1, gamma2). Some EEG files contain very few bad signals, which are not recorded

in the file because of hardware limitations. In most cases, the bad signals between consecutive samples were five seconds at most and were symbolized in the file as NA (not available).

On the other hand, each ECG file consists of three groups of data. The first group, called samples, represents the information about each heart pulse and contains the sample number, sample time, and raw heart data, which represent the ECG signal value. The second group of data contains information about the R peaks, which is the maximum amplitude during a specific interval of the ECG wave. The information about the R peaks includes peak time, peak value, and the interval between successive R peaks. The third group of data contains the heart rate variability value (i.e. SDNN), which is calculated every sixty seconds. Figure 10 shows the ECG data representation in the file.

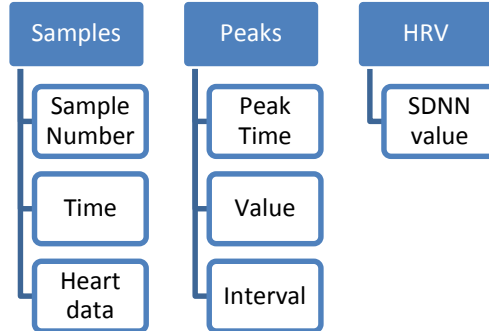


Figure 10: ECG Data Representation

#### 4.4 Stage 3: Features Extraction

Using the collected raw physiological signals to detect the insider violations for the complexity of these signals and for the associated noise is difficult. Therefore, these signals must be preprocessed before use in the proposed system. Moreover, some characteristics, called features, of these signals have the main role of identifying major changes during

suspicious behaviors. Features are the measurable, independent, and informative characteristics extracted from raw signals [23], which makes these signals more effective in identifying internal violations. Therefore, it is necessary to change the format of raw signals into features using the appropriate equations that will increase the strength of these features in determining internal violations.

Obviously, not all features have the same effect in detecting changes in bio-signals during attacks; some of these features should be nominated. On the other hand, the nominated features may play different roles in determining internal violations because of the changing rate of physiological signals in normal and suspicious activities. Therefore, it is necessary to determine the impact of each nominated feature on the accuracy of the proposed system. To achieve this, several evaluation criteria were used. Chapter 6 explores the importance of each nominated feature.

#### **4.5 Stage 4: Proposed System**

The idea of the proposed system is to detect malicious activities by distinguishing whether the changes in the bio-signals (EEG and ECG) are due to malicious behavior or the normal behavior of the employee. Figure 11 shows the schematic diagram of the proposed system. The proposed system consists of eight units, each of which has its turn in detecting suspicious behavior and mitigating the risk of insider attack. Each unit is discussed individually in the following subsections.

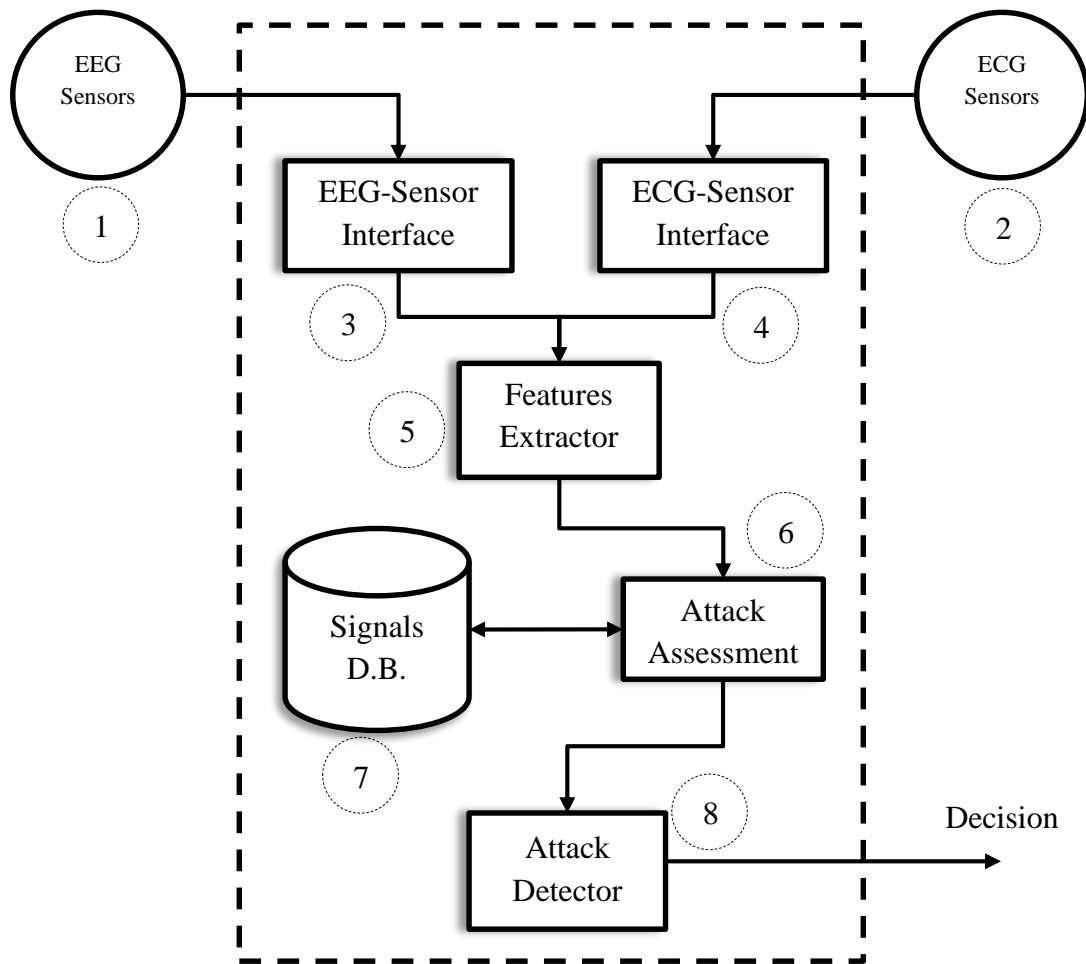
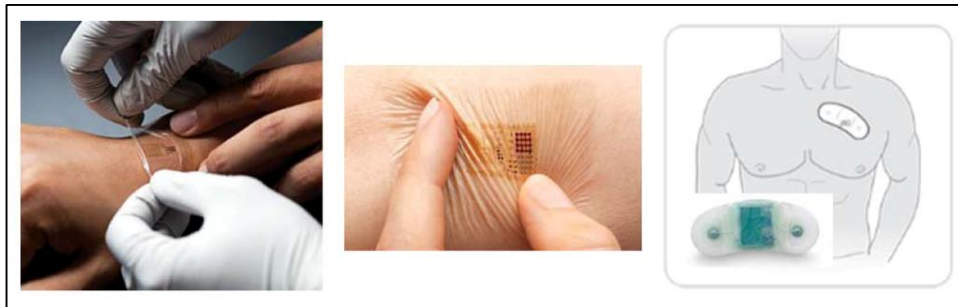


Figure 11: Schematic Diagram of The Proposed System

#### 4.5.1 Sensors

From the schematic diagram of the proposed system, units 1 and 2 represent the sensors, which are devices that are frequently used to detect signals or measure a property [124]. Sensors have been utilized to capture human bio-signals. Two types of sensors were used in our experiments to collect the human bio-signals. The first type is the EEG sensors that are part of the NeuroSky MindWave device, whereas the second type is a part of the Wild Divine device that is used to collect ECG signals. More details about these devices are discussed in Chapter 5, Section 5.7.1 and 5.7.2.

This research did not put any restriction on the types of sensors to collect human bio-signals. Any type of sensor could be used as long as it has the ability to collect bio-signals accurately. With the rapid development of science and technology, the shapes, sizes, and techniques of these sensors have changed greatly. Several kinds of small, convenient, easy-to-use sensors have been developed. These sensors can be used extensively by corporate employees. An example is the Motorola MC10 tattoo sensor [125], Figure 12 Shows several types of sensors to collect ECG signals.



**Figure 12: Wearable ECG Sensors**

Furthermore, depending on the sensor types, transforming the received signals from one format to another could be done either in the interfaces or in the devices that contain the sensors, such as in our case. The NeuroSky MindWave device transforms EEG signals from the time domain to the frequency domain using the standard fast Fourier transform (FFT) [126]. Because an EEG signal consists of different frequencies, FFT converts the EEG signal to its frequency component, i.e., delta, theta, alpha, etc. Figure 13 illustrates the process of transforming EEG signals from time to frequency domain.

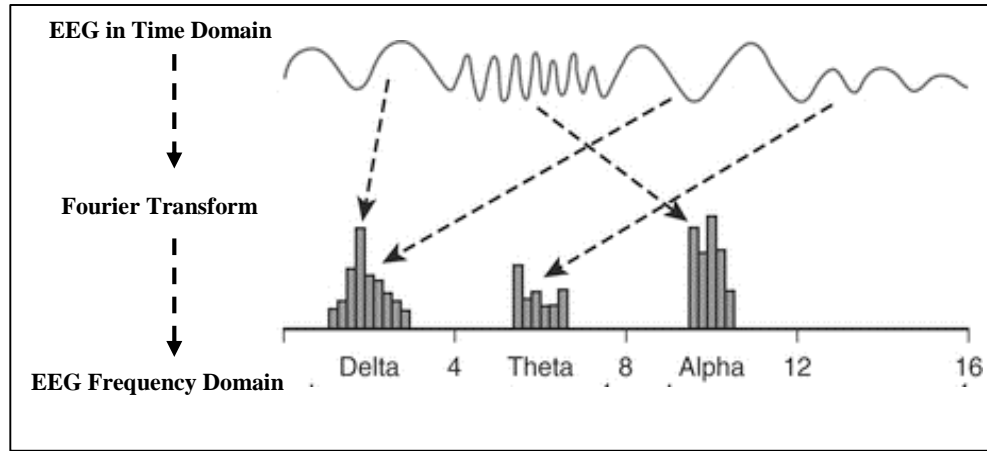


Figure 13: Transforming EEG From Time Domain To Frequency Domain

## 4.5.2 Interfaces

An interface is hardware and/or software that enables sensors to communicate with a computer. The aim of using these units in the proposed system is to receive the collected signals from the sensors and record these signals in the computer for analysis. In this research, we used two different interfaces: the LightStone Monitor v0.8, to collect the ECG signals, and the NeuroExperimenter, to collect the brain signals [127]. These interfaces are free, easy to use, and open-source. Thus, the code for collecting the data from the device could be traced and modified.

## 4.5.3 Features Extractor

In this unit of the proposed system (i.e., Unit 5), the features of the collected signals are extracted. The extracted features will have an influential role in distinguishing between activities. The purpose of this unit is to extract the features from both EEG and ECG signals. To extract such effective features for distinguishing the inside attacker during suspicious behaviors, the process of converting signals into features passes through several equations. These equations have been utilized to maximize the differences between

extracted features during normal and malicious activities. MATLAB R2014b was used to implement the features extractor unit. The output of the features extractor unit is thirteen features which illustrated in Table 8. After the output of this unit, the extracted features will be transmitted to the next unit, which is the attack assessment unit. The feature extraction process is discussed in Chapter 5, Section 5.10.

**Table 8: Features Extractor Output**

<b>Type</b>	<b>Features</b>
<b>EEG Features</b>	Delta
	Theta
	Alpha1
	Alpha2
	Beta1
	Beta2
	Gamma1
	Gamma2
	The difference of alpha1 and Alpha2 (AD)
	Total power of EEG signal
<b>ECG Features</b>	R-Peak (heart power)
	R-Peak to R-Peak Interval
	Heart rate variability

#### **4.5.4 Attack Assessment**

The attack assessment unit is responsible for distinguishing between normal and suspicious activities performed by the employee during his/her daily work. This unit uses the extracted features from EEG and ECG signals to assess whether the difference in features amounts to being classified as an internal attack. The attack assessment process is carried out using the employee's bio-signals, which have been assembled during normal activities and stored for the purpose of comparison. So, the attack assessment unit also utilizes the stored features in the signal database (i.e., Unit 7 in Figure 11).

The attack assessment unit distinguishes between normal and suspicious activities using supervised machine learning to monitor and direct the execution of a task, project, or



activity. Thus, the supervised machine-learning model is to teach the model or load the model with enough knowledge to predict decisions for future instances by training it with some labeled data. Therefore, the model can classify and predict the type of future data based on knowledge from the previous examples of labeled data.

The supervised model reduces the time for analyzing the incoming data, classifying, and decision making as the process of training the model occurs only once. The classification algorithms used in this research are implemented using Java and WEKA API, where WEKA is an open-source software that contains a set of machine-learning algorithms for research purposes [128, 129]. The output results of this unit will be transmitted to the attack detector unit. The classification algorithms used in the experiments are presented and described in chapter 5, section 5.11.

#### **4.5.5 Comparative Signal Database**

This unit of the proposed system will contain samples of the human bio-signals collected from the organization's employees during normal activities. The aim of this unit is to store the employees' normal bio-signals to be compared with the collected bio-signals to detect the insider attack. The stored signals in this unit are already converted into features. These stored features play the role of the fingerprint for the employees' physiological signals in the cases of normal activities. As explained in the previous section, this unit will be utilized by the attack assessment unit to train and build the supervised model, used for detecting malicious activities.

#### **4.5.6 Attack Detector**

The period of suspicious activity is very small compared to the time spent by the attacker in normal work. Moreover, an inside attacker, as any thief, will be keen not to be detected or tracked, so he will try to hide the trace of his malicious act by executing the attack in several stages at different times. The amount of suspicious incoming data will vary among attackers and among suspicious acts. Therefore, the amount of detected suspicious signals from the total incoming signals varies from person to person. The decision-making process must have a certain boundary or threshold. The attacker detector unit holds the threshold to judge if the incoming data is considered as an attack. The threshold will vary based on the type of protection; the lower the threshold, the higher the level of protection. The threshold is determined by the organization, depending on the nature of the organization's work and security. When the organization aims at a safer system, it should use a low threshold level, which means that the proposed system will raise alarm and suspect harmful activities even with high error rate.

#### **4.6 Evaluating the Proposed System**

To address the fifth research objective, the potential of the proposed system to differentiate between normal and suspicious activities has been evaluated. Around eight evaluation criteria have been used to measure and summarize the quality and specify the efficiency of the proposed system from different points of view.

For instance, the accuracy, recall and precision of the proposed system have been used to measure the ability of the system to discriminate between normal and malicious activities. The area under the ROC curve has been used to measure the ability of the proposed system

to achieve high accuracy in case the input data to the proposed system vary in size—i.e., the normal activity period varies from that of malicious activity.

The evaluation of the proposed system was not limited to the use of different measurement criteria, but the system was evaluated using several classification algorithms. To demonstrate the quality of the extracted features from the bio-signals, three classification algorithms have been used, each algorithm with a different learning technique. Despite the type of algorithm used for building a model during the training and performance of each algorithm, extracting good features should give better results. The proposed system has proved that it can recognize malicious activity despite using different classification algorithms and the small malicious activity period compared to that of normal activity. Evaluating the results of the proposed approach is discussed in Chapter 6.

## **4.7 Summary**

The methodology of this research has been divided into five stages, each stage intended to address one of the research objectives. The stages are the literature survey, the construction of the bio-signal data set, the feature extraction, a description of the proposed system, and the evaluation of the proposed system. A literature survey revealed considerable diversity in the methods of protection from insider threats. The outcome of the literature survey stage was the development of a categorization of the existing methods, as well as the discovery of the shortcomings of these methods, which played an important role in improving the proposed system.

When collecting the bio-signals for the creation of the data set, some criteria were taken into consideration, such as collecting the bio-signals for a sufficient number of samples to

produce reliable results and developing the scenarios of insider threats to be as realistic as possible. Moreover, the feature extraction stage is an important stage intended to convert the collected signals into measurable characteristics that play a vital role in detecting malicious activity.

Furthermore, in this chapter, the main units of the proposed system have been discussed: the sensors, the interfaces, the feature extractor, the attack assessment, the comparative data set, and the attack detector. The last stage of this research methodology—the evaluation of the proposed system—utilized several techniques to assess the potential of the proposed system for detecting insider threats. Examples of these techniques include using several metrics to assess the accuracy of the proposed system, evaluating the system using different classifiers, and using different frame sizes.

## **CHAPTER 5**

### **EXPERIMENTAL WORK**

#### **5.1 Overview**

This chapter mainly discusses in detail two stages of the research methodology: the construction of the bio-signal data set and the feature extraction (i.e., stages two and three in the research methodology). Also, this chapter partially deals with the proposed system (i.e., stage four of the research methodology), by presenting the machine learning algorithms used in the attack assessment unit.

On the database construction stage, this chapter presents the experiments' preparation process as well as the experimental environment and discusses the experimental scenarios for collecting the human bio-signals. Also, this chapter presents the devices that are used for assembling the bio-signals and discusses components of the collected brainwaves and ECG signals, for instance Delta, Theta, Alpha Beta, Gamma, and the heart rate variability.

On the other hand, this chapter discusses the feature extraction stage by illustrating the procedure of extracting the features from the collected bio-signals. In addition, it demonstrates the processes and equations that have been used in constructing the final feature-frame. Also, this chapter discusses the field of machine learning and presents the classification techniques that are utilized by the attack assessment unit of the proposed system.

## **5.2 Experiment Setup**

The main objective of our experiments is to distinguish between insider threats and benign activities by utilizing bio-signals. To achieve this objective, the participants' brain and ECG activities were recorded while they were performing two different scenarios: normal and malicious. The duration of each scenario is between eight to twelve minutes. The first scenario simulated the normal activities of the employee during his daily work. In contrast, the second scenario simulated as much as possible real insider threats performed by authenticated users who do not have any authorization to conduct such activities. To this end, during the second scenario, we sought to fulfill the following criteria:

1. The participant is aware that his malicious work is not permitted and that he is responsible for any consequence of that work.
2. The participant is the only person who has the decision to conduct malicious activity.
3. The participants have a suitable environment for conducting malicious actions, but where they are not aware that their malicious acts are part of the experiments and are being monitored by the researcher.

## **5.3 Experiment Environment**

The experiments were conducted at the campus of Hadhramout University, where the researcher works as a lecturer. The necessary approvals were obtained from the concerned authorities at Hadhramout University as well as the colleges that provided us with facilities for using their resources and their own labs to conduct these experiments. Moreover, in

order to conduct and record the malicious activities during the second scenario, the university offered the necessary facilities to obtain mock exams and fake documents with the collaboration of the subjects' professors.

To provide the appropriate environment, the experiments were conducted in a dedicated lab. Figure 14, shows the environment of the experimental lab. The participants sit alone in the lab, and the researcher monitors the computer's screen and recording the bio-signals while he was sitting in a separate room (i.e., Lab technician room).

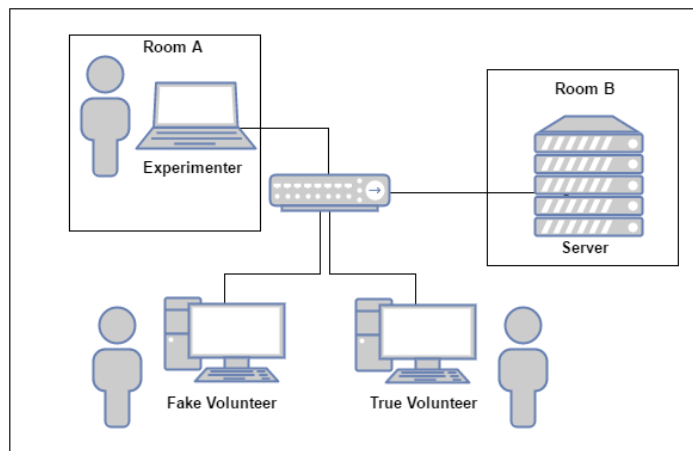


Figure 14: The Experimental Lab Environment

## 5.4 Experiment Scenarios

To simulate a realistic insider threat scenario, we tried to provide an appropriate environment that would attract the participants to perform a malicious activity without their knowledge that the attack was part of the experiment. To achieve this, in agreement with the colleges where the experiments were carried out, we announced fake goals for this experiment. It was announced that the main goal of this experiment was to develop software that would allow users to write on a computer using their bio-signals and brain

reactions without relying on a keyboard. The announcement claimed that bio-signals and brain reactions would be recorded while the participants were writing simple paragraphs using any word editor; the recordings would help the alleged software to distinguish the changes in bio-signals during the writing of each character or number. Moreover, the participants were notified that the experiment contained two scenarios. In the first scenario, the participants would write Arabic paragraphs, whereas in the second scenario, the participants would write English paragraphs. Moreover, the Arabic and English scenarios would take place on different days.

Each time, the scenarios were conducted on a group of two participants. During the normal scenario, the bio-signals were collected from both participants. But during the malicious scenario, one participant (i.e., a fake participant) was hired by the researcher to play the role of a motivator as we would discuss in the following paragraphs.

#### **5.4.1 First Scenario: Normal Activities**

In this scenario, the participants were given some paragraphs written by hand in **Arabic**, which included some simple questions asking the participants to do the following tasks:

1. Use any word editor to rewrite the given Arabic paragraphs.
2. Answer the simple questions and write the results.
3. Save the document in a specific folder on the server machine (the folder path was given).

The bio-signals recorded during this scenario were stored in the data set as normal activities. This scenario included three regular office tasks: typing paragraphs, thinking



about solving problems, and using the organization's network to save the document on the server.

#### **5.4.2 Second Scenario: Malicious Activities**

In the second session, the participants were asked to rewrite some paragraphs already written by hand in **English**, which included some simple questions requesting that the participants do the same tasks as in the first scenario.

Prior to the second scenario, some extra folders were added to the server, exactly in the path where the participants would store their documents. These folders contained some fake private data that would pique the participants' curiosity to read them or to copy these data from the remote server to the local machine or to their flash drives. Examples of the counterfeit data are exam models and solutions, lecturers' private data, lists of staff salaries and allowances, and a list of candidates for bonuses. The variation in folders' titles was intended to provoke the curiosity of participants of different ages covered by the experiments. All counterfeit data were obtained with the consent of the parties involved; for example, the old exams were obtained from the actual professors of the subjects.

To make the participants spend a long time searching for the data, the folders did not contain the counterfeit data directly. The counterfeit data were saved in subfolders that were scattered in other subfolders. Moreover, in order to make the process of copying the entire folder to the local machine take a long time, the size of the subfolders that contained the fake data was increased by compressing these subfolders with additional large files. The goal of making the process of searching and copying the data quite long was to record

as many bio-signals as possible from the participants while they were doing such unauthorized activities.

To ensure that the participants had observed the presence of the fake files, the fake participant who was hired by the researcher would play the role of a motivator, especially in the second malicious scenario. In case the real participant did not notice the existence of the files, the fake participant would pretend that he had discovered some important files on the server while he was storing the document. The real participant would be notified about the existence of the files cautiously and confidentially by the fake participant. The bio-data collected from the fake participant were ignored. To maintain the secrecy of the actual goal of the malicious scenario, it was conducted after completing the normal scenario, and the fake participant was asked to play the role of motivator only one day before conducting this scenario. The fraudulent participant was carefully chosen from the same age and academic level for the real participant.

The second scenario simulated a real insider threat in which the user had authentication to access some files on the network, but he did not have any authorization to reach all folders, read the files, and copy private data. In addition, the user had the choice to conduct the malicious activity or to ignore the fake files. During the second scenario, the network devices were monitored using Net Monitor for Employee [130]. This software allowed the researcher to monitor the computer screens in the network, which helped to determine the time that the participant began to conduct malicious activities. When the real participant did not conduct any harmful activities, the collected bio-signals of that participant were ignored and were not stored in the data set as suspicious signals.

## **5.5 Bio-signals Data set**

The experiments were conducted on 172 participants. In the normal scenario, all participants' bio-signals were collected. However, during the second scenario, only 84 participants noticed the existence of the fake exams and decided to view or copy these files. These acts were considered malicious, so the bio-data of these 84 participants were stored in our data set and will be used as incoming data. Thus, the data set contains the bio-signals of only 84 participants, of whom 43 were male and 41 were female. All participants were in the range of 19 to 36 years old.

## **5.6 Ethical Considerations**

The experiments were conducted with the approval of Hadhramout University, which provided such facilities as official letters to the colleges where the experiments were performed. Also, written permission was obtained from the colleges to carry out these experiments in their laboratories. Participating in the experiments was voluntary, where the participants had rights to attend the scenarios when they had free time. To maintain the secrecy of the second scenario, the real objective of the experiment was revealed to the participants after completing the entire experiment. Participants were informed about their right to participation and use their bio-signals for that objective. They were also informed that their personal data would be protected. The consent of 84 participants was obtained to use their bio-signals for research purposes.

## 5.7 Experiment Devices

Two types of devices were used for collecting the EEG and ECG bio-signals: the NeuroSky MindWave and the Wild Divine. The devices have many advantages such as being small, wearable, easy to connect to a computer, comfortable, cheap, have several support programs, and having been used in several studies [131–133]. The NeuroSky transmits the signals wirelessly to the computer, whereas the Wild Divine uses the universal serial bus (USB) to connect with your computer.

### 5.7.1 The NeuroSky MindWave

The NeuroSky is a headset consisting of two sensors located on the forehead and earlobe. The NeuroSky monitors the brain activity; particularly those signals pertain to the attention. Unlike the medical EEG devices that use a paste or electrolyte gel to improve the contact between the scalp and the electrode, NeuroSky uses dry sensor technology which sensitive enough to pick up electrical activity without using the paste. [126]. Figure 15 illustrates the NeuroSky headset.

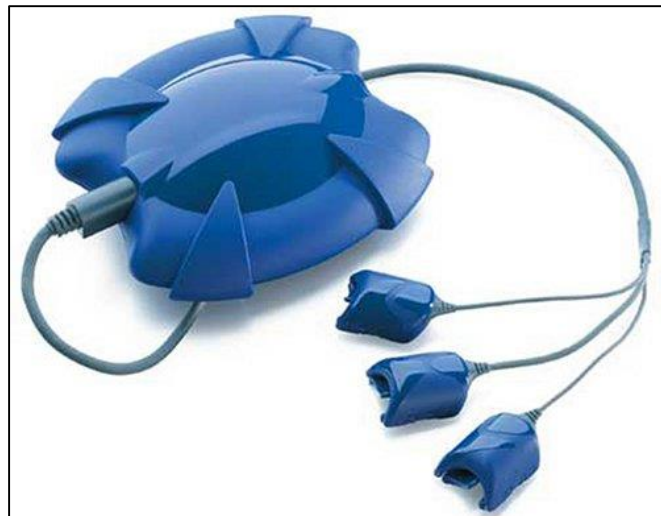


Figure 15: NeuroSky MindWave Headset

To improve EEG clarity, the device collects EEG brainwaves by calculating the potential difference between the forehead and the ear electrodes. The reference electrode was chosen in the earlobe because there's none EEG activity in the earlobe; which leads to calculate the voltage difference between accurately [126].

### **5.7.2 Wild Divine**

The Wild Divine device is a biofeedback component that can measure human biological data from the autonomic nervous system, which responds to a person's inner world of thought, perception, and indicators of positive emotions, like excitement, and negative ones, like nervousness. The Wild Divine device has three finger rings, as illustrated in Figure 16. To measure human bio-signals, three fingers, i.e., ring, middle, and index, are inserted into the rings. The middle ring measures HRV, whereas the other two rings measure the skin conductance level [134].



**Figure 16: The Wild Divine Device**

## **5.8 Brain Waves**

### **5.8.1 An Electroencephalogram (EEG)**

The human brain consists of billions of neurons which emit electrical impulses to communicate with each other. The changes of electrical impulses are represented in the form of brainwaves that are strong enough to be recorded by using electrodes on the scalp with a technique called electroencephalography (EEG) [135]. Researchers have made considerable strides to link the brainwaves with memory, consciousness and even certain diseases such as Epilepsy [136–138].

### **5.8.2 Brainwaves: Types and Functions**

Brainwaves have different patterns based on their frequencies and amplitudes. The frequency of a brainwave is directly proportional to the number of times that neurons are firing per second, whereas the amplitude is directly proportional to the number of neurons fired synchronously. There are five types of brainwave patterns, namely Delta, Theta, Alpha, Beta and Gamma [139, 140]. Figure 17 illustrates the frequency spectrum of normal EEG signals.

#### **5.8.2.1 Delta Brainwaves**

Delta brainwaves have the frequency range of (0.5 to 4 Hertz), which is the lowest frequency among the different types. It has relatively high amplitudes in the range of (75 – 200 $\mu$ V). Delta activities are typically linked to a deep sleep or unconscious state and are predominantly found in human beings who hold a strong sense of empathy and intuition.

Moreover, Delta brainwaves allow us to access subconscious activity if generated in the waking state [141, 142].

### **5.8.2.2 Theta Brainwaves**

Theta waves are slightly higher in frequency than Delta waves. Theta waves travel between neurons in the frequency range of 4 to 8 Hertz. They are often associated with the meditative or deep relaxation state of consciousness. The deeper the meditation, the higher the creativity and the faster the learning displayed by a person. Theta activities increase in the daydreaming state when a person dreams consciously or is in a light sleep. They are linked with highly monotonous daily tasks such as walking on the same road, wearing clothes and combing hair [142].

### **5.8.2.3 Alpha Brainwaves**

Alpha waves are formed when neurons fire signals at frequency ranges of (8 to 13 Hz), which is faster than in the case of Theta waves. The typical amplitude of Alpha waves is around (50 $\mu$ V peak-peak). Alpha waves are usually dominant in normal situations, especially with closed eyes, and in the relaxed but wakeful state. On the other hand, Alpha waves are attenuated when the eyes are open and when a mental effort is required to solve difficult problems [135, 142].

### **5.8.2.4 Beta Brainwaves**

Beta waves travel between neurons at frequency ranges of 13 to 30 Hertz. Beta waves are dominant with the state of awareness, concentration and the active state of wakefulness when the eyes are open. Beta activities are associated with a state of increased alertness

and a focus on the task at hand. Furthermore, when Beta activity increases, the brain is working efficiently and can develop new ideas and generate solutions [141, 142].

### 5.8.2.5 Gamma Brainwaves

Gamma waves have the highest frequency compared with the rest of brainwaves. Gamma waves travel between brain neurons in the frequency of more than 30 Hertz. Gamma waves are associated with the high energy moments, the concentration and focus. Moreover, Gamma waves are linked to language processing, memory, and regional learning [142].

It is worth mentioning that, always the five brainwaves are generated together. However, when one of the brainwaves dominates, it means that the other brainwaves are weak but still can be distinguished. Thus, the strength of the brainwave is related to the activity type and a state of the person [143].

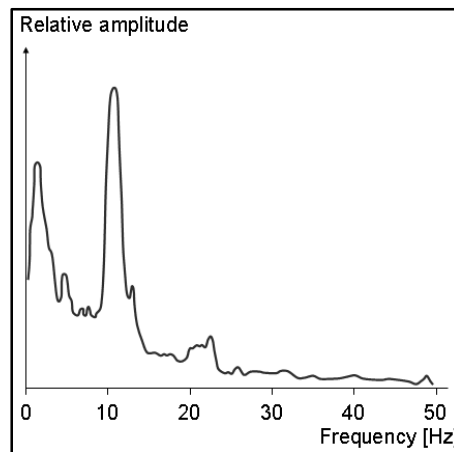


Figure 17: Frequency Spectrum of Normal EEG

## 5.9 An Electrocardiogram (ECG)

The human heart is capable of generating electrical signals, which it uses to create the muscular contractions that are needed to move the blood through the body's blood vessels.



In fact, physicians can study and analyze the way that the heart produces electrical signals and determine different types of abnormalities that might exist within the heart simply by using a tool known as an electrocardiogram (ECG). An electrocardiogram measures the electrical signal generated by the heart. ECG is not a tracing of a single action potential but an amalgamation of the many action potentials that constitute the electrical activity of the heart [144, 145].

### 5.9.1 Types of ECG Wave Components

In this research, three components are extracted from the ECG wave to be used in the proposed system: heart rate variability, R peaks, and the interval between R peaks.

#### 5.9.1.1 Heart Rate Variability HRV

Heart rate is controlled by the balance of sympathetic and parasympathetic of the autonomic nervous system. Heart rate variability HRV is the variation of heart period or inter-beat interval. It is the time between successive R peaks in the electrocardiogram as illustrated in Figure 18. HRV is measured by calculating the standard deviation (SDNN) of all normal to normal inter-beat intervals (i.e., all normal R to R intervals) [146].

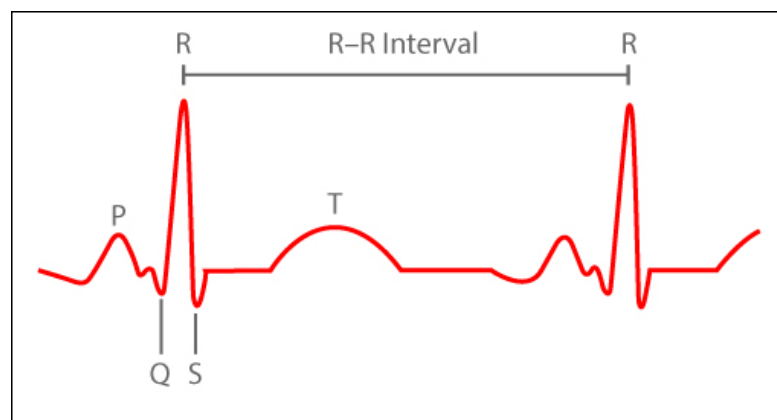


Figure 18: R Peak To R Peak Interval of ECG [147]

Heart rate variability is used as an indicator of mental workload. Moreover, it can be used as a measure of emotional response to certain stimuli. Clinical research has proved that the heart rate variability HRV level becomes low when a person experiences high stress, whereas the increases in HRV level is an indication of a high resilience. Figure 19, shows low and high HRV [148].

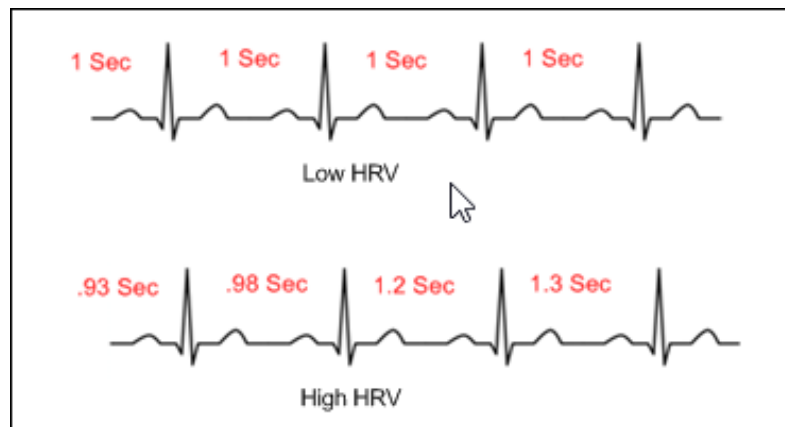


Figure 19: Heart Rate Variability (HRV) [149]

### 5.9.1.2 R Peak (Heart Power)

The electrocardiogram R wave which is denoted by QRS complex in Figure 18, is used in the analysis of the irregularities of heart rhythm. Furthermore, the R wave has a vital role in determining HRV. The R wave is a positive deflection upwards that represents depolarization of the left ventricle and myocardium. The maximum amplitude in the R wave is known as R peak. R peak is measured in millivolts [150, 151].

### 5.9.1.3 R to R Interval

The R–R interval represents the interval between two consecutive R peaks, i.e., the interval between two heartbeats. The R–R interval is an important component to differentiate between malicious and normal activities. We can assume that the heart rate is affected when a person conducts malicious activities, and the R–R interval is a good indicator of the

amount the heart rate is increasing. The R–R interval is used to calculate the heart rate in beats per minute by multiplying by 60 the invert number of R–R interval per second [144].

Figure 18, illustrates the R-R interval.

## **5.10 Feature Extraction**

Features are the characteristics of signals that have the primary role in identifying the major changes these signals. Features will increase the effectivity of the raw signals for distinguishing the malicious activities [152]. This section discusses Feature Extraction which is the third stage of the research methodology (Chapter 4, Figure 7). It presents the signals' preprocessing, the procedures of extracting features from the raw signals, and the equations used for converting the raw signals into features. Moreover, this section presents the process of producing the final feature-frame that used by the proposed system to detect the malicious activities. Figure 20 illustrates the main diagram of the features extraction procedures; starting with collecting of raw signals to the process of configuring the final feature-frame.

### **5.10.1 Signal Preprocessing**

The preprocessing is taking place by two units of the proposed system, the sensors and interfaces. The sensors collect several signals, separate them, and send these signals to the interfaces. The interfaces recorded the received signals and utilizing these signals for calculating additional parameters.

The device collects EEG brainwaves by calculating the potential difference between the forehead and the ear electrodes. The signals are amplified by the device 8,000 times in order to improve the faint EEG waves. Moreover, the EEG signals are filtered using low-

and high-pass filters to correct any possible distortion. Then, the standard fast Fourier transform (FFT) is performed on the filtered signals to convert them to the frequency domain, i.e., EEG bands [126].

On the other hand, The ECG signals are processed and converted from analog to digital signals in the ECG sensor (i.e. Wild Divine device). The device extracts the main components of the signal such as heart rate, the peak of ECG wave and the intervals between peaks [134]. Then those components are sent to the ECG interface to be recorded and utilized for calculating additional features such as the intervals between signals, and HRV heart rate variability.

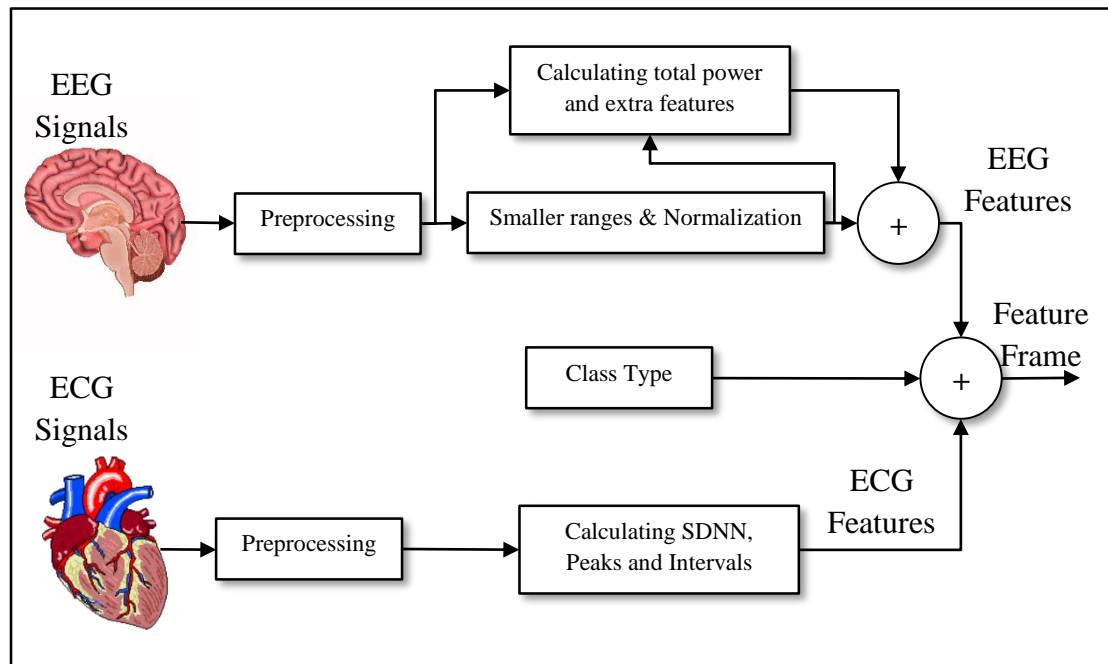


Figure 20: The Main Diagram of The Features Extraction Process

### 5.10.2 EEG Features

The proposed EEG features are based on three factors which are: dividing the EEG waves into smaller ranges to extract features, Normalizing the features and extract additional features from the influential EEG waves.

### 5.10.2.1 Small Ranges of EEG Waves

The NeuroSky device sends the brainwaves' power in the form of sessions of one-second duration. In this research, the frequency ranges of brainwaves were divided into smaller bands as illustrated in Table 9. Since the proposed approach uses machine learning techniques to detect malicious activities, dividing the brainwaves into smaller frequency bands would increase the number of features used in the detection. Moreover, although the brainwaves would not have the same impact on the results, using smaller frequency bands allows for an extensive study of each part of the brainwave and its impact on the detection of internal threats.

**Table 9: Frequency Ranges of EEG Bands**

<b>EEG Band</b>	<b>Frequency Range</b>
Delta	1-3 Hz
Theta	4-7 Hz
Alpha1	8-9 Hz
Alpha2	10-12 Hz
Beta1	13-17 Hz
Beta2	18-30 Hz
Gamma1	31-40 Hz
Gamma2	41-50 Hz

### 5.10.2.2 Normalization

In the normalization stage, we normalized each EEG component (i.e., Delta, Theta, Alpha1, Alpha2, Beta1, Beta2, Gamma1, and Gamma2) in a single session using Equation 5.1. Normalizing the power of EEG bands has two major advantages to the efficiency of the extracted features [153]; which are:

1. Reduce the impact of the signal variability.

2. Allow the proposed features to distinguish between the active and idle EEG activities by increasing the separation between these activities.

$$NW_{x,i} = \sqrt{\frac{w_{x,i}}{tot\ pwr_i}} \quad (5.1)$$

$NW_{x,i}$ : Normalized EEG component  $x$  in session  $i$ .

$totpwr_i$ : the total power of session  $i$ .

### 5.10.2.3 Calculating Additional Features

Two more features were calculated during this stage: the total power and the alpha difference (AD). The total power ( $totpwr$ ) of each EEG session is the summation of the power of the eight components in that session. Unlike the normalized EEG signals, which can facilitate comparison among the EEG components, the total power provides an opportunity to study the overall change in the EEG signals by comparing the total power of signals during both malicious and normal scenarios. Equation 5.2 demonstrates the calculation of total power.

$$totpwr_i = \sum_{x=1}^8 w_x \quad (5.2)$$

$i$ : the frame number.

$x$ : the brainwave number.

$w$ : the brainwave power.

Moreover, Alpha wave plays an important role in determining the transition of a person's condition from the relaxing or calm state to the thinking and concentration state [141, 154].

The concentration level of the participant will change when conducting the unauthorized actions. So, Alpha wave can provide more information when the participant starts the harmful activities. Therefore, we studied the effect of the Alpha difference (AD) in the training and testing data. From the collected bio-signals, we found the percent change of

Alpha difference increased by around 46.6% more than Alpha1 in the training and testing data, and it increased by around 66% more than Alpha2. So, using AD as an additional feature would improve the accuracy of the proposed system. The AD and the percent change illustrated in Equations 5.3 and 5.4.

$$AD = |\text{Alpha2} - \text{Alpha1}| \quad (5.3)$$

*Alpha1 and Alpha2 are normalized*

$$\text{Percent Change} = \frac{(\text{Average}(AD))_{\text{Test}} - (\text{Average}(AD))_{\text{Train}}}{(\text{Average}(AD))_{\text{Train}}} \times 100\% \quad (5.4)$$

### 5.10.3 ECG Features

In this research, three features were extracted from the ECG waves: The R peak, the R–R interval, and heart rate variability. The R peak is the maximum amplitude of the electrocardiogram (ECG) signal, whereas the term interval refers to the time difference between the consecutive R peaks [144]. Moreover, heart rate variability, or SDNN, is the standard variation of heart period, which is calculated using Equation 5.5. It is worth mentioning that the ECG interface calculates the value of SDNN once every minute. Therefore, every 60 seconds, the application calculates the standard deviation of the intervals from the last 60 seconds.

$$SDNN = \sqrt{\frac{1}{N-1} \sum_{i=2}^N (RR_i - \overline{RR})^2} \quad (5.5)$$

*RR<sub>i</sub>: denotes the time from *i*th to the *i*+1st R peak.*

*RR: the average interval.*

*N: intervals in total.*

### 5.10.4 Feature Frame

The final feature frame consists of a combination of the EEG and the ECG feature frames. The EEG feature frame contains 10 features, which are illustrated in Figure 21. Only the total power (totpwr) is not normalized, whereas the other nine features are normalized.

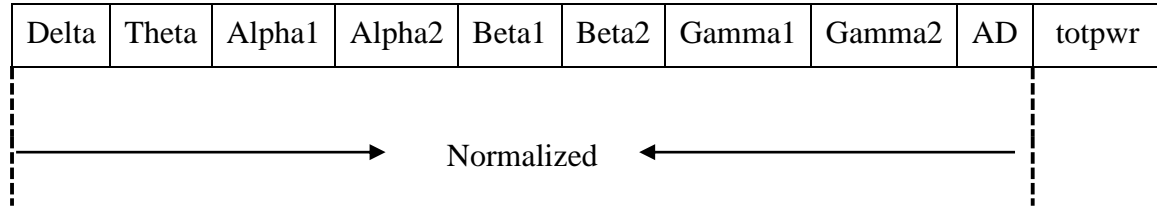


Figure 21: EEG Feature Frame

On the other hand, the ECG feature frame consists of three features: R peak, interval, and SDNN. The class field contains the label that is used to train the machine learning algorithm. The label or the class determines the type of frame that has been extracted from either the normal or the suspicious activities. Figure 22, illustrates the final feature frame.

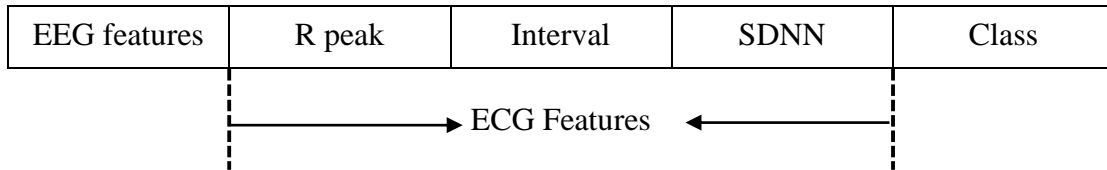


Figure 22: Feature Frame

*R peak: the maximum amplitude of ECG R peak in second.*

*Interval: the time between R to R peaks.*

*SDNN: standard deviation of all R to R intervals.*

## 5.11 Machine Learning

Machine learning is an area of computing that is improving the ability of the computer systems to learn from past experience. Tom M. Mitchell, who is a professor at Carnegie Mellon University, defined the machine learning as:



*"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."* [155]

In other words, the computer program has learned when it utilizes past experience to improve how it performs a particular task. A machine learning program is fairly different compared to a normal computer program, because in the normal computer program, all the data and the parameters needed to perform a certain task have already been defined by its programmer.

Based on the learning technique, the field of machine learning can be divided into categories: supervised, unsupervised, and reinforcement learning [156]. Supervised learning is where the machine is trained and taught using labeled data, i.e., tagged data with the correct answer, to utilize this past experience to give an outcome for new, previously unseen data. Examples of supervised learning are classification and regression. On the other hand, unsupervised learning is where the machine is trained with unlabeled or untagged data to draw inferences and create a model for the data set. Examples of unsupervised learning are clustering and frequent-patterns. The third category of machine learning is reinforcement learning, which deals with how the software agents can maximize some concept of cumulative reward by taking necessary actions in a specific environment. For example, a game agent can utilize cumulative effect in order to create a winning strategy by playing the game many times [22, 157].

Detecting the attacker from inside organizations, which is the main problem of this research, tends to be a classification problem, where the machine should classify the

incoming data as malicious or normal. Therefore, in this research, supervised learning will be utilized where the machine will be trained using labeled data with correct answers to teach the machine how to classify new, previously unseen data. In this research, three classification algorithms were used: Random Forest, Support Vector Machine and Neural Network. Each algorithm has its technique of learning from labeled data.

### **5.11.1 Random Forest**

The Random Forest (RF) method uses a random selection of data and a random selection of variables to create many decision trees. The random selection of data leads to the creation of several subsets of data that have different sizes and may have overlapped data, so the sizes of the decision trees will be random. On the other hand, the random selection of variables or features leads to the creation of decision trees with a different number of variables [158]. So, the RF technique consists of many decision trees. The idea behind the correctness of the RF method is that the huge number of trees leads to a better prediction because of the following:

1. The decision tree is usually correct, but it may have some parts of the data which are incorrect.
2. The huge number of decision trees will never have the same incorrect parts of data.

It worth to mention that each decision tree will classify the new incoming data and vote for this classification. The RF will choose the classification that has the most votes over all the other trees. Figure 23 the RF procedures for selecting data and variables, and voting for the results. From the Figure, it can be noted that the sizes of the selected data and the number of variables is varied. This variation leads to creating decision trees of different sizes. Also,

we can notice that the number of decision trees which vote for the suspicious activity is more than the trees which vote for the normal activity; in this case, the incoming signals will be classified as suspicious [158, 159].

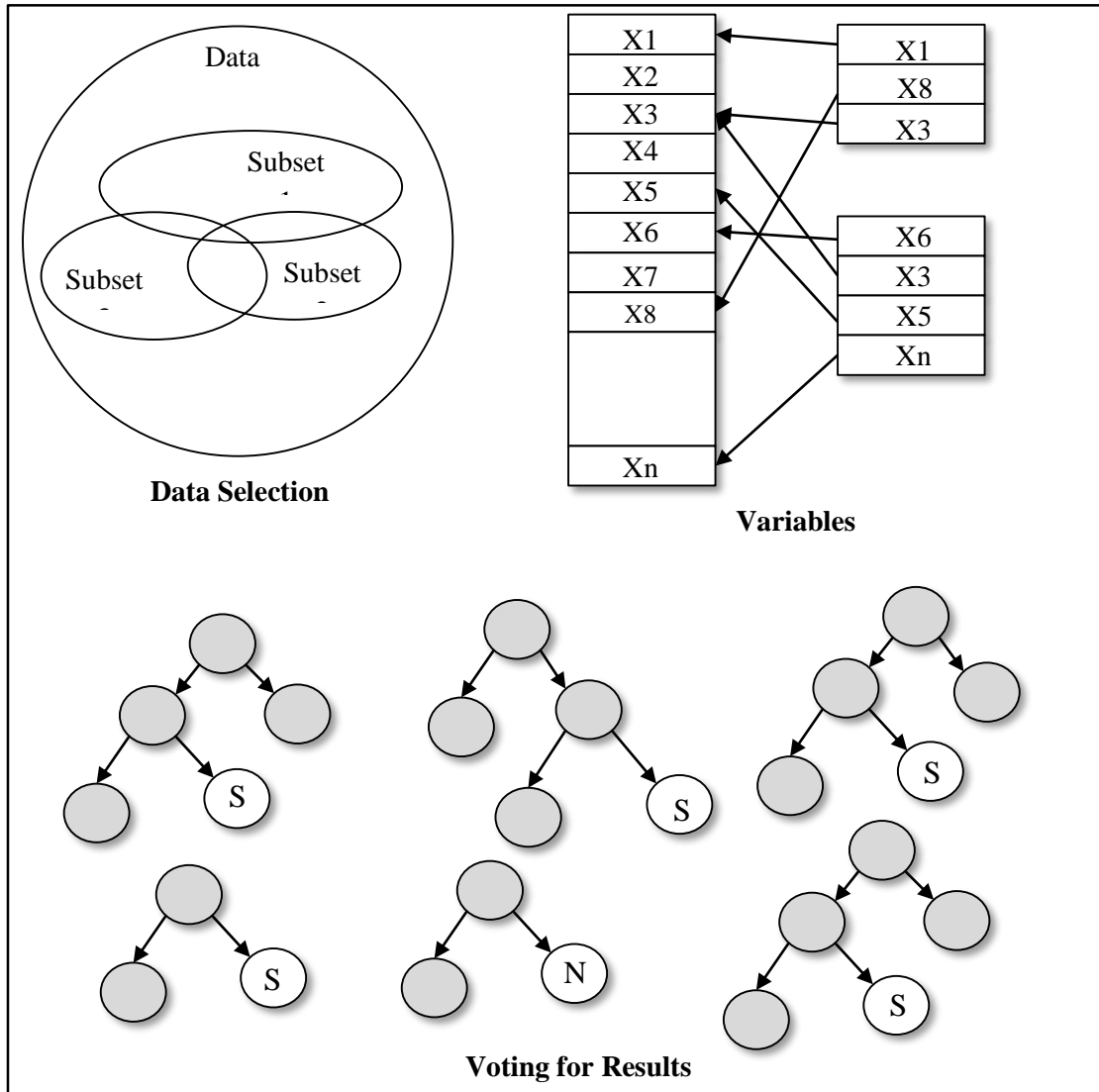


Figure 23: Random Forest Procedures

### 5.11.2 Support Vector Machine

Support Vector Machine (SVM) uses support vectors to draw the decision boundary (hyperplane) to segregate the classes of data. The best hyperplane is one that leaves the maximum margin from the classes. The support vectors are the extremes of the data sets

where the SVM basically implies that only these extreme points in the data set distinguish between different classes, whereas the other training examples are ignored [160, 161].

The SVM algorithm segregates the classes of data by drawing the decision boundary or a hyperplane. The best hyperplane is the one that leaves the maximum margin from the classes of data. Therefore, the SVM algorithm uses the support vectors which are the extremes of the data sets to draw the hyperplane. The SVM algorithm implies that only these extreme points in the data set distinguish between different classes, whereas the other training examples are ignored. Figure 24 illustrates the process of selecting the best hyperplane. This Figure shows how the SVM algorithm segregates between two classes, which are the black and white squares [160].

From Figure 24, the SVM selects the extreme points of the classes, i.e., the extreme elements of white and black squares, to draw the hyperplanes. Also, it can be noticed from the same Figure that the margin  $Z2$  of the second hyperplane  $H2$  is wider than the margin  $Z1$  of the hyperplane  $H1$ . Therefore, the second hyperplane  $H2$  can distinguish between the classes of data better than the first hyperplane  $H1$ .

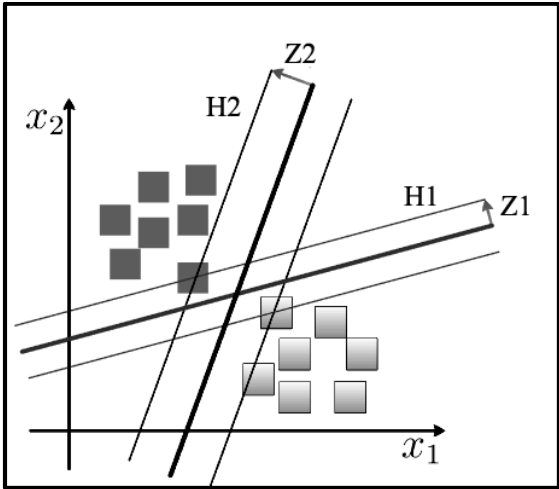


Figure 24: Selecting The Best Hyperplane

### 5.11.3 Neural Network

The principle of neural networks lies in trying to emulate how the human brain works. A neural network is made up of several nodes, which are called neurons. The connections between these neurons are called synapses. So, information is passed between neural network nodes via the synapses. When a neural network node receives information, it can process it and transmit it to the connected nodes. The connections or synapses have values that represent their weights. The higher the weight of the connection, the more important the information of that connection. Therefore, when a node is stimulated by several nodes, it can decide which node is more important [162, 163].

The neural network consists of several layers of nodes, where the information travels from the input layer to the output layer of the network. The layers between the input and output layers are called the hidden layers. Figure 25, illustrates the structure of neural network.

Each node of the neural network judges its input and sets its value using a transfer function, which is a simple math equation that allows the node to accept or reject the triggers. The transfer function produces a combination value from the trigger value and the current value of the node [162, 164].

The neural network learns through a process called backpropagation. The network starts with random connection weights, then calculates a set of outputs for a given set of inputs. These outputs are compared with the desired output, but because the network starts with random connections, there will be an error, or a difference between the outputs. So, the neural network will adjust the connection weights to reduce the error. This process is

known as backpropagation. In it, the output nodes tell the nodes in the previous layers about the error and try together to adjust their connection weights to reduce the output error [165].

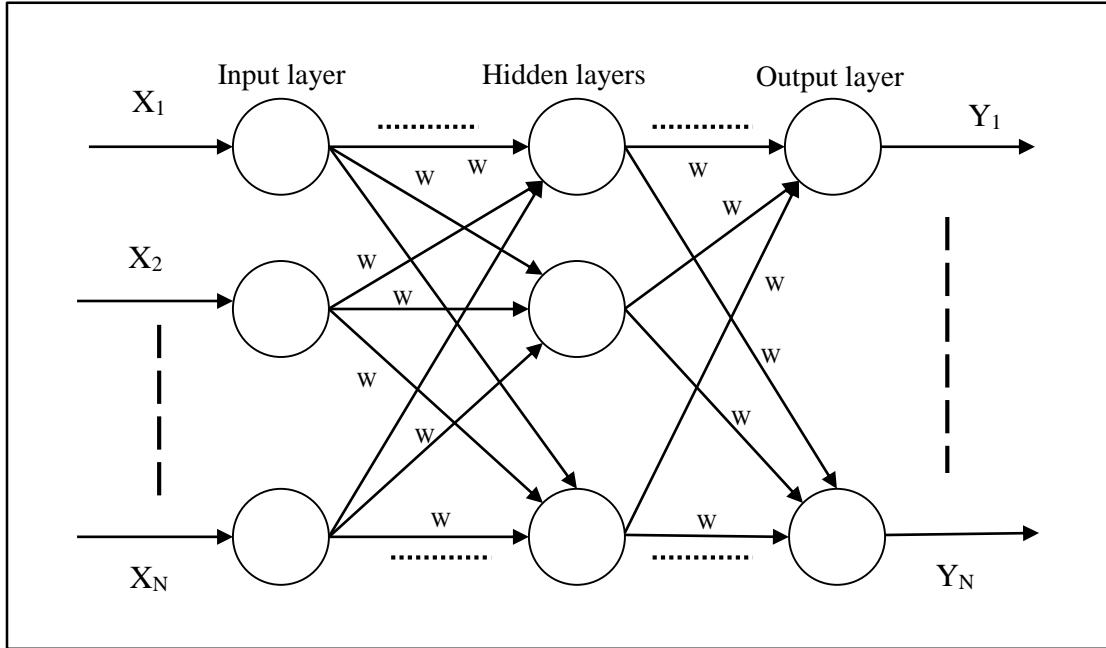


Figure 25: Neural Network Structure

## 5.12 Summary

Collecting the human bio-signals are essential in two phases of this research: Building the bio-signals data set and feature extraction. So, two experimental scenarios have been developed to collect the bio-signals from volunteers during the normal and suspicious activities. The proposed scenarios aimed at creating a realistic environment for the normal and insider threats where the volunteers have the decision to conduct the malicious activity without knowing that their malicious activities are part of the experiment.

Moreover, the procedure of feature extraction from the collected human bio-signals have been illustrated in this chapter, including presenting the EEG and ECG signals, demonstrating the devices that have been used for collecting the signals and discussing

equations that have been used in extracting the features. The extracted features have been illustrated in the form of a frame, which contains both EEG and ECG features. The feature frame contains ten EEG features and three ECG features, in addition to a class field which was used as a label for the machine learning.

Since the classification is a crucial phase in the proposed system for distinguishing between the insider threats and normal activities, three machine learning algorithms that have been used to evaluate the proposed system, these algorithms are the random forest, the support vector machine, and the back propagation neural network. Each of these algorithms has its learning method and model. Using several machine learning algorithms is intended to assess the impact of these algorithms on the proposed system results.

## CHAPTER 6

### RESULTS AND EVALUATION

#### 6.1 Overview

This chapter presents the research hypotheses and the procedures to address them. Moreover, it discusses the results and efficiency of the proposed system using several metrics. Selecting those metrics takes into account the evaluation of the proposed system from various aspects such as the accuracy of detecting the insider attacks, and to what extent this accuracy is correct.

Also, this chapter deliberates the feasibility of using only the brainwaves to identify the insider threats, where the accuracy of the proposed system would be examined using only the EEG features. Furthermore, it clarifies the most influential features on the results.

Moreover, this chapter discusses the performance of the proposed system to detect malicious threats by using three of the machine learning classification algorithms. The aim of using different algorithms to illustrate that the results are not affected by the models and learning methods of several machine learning techniques.

#### 6.2 Hypotheses

To address the main research problem, detecting insider threats using human bio-signals, the proposed approach will be evaluated by assessing its ability to distinguish between



normal and malicious activities. To achieve this goal, the following main null hypothesis has been developed:

**H0)** *The proposed approach has no potential to distinguish between the normal and malicious activities by using human bio-signals; thus, it is not able to detect the insider attackers.*

In order to address the main null hypothesis, the following support hypotheses have been developed:

**H0-1)** *The extracted features from the brainwaves (EEG signals) have no potential to detect an insider attacker by differentiating between normal and suspicious activities.*

To address the hypothesis H0-1, we have generated only the EEG features frame, which is illustrated in Figure 21. Then, using machine learning techniques, the extracted EEG features have been tested for their ability and efficiency to distinguish between the two experimental scenarios, i.e., malicious and normal activities.

**H0-2)** *The extracted features from the electrocardiogram (ECG) signals have no potential to detect an insider attacker by differentiating between normal and suspicious activities.*

To address the second null hypothesis H0-2, we have generated the full features frame (i.e., the frame that contains both EEG and ECG features), as illustrated in Figure 22. Then, utilizing machine learning techniques, the extracted features (i.e., ECG and EEG together) have been tested for their ability and efficiency to distinguish between the two experimental scenarios. Moreover, in order to study the effect of the ECG features on the results, the

results using the full frame have been compared with the EEG frame results. The full frame will be abbreviated as EEG+ECG.

To determine the results' significance, and to test the validity of the null hypotheses, the Paired t-Test has been used [166, 167]. The t-Test uses when the observation on the two populations are collected in pairs. The following expression has been used to test the validity of the main null hypothesis using the Paired t-Test:

$$t_0 = \bar{d}/(S/\sqrt{n}) \quad (6.1)$$

$\bar{d}$  and  $S$  are the sample average and standard deviation of the difference.  $n$ : Sample number. The one tail t-Test was used with significance value  $\alpha$ : =1% as illustrated in Figure 26. So, the null hypothesis will be rejected if the t-Test is in the rejection area.

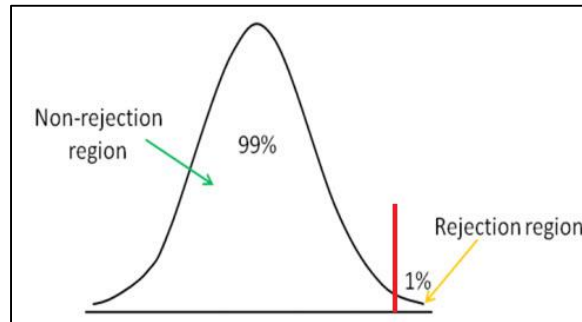


Figure 26: One Tail t-Test

### 6.3 Evaluation Metrics

Evaluation metrics play a vital role in assessing the performance and efficiency of the proposed system for distinguishing between benign and malignant bio-signals. Therefore, several metrics have been used each of which evaluates the system for different objectives.

Since the research problem—identifying the malicious activities—is a classification problem, the confusion matrix is utilized. The confusion matrix is a technique that

evaluates the performance of the classification model and illustrates the errors of the classification problem by summarizing the prediction results. The confusion matrix of the binary classification problem that distinguishes between two different groups is represented by a 2-by-2 matrix [166] as illustrated in Table 10. Also, the confusion matrix plays an important role in deriving several of the following evaluation metrics.

**Table 10: Confusion Matrix**

Actual Class	Predicted Class	
	Negative	Positive
Negative	TN	FP
Positive	FN	TP

**True Negative (TN):** TN are negative cases that are correctly predicted to be negative [166]. In this research, TN represents the bio-signals that are collected during a first scenario while the participants are doing a regular activity, which is correctly predicted by the classifier as a regular activity.

**True Positive (TP):** TP are positive cases that are correctly predicted to be positive [166]. TP represents the bio-signals that are collected during a second scenario (suspicious activity), which is correctly predicted as a suspicious activity.

**False Negative (FN):** FN are positive cases that are incorrectly predicted to be negative [166]. FN represents the bio-signals that are collected during a suspicious activity, which is incorrectly predicted as a regular activity.

**False Positive (FP):** FP are negative cases that are incorrectly predicted to be positive [166]. FP represents the bio-signals that are collected during a regular activity, which is incorrectly predicted as a suspicious activity.

### 6.3.1 Accuracy

Accuracy is the ratio of correctly classified bio-signals (TP and TN) to the total number of bio-signals collected from the participant. It Indicates the degree of conformity and correctness of the results obtained when compared to the true value [166]. The accuracy is calculated using Equation 6.2.

$$Accuracy = \frac{\text{correct}}{\text{total}} = \frac{TP+TN}{N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6.2)$$

Given that the accuracy metric has limitations when used with unbalanced data (i.e., usually, the number of normal frames is greater than the number of malicious frames) [166]. For example, negative (normal) cases account for 9850 frames of data, whereas positive (suspicious) cases account for only 150 frames of the data. If the classifier always predicts the majority class (i.e., predicts all the data as TN), the accuracy will equal 98.5%. However, the classifier did not ever predict any of the TP (suspicious). So, accuracy may not be a good measurement when the data has unbalanced classes. To ensure that classifier performance is correct when using the proposed features, several support metrics are used.

### 6.3.2 Precision

Precision is the proportion of correct positive prediction (true positive) to the total classified positive cases. Precision is a way of describing how multiple measurements are close to each other. It ensures that the test shows the same results when repeated several times under stable conditions. High precision leads to fewer false positive cases [166]. In this research, the precision measures how often the classifier correctly predicts the suspicious feature-frames.

$$Precision = \frac{TP}{Predicted\ Positive} = \frac{TP}{TP+FP} \quad (6.3)$$

### 6.3.3 Recall

Recall is the proportion of correct positive classification (true positive) to the total actual positive cases. Recall measures the true positive rate. High recall leads to fewer false negatives cases [166].

$$Recal = \frac{TP}{Actual\ Positive} = \frac{TP}{TP+FN} \quad (6.4)$$

### 6.3.4 F-score

F-score or F-measure is a metric that returns a value between zero and one and utilized to assess the usefulness of the classification technique. The higher the F-score, the better the prediction of the classifier. F-score is the harmonic mean for the recall and precision [166, 168]. The value of the F-score is calculated as:

$$F = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (6.5)$$

### 6.3.5 Area Under the Curve AUC

Receiver Operating Characteristic (ROC) curve is commonly used with Area Under the Curve (AUC), to evaluate the binary classification that is the classification of two classes. ROC curve summarizes the performance of the binary classifier by plotting the correctly classified instances, the True-Positive rate against the False-Positive rate. The misclassification rate is a type of simple classification measure, which visualizes the error rate for the single threshold. ROC curve has several advantages over this as it deals with all possible classification thresholds when representing the error rate. The AUC, which is

just the percentage of area that is under the ROC curve is an effective metric with data sets. This is the case even when the classes are highly unbalanced [166, 169].

### 6.3.6 Kappa

The performance of the classifier, which is the ‘Observed Accuracy’ can be compared with the possibility of achieving this performance by random chance, also called ‘Expected Accuracy’. The chance-corrected metric that is utilized for this comparison and assessment is ‘Kappa’. The Kappa score is directly proportional to the difference between the accuracies. So, we can say that a model will have a high Kappa score if there is a big difference between the observed accuracy and the expected accuracy. As Kappa considers random change in its calculation, it is more reliable than using the accuracy [170, 171].

Kappa is calculated using the following equations:

$$kappa = \frac{Observed\ Accuracy - Expected\ Accuracy}{1 - Expected\ Accuracy} \quad (6.6)$$

$$Expected\ Accuracy = \frac{Exp\ Acc\ Normal + Exp\ Acc\ Suspicious}{n} \quad (6.7)$$

$$Exp\ Acc\ Normal = \frac{(TN+FP)*(TN+FN)}{n} \quad (6.8)$$

$$Exp\ Acc\ Suspicious = \frac{(TP+FP)*(TP+FN)}{n} \quad (6.9)$$

*Observed Accuracy:* Instances classified correctly.

*Expected Accuracy:* Expected accuracy by any random classifier.

*n:* Number of instances in the data.

### 6.3.7 Matthews Correlation Coefficient

B.W. Matthews had developed the Matthews Correlation Coefficient (MCC), a type of machine learning metrics used to assess the quality of the binary classifiers [172, 173].

Biomedical research widely makes use of MCC and it has been selected as one of the

elective metrics in the US FDA-led initiative, MAQC-II [174]. MCC is also suitable to assess the binary classifier when dealing with data that is imbalanced that means classes of different sizes of data. The coefficients are returned between -1 to 1 by using MCC. The higher the coefficient, the better the classifier is considered to be.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (6.10)$$

### 6.3.8 Percent Difference

To compare two experimental results when both results are obtained using different approaches, the Percent Error or Percent Difference (PD) is usually used. When the value is equal to zero or when the two values have a big difference, the maximum PD will never be higher than 200% [175]. Percent Error is calculated using Equation 6.11.

$$Percent\ Difference(PD) = \left| \frac{1st\ value - 2nd\ value}{(1st\ value + 2nd\ value)/2} \right| \quad (6.11)$$

### 6.3.9 Cross-Validation

The k folds cross-validation is an evaluation technique for the machine learning predictive model. The cross-validation separates the data into k equal-size subsets. One of these subsets is used to test the machine learning model, whereas the k-1 is used to train the model. The k folds cross-validation repeats the learning algorithm k times, and then the average performance of the k different testing subsets is calculated [166]. In this research 10 folds cross-validation is used.

### 6.3.10 Confidence Interval

To ensure that the achieved results for the proposed approach are correct, and to describe the uncertainty in the estimates, the 95% confidence interval of the accuracy that achieved

by the classifiers has calculated. The confidence interval of the results was calculated as following:

$$\text{confidence interval} = \bar{x} \pm E \quad (6.12)$$

$$E = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (6.13)$$

$$\sigma = \sqrt{S^2} \quad (6.14)$$

$$S^2 = \frac{1}{n-1} [\sum_{i=1}^k f_i x_i^2 - n\bar{x}^2] \quad (6.15)$$

Where  $\bar{x}$  is the mean,  $E$  is the error,  $n$  is the number of participants,  $z_{\frac{\alpha}{2}}$  is a constant,  $\sigma$  is a standard deviation,  $S^2$  is the variance, while  $f$  is the frequency of participants in each range and  $x$  is the midpoint of each range.

## 6.4 Hypotheses Testing and Validity

We validate the null hypotheses by evaluating the proposed system and the extracted features using the following procedures. First, the accuracy of EEG features is presented to assess hypothesis H0-1. Then, the accuracy of the proposed EEG+ECG method is discussed and compared with the EEG method to test hypothesis H0-2 and illustrate the impact of adding ECG features on results. Validation of the research hypotheses using the accuracy metric does not provide high credibility, given that the accuracy metric has limitations when used with unbalanced data (i.e., usually, the number of normal frames is greater than the number of malicious frames). Therefore, additional support metrics mentioned in section 6.3 have been used to assess the correctness of the accuracy metric. Figure 27 shows the procedures for testing the validity of the research hypotheses.



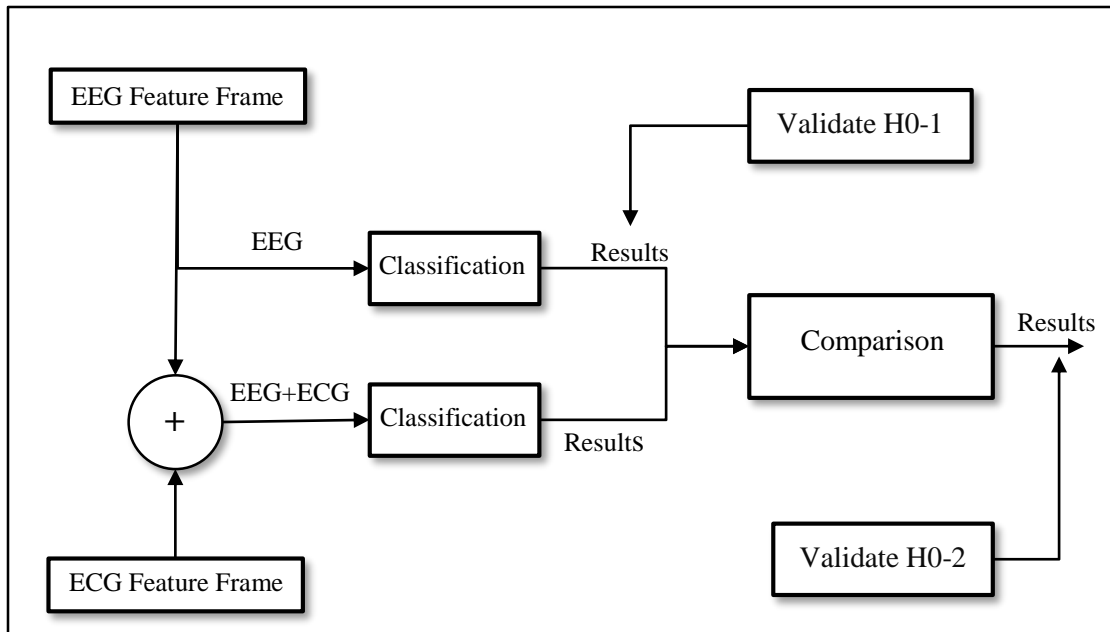


Figure 27: Testing The Validity of The Null Hypotheses

## 6.5 Presenting Results

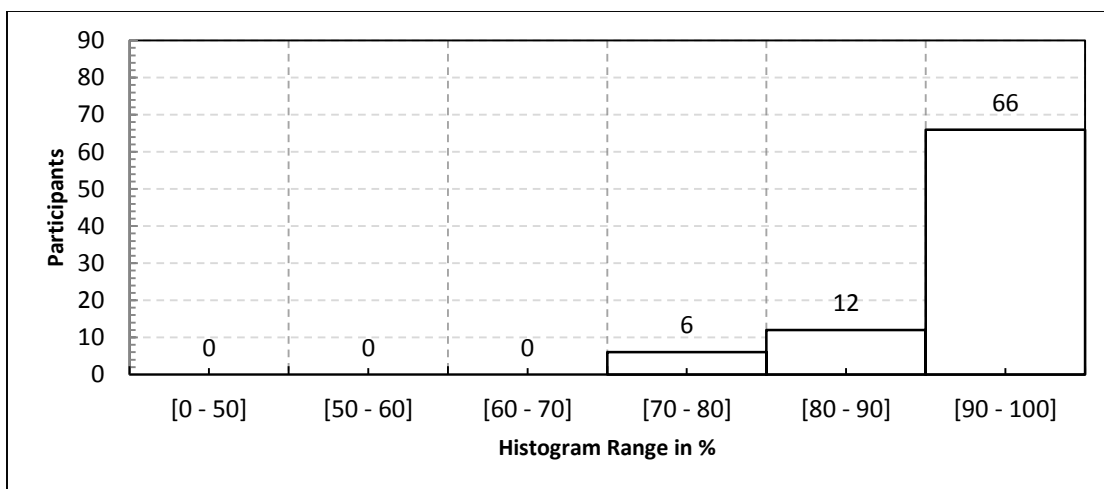
Bio-data were collected for 84 participants during both scenarios (normal and suspicious). In order to illustrate the results for this large number of participants, the results will be presented in the form of a histogram. Histograms provide a visual display of vast amounts of data that are hard to understand in a tabular format. The values of each metric (ex. accuracy) are divided into ranges, which are used as the histogram ranges. Each range is determined by the highest and lowest values of the range. The histogram separates the results into groups, each of which contains the number of participants that fall within the same range of the histogram and thus, the same range boundaries of the metric.

Figure 28 shows an example of the results presentation. From the figure, the X and Y axis of the graph shows the metric's ranges and the number of participants, respectively. Clearly, the figure illustrates the results of 66 participants fall within the range of 90–100% of the histogram, meaning that these participants achieved results higher than 90% and less than

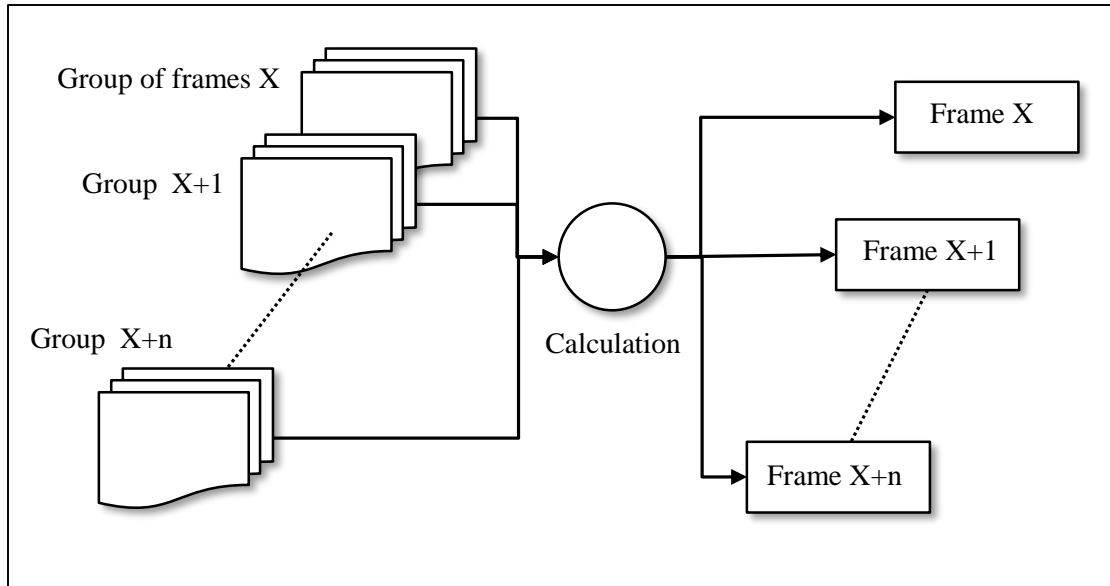
100% when using the specified metric. Moreover, in our experiments, no volunteers fall in the range from zero to 50%. Therefore, ranges with low values are ignored to abbreviate the empty results.

In addition to the histogram results, the results of each participant will be presented in scatter graph. Unlike the histogram that represents a range of accuracy values, each point in the scatter graph represents only one value of accuracy and the number of participants achieved that value of accuracy.

Moreover, evaluating each frame of incoming data individually to distinguish insider threats would keep the system busy all the time and waste the resources of the device containing the proposed system. Therefore, the average of each five feature-frames is calculated and packaged as a single frame. So, instead of evaluating each frame, a group of frames will be evaluated by converting them into a single frame as shown in Figure 29. Moreover, the average of ten and the average of fifteen feature-frames will be tested to illustrate if there is a degradation in the quality of the detection when the multiple frames compressed together.



**Figure 28: Results Presentation**

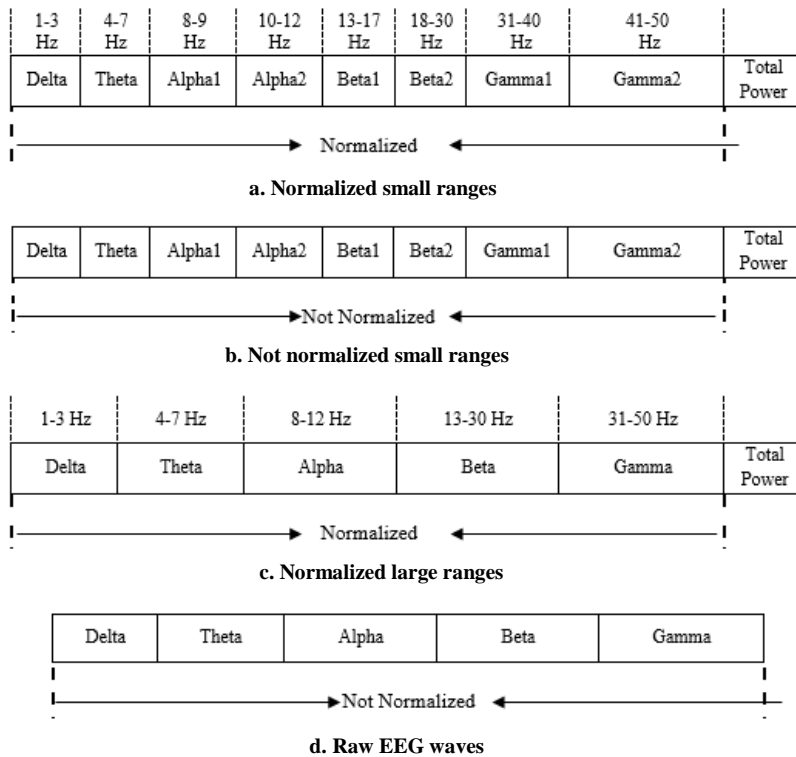


**Figure 29: Generating a Group of Frames**

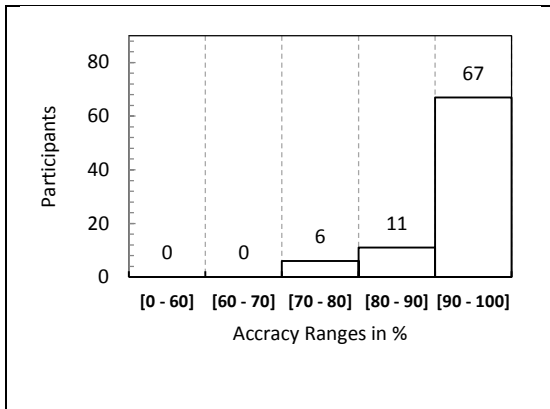
## 6.6 Results of EEG Features

This section justifies accepting or rejecting research hypothesis H0-1. To this end, the results of using only the EEG features are presented and discussed. Moreover, the proposed EEG approach is compared with different feature-frames illustrated in Figure 30 to demonstrate the effect of each factor of the EEG feature-frame (i.e., using small ranges, normalization, and additional features). The results are compared with the raw EEG waves (i.e., Figure 30, frame d) in detail, to assess the effectiveness of the extracted features. The accuracy metric is used as a reference to evaluate the validity of H0-1 because accuracy describes how the results of the proposed approach are close to the true results. However, due to the accuracy limitations when used with unbalanced data as in this research, the other metrics mentioned earlier in this chapter were used to support and measure the correctness of the reference metric.

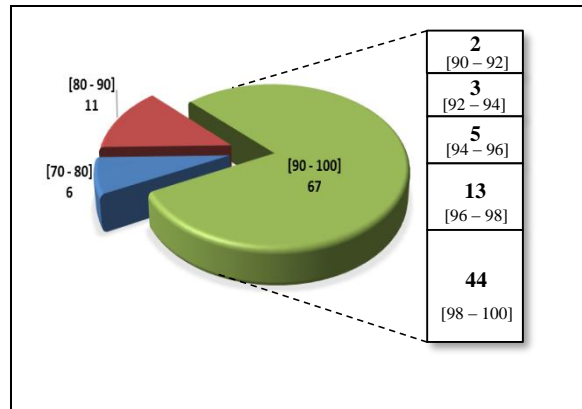
Figure 31 illustrates the classification accuracy of the proposed EEG approach using the Random Forest (RF) classifier. The figure shows that most of the participants are correctly classified, with an accuracy of more than 90%. In other words, the classifier can differentiate between normal and malicious activities for those cases with an error less than 10%. On the other hand, accuracy is lower than this for 17 participants. To investigate these results further, Figure 32 illustrates more details about the accuracy of the proposed approach. The figure shows the distribution of classification accuracy of the range 90–100% in more detail. The classifier can detect malicious activities in 44 cases with an accuracy of more than 98%, which indicates the efficiency of the proposed approach for identifying such threats.



**Figure 30: Features for Comparison**



**Figure 31: Classification Accuracy using RF**



**Figure 32: Accuracy of Proposed Approach in Detail**

The timeframe of suspicious activity is always very small compared to the time spent by the insider attacker in normal work. Like any thief, an insider attacker will be keen not to be detected or tracked. He will try to hide the traces of his malicious acts by executing the attack in several stages at different times. So, the amount of suspicious incoming data will vary between attackers and between suspicious acts. Therefore, the number of the detected suspicious signals among the total incoming signals varies from one person to another, resulting in different detection accuracy.

The attribute selection method was used to illustrate the effect of the alpha difference AD factor on the results. During the classification of data, the classifier may rely on some features (trusted or selected features) more than others. These trusted features have a significant impact on the results. Thus, the classifier can obtain almost the same classification results using only the trusted features. The procedure of selecting the trusted features is called the attribute selection method. In this research, the supervised attribute selection is implemented using Weka API with Java code [128, 129]. The supervised attribute selection requires a search algorithm and an evaluation method. The GreedyStepwise [176] and the WrapperSubsetEval [177] were used as the search and

evaluation respectively. The WrapperSubsetEval generates all possible subsets of features and uses an induction algorithm to select the subset of features that achieved the highest evaluation [178]. Figure 33 shows the impact of each feature on the results of the proposed approach. The figure shows that AD was one of the most influential features in more than half of the classification cases, because the concentration level of the participant will change when conducting the unauthorized actions as well as the AD level.

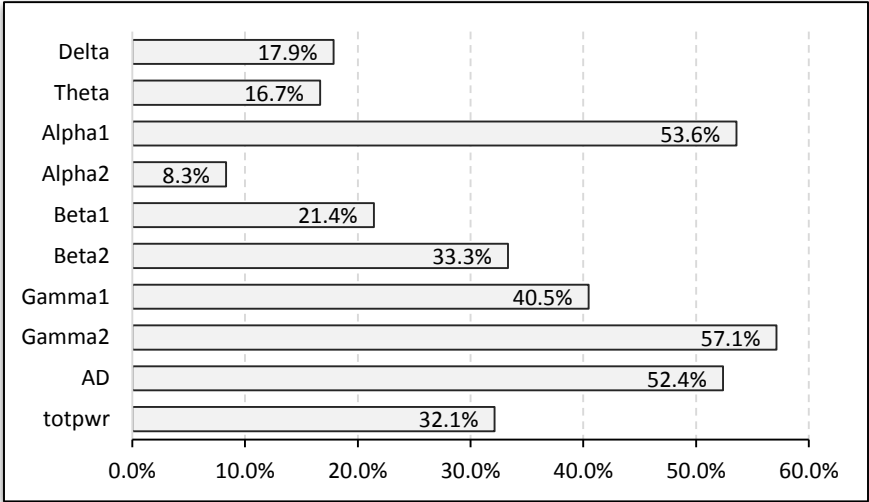


Figure 33: Impact of Features on the Results

Furthermore, to demonstrate that the addition of the AD factor increased the accuracy of detecting insider threats, Figure 34 illustrates the accuracy comparison of the proposed approach with and without using the AD factor. The figure shows that there is a significant increase in cases detected with accuracy greater than 90% when adding the AD factor to the proposed approach. The number of malicious cases discovered by the classifier increased by around 55.8% with accuracy above 90% when using the AD factor. Moreover, Figure 35 illustrates the effect of AD factor in for each participant using scatter chart. Each point in the scatter graph represents the number of participants achieved that accuracy.

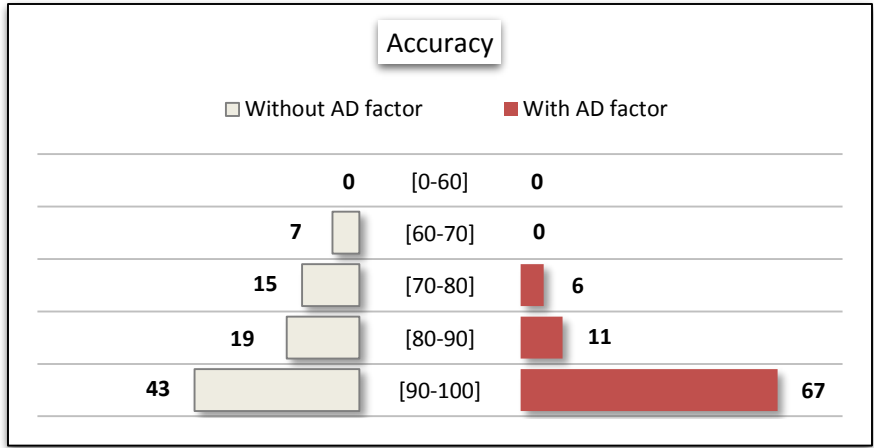


Figure 34: Effect of AD Factor

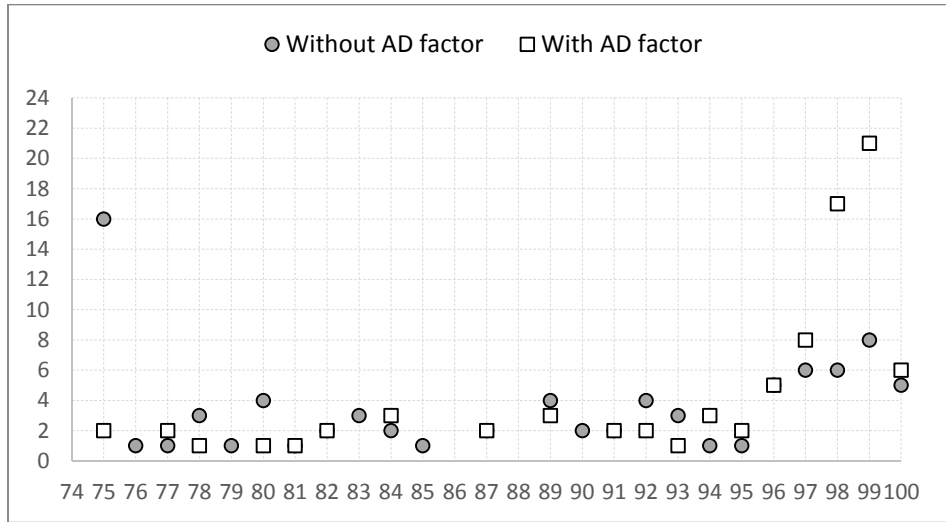
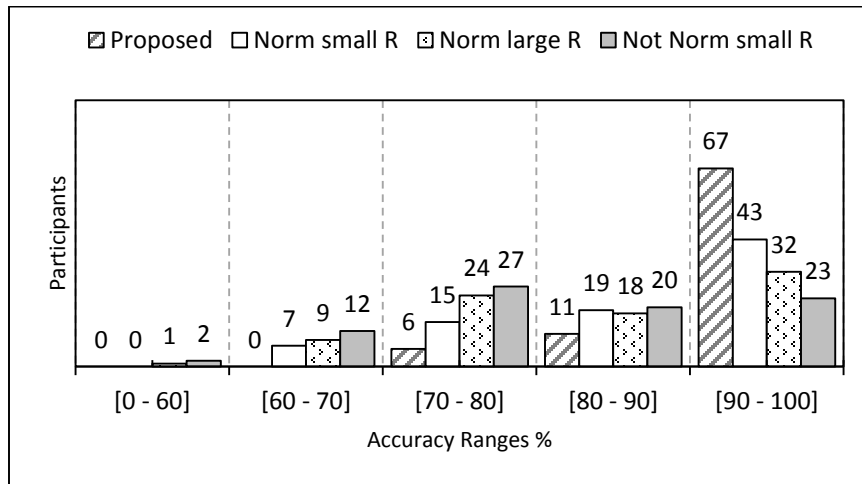


Figure 35: Effect of AD Factor in Details

In order to clarify the effect of using small frequency bands and feature normalization, we compare the proposed EEG feature-frame with the feature-frames in Figure 30 (i.e., a, b, and c). Figure 36 illustrates that the normalized frame with small frequency ranges achieved better results than the normalized frame with large ranges. Using small ranges, the classifier detected 43 cases with accuracy above 90%, compared to 32 cases when using large ranges. The small ranges of frequency provide more features for comparison, thus improving the detection of harmful activates. Moreover, Figure 37 shows the results in

details for each participant. Each point in Figure 37 represents the number of participants achieved the corresponding accuracy.

Furthermore, Figure 36 shows a comparison of the normalized and un-normalized frame with small frequency ranges. Using the un-normalized feature-frame, the classifier detected only 23 cases with accuracy above 90%, which is the worst result compared to the rest of the feature-frames. The normalization reduced the impact of EEG signal variability and increased the ability of the proposed approach to distinguish between normal and malicious activities [153]. This result demonstrates the effect of normalizing the extracted features. In contrast, the proposed EEG approach significantly outperformed in this comparison.



**Figure 36: Accuracy Comparison of Feature-frames**



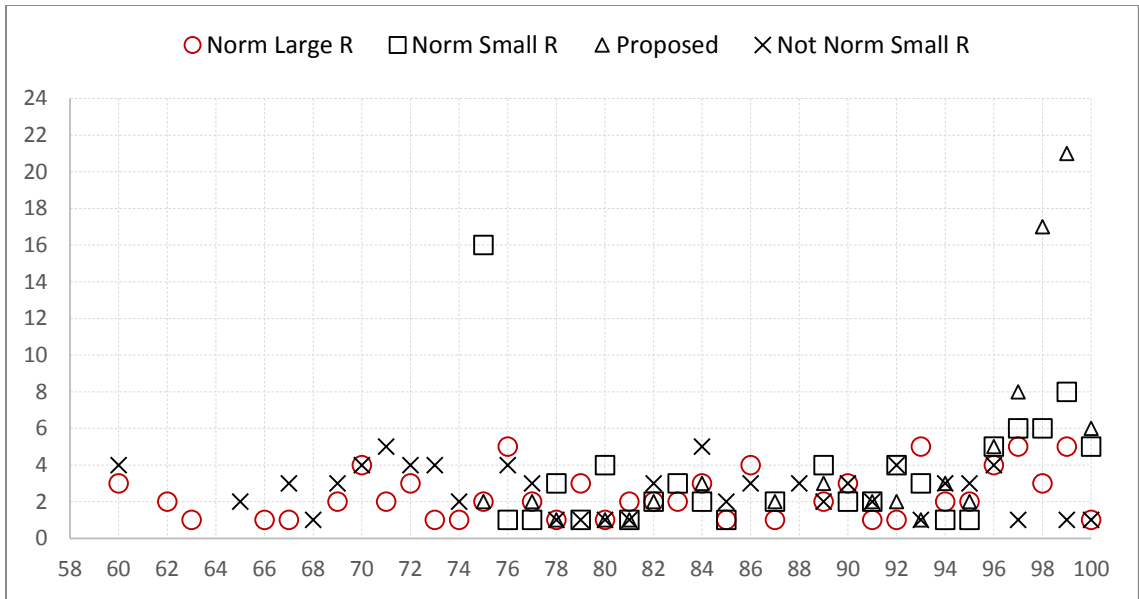


Figure 37: Accuracy Comparison of Feature-frames using Scatter Chart

To illustrate the changes in the smaller EEG frequency bands during normal and malicious acts, Figure 38 shows the lower and higher values of every frequency range. Each value represents the median of all participants for the range. For example, the lower value of delta is the median of the lower delta values of all participants. The changes are the result of changing the concentration of each EEG wave during different scenarios. Also, using smaller ranges highlights the differences more clearly.

Furthermore, to demonstrate that the extracted EEG features have much better performance than the raw EEG waves (i.e., frame d in Figure 30), Figure 39 illustrates the comparison of the two methods' accuracy. This comparison shows the number of participants for whom a classifier could correctly detect that they engaged in malicious activities and demonstrates the accuracy of this detection. The figure illustrates that when using the proposed EEG features, the classifier can correctly detect insider threats more than two times better than when using the raw brainwaves: Using the proposed EEG features, the classifier detected 67 cases with accuracy above 90%, compared to 24 cases when using

raw EEG waves. Moreover, unlike the raw EEG waves, the proposed EEG features achieved fewer cases within the smaller accuracy ranges. Furthermore, Figure 40 provides more details about the comparison between the proposed EEG features and the raw brainwaves by illustrating the accuracy of individual participant in a scatter chart.

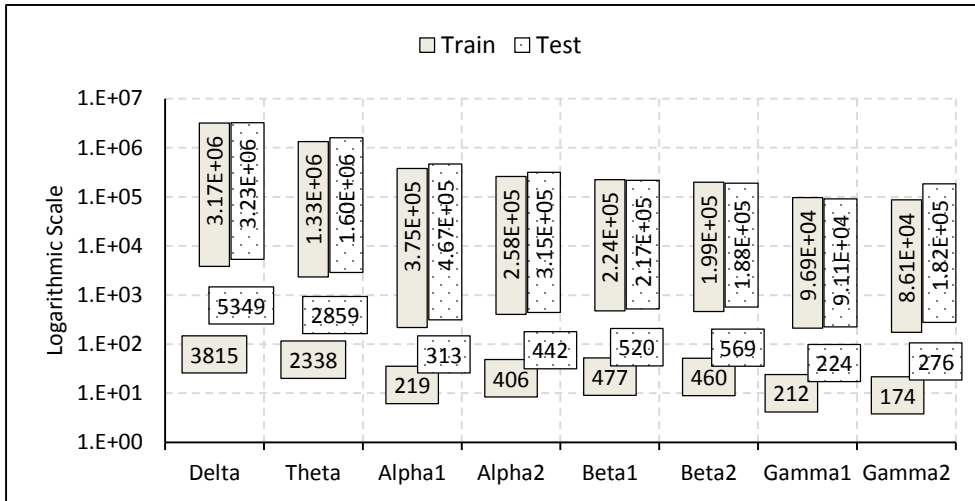


Figure 38: EEG Frequency Bands During Normal and Malicious Acts

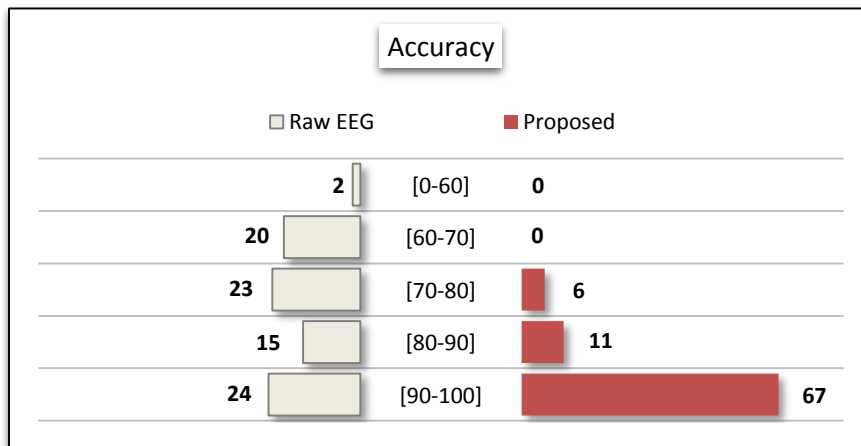


Figure 39: Accuracy of The Proposed EEG And Raw EEG Data Using RF

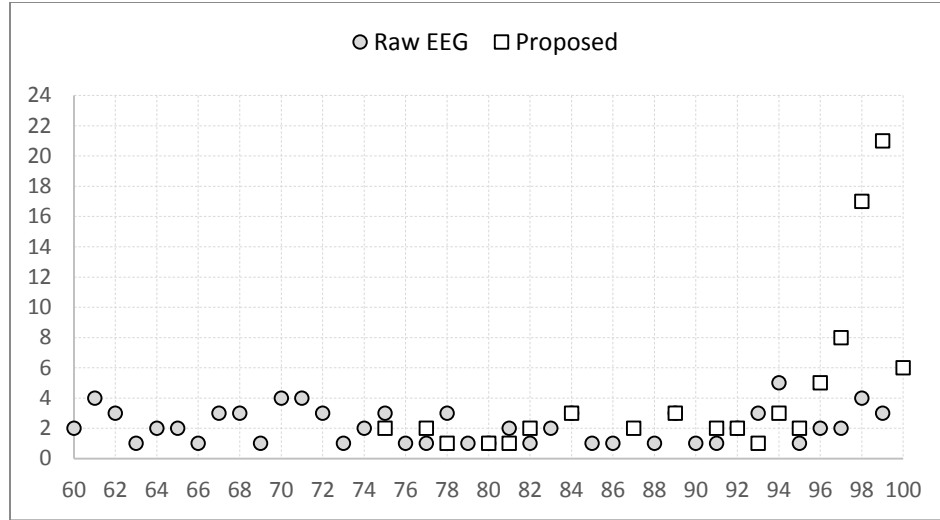


Figure 40: Scatter Chart of Accuracy for The Proposed EEG and Raw EEG

### 6.6.1 Rejection of Null Hypothesis H0-1

To reject the first null hypothesis H0-1, the proposed EEG features must increase the detection values of malicious activities more than using the raw brainwaves. To demonstrate this, the Z-score test (equation 6.1) is used as following:

$\bar{d}$ : The average of difference = 15.11.  $S$ : Standard Deviation = 12.28

$n$ : Sample Number = 84.  $\alpha$ : Significance Value =1%. The t state  $t_0 = 11.28$

The t Critical one-tail (cutoff point) is 2.372. In right-tail hypothesis testing, any  $t_0$  greater than the critical value will be used to reject H0-1. Since t state =11.28 which is greater than 2.372, we reject the first null hypothesis H0-1 and developed an alternative hypothesis, which we will refer to as H1-1.

**H1-1)** *The extracted features from the brainwaves (EEG signals) have a positive effect on detecting an insider attacker, by differentiating between normal and suspicious activities.*

## 6.7 Results of EEG+ECG Features

This section aims to validate the second research hypothesis H0-2 and illustrate the potential improvement of using the ECG features for identifying insider threats. To achieve this end, the classification results of the extracted EEG+ECG features will be presented, then the EEG+ECG accuracy is compared with only the accuracy of the EEG features. The improvement in the EEG+ECG accuracy over the EEG-only features illustrates the effect of using ECG on detecting insider threats.

Figure 41 shows that when using EEG+ECG features, the classifier can distinguish between normal and malicious activities in 81 cases with an accuracy of more than 90%, compared with 67 cases when using only the EEG features. The number of cases in which the classifier was confident that the participant was doing malicious work increased when using EEG+ECG. This improvement demonstrates the effect of adding the extracted ECG features to the proposed EEG frame. Moreover, Figure 42 provides more details using the scatter chart.

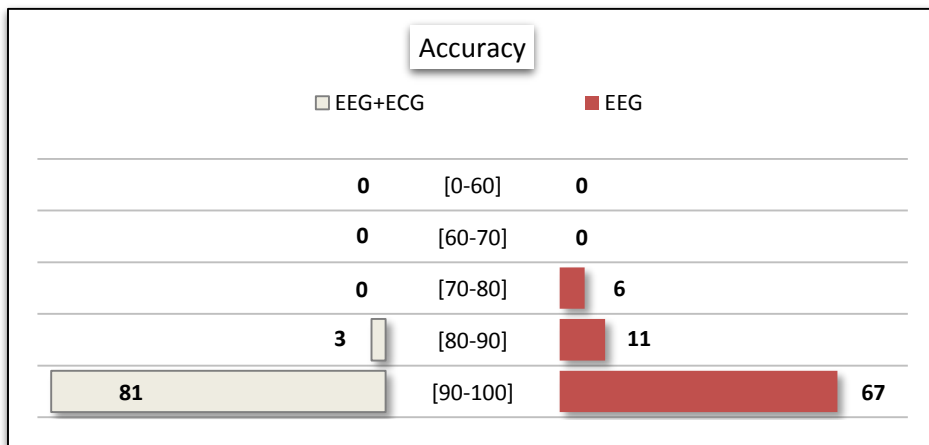
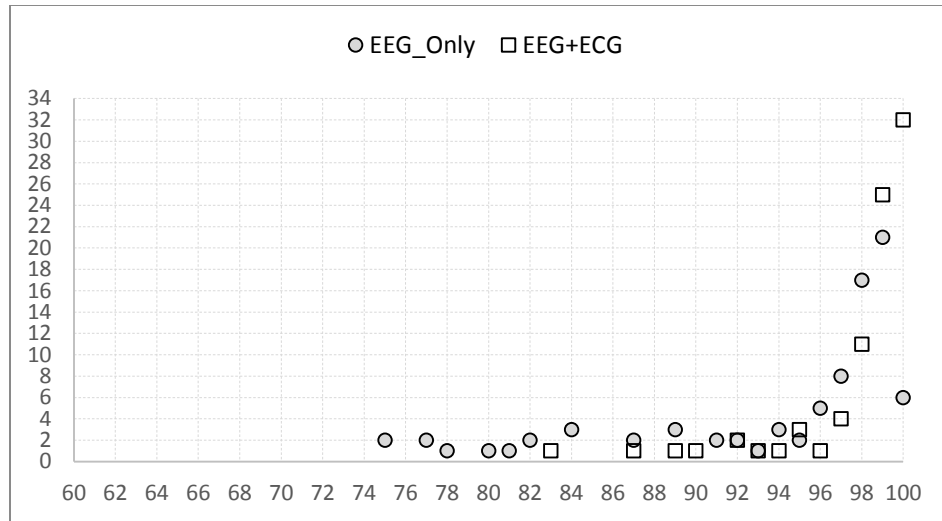


Figure 41: Accuracy of The Proposed EEG+ECG and EEG Features Using RF



**Figure 42: Scatter Chart of Participant' Accuracy using EEG+ECG and EEG features**

Percent difference (PD) is usually used when comparing two experimental results when the results have been obtained using two different approaches. Table 11 illustrates the PD of EEG and EEG+ECG accuracy. With accuracy of more than 90%, cases detected using the EEG+ECG method increased by around 18.9% over EEG only. When the classifier is confident with more than 90%, the classifier can detect 81 participants using EEG+ECG compared with 67 participants when using only EEG features. Using EEG+ECG signals increases the variation between normal and malicious signals. Thus, it can achieve better results when detecting insider threats than EEG can.

**Table 11: PD of Accuracy Ranges**

Accuracy Ranges in %	Participants		(PD)	
	EEG Features	EEG+ECG Features		
[70 - 80]	6	0	200%	140%
[80 - 90]	11	3	114.3%	
[90 - 100]	67	81	18.9%	

### 6.7.1 Rejection of Null Hypothesis H0-2

The results of the proposed EEG+ECG approach illustrate that the ECG features had a positive impact on detecting insider threats, by increasing the variation between normal and suspicious activities. Adding ECG features to the EEG feature-frame leads to accuracy improvement of the proposed system. To test H0-2, the Z-score test (equation 6.1) is used.

$\bar{d}$ : The average of difference = 3.623  $S$ : Standard Deviation = 5.85

$n$ : Sample Number = 84.  $\alpha$ : Significance Value =1%. The t state  $t_0 = 5.67$

The critical value (cutoff point) is 2.372. The t state of 5.67 is in the rejection area. We reject the second null hypothesis H0-2 and develop an alternative research hypothesis which we will refer to as H1-2.

**H1-2)** *The extracted features from the electrocardiogram (ECG) signals have a positive effect on detecting an insider attacker, by differentiating between normal and suspicious activities.*

## 6.8 Classification Accuracy Assessments

Given that the accuracy metric has limitations when used with unbalanced data (i.e., usually, the number of normal frames is greater than the number of malicious frames) [152], we justified the results using six additional metrics illustrated in Figure 43. These metrics are effective with data sets containing highly unbalanced classes. For instance, the area under the curve, and Matthews correlation coefficients that are widely used in the biomedical research and selected in the US FDA-led initiative MAQC-II as one of the elective metrics [174]. Figure 43 shows the evaluation of the proposed EEG and

EEG+ECG approaches when using the RF classifier. It can be noticed that both approaches achieved high results within the range of 90–100%, which proves the accuracy and quality of the extracted features to detect insider attacks.

Precision represents the ratio of the true number of malicious frames detected by the classifier to the total frames that were predicted as malicious. Figure 43.a shows that when using the proposed EEG+ECG features, the RF classifier can correctly detect 80 cases with precision range above 90%. This result indicates that the proposed system is accurate for detecting malicious threats, and the change in the detection accuracy will not exceed 10% when the test is repeated several times. On the other hand, within the same precision range, the classifier can correctly detect 66 cases using proposed EEG features. This high level of precision value illustrates that the classification accuracy of the proposed features is stable and is not achieved randomly.

Recall demonstrates the ratio of the number of malicious frames detected to the total number of true malicious frames for a participant. Figure 43.b illustrates that the proposed EEG+ECG detected 79 cases with recall range above 90%. This result demonstrates the number of true positives (malicious frames) that the RF classifier can detect from the participant data. In other words, the classifier can predict correctly 90–100% of the total suspicious signals for 79 participants. Conversely, using EEG features, 63 cases were detected with recall range above 90%.

Matthews's correlation coefficient (MCC) returns coefficients in the range of -1 to 1. Regardless of the size of unbalanced data, the higher the coefficients are, the better the classifier's prediction is. From Figure 43.c, both EEG and EEG+ECG methods achieved

positive MCC values much higher than -1. Using EEG+ECG, the classifier detected 75 cases with MCC above 90%.

The area under the curve (AUC) demonstrates how good the classifier is to distinguish between normal and malicious activities. The high value of AUC proves that the FP rate (i.e., detected malicious data) is much more than the FN rate (i.e., normal data identified as malicious). From Figure 43.d, when using the EEG+ECG approach, the classifier can detect all the cases within the AUC range above 90%. This result proves the quality of the extracted features which reduce the number of normal frames that are detected as suspicious. In contrast, using the proposed EEG approach, RF classifier detects 74 cases with AUC above 90%, which is considered a high result.

F-score is a single measure which represents the harmonic mean for recall and precision. F-score is utilized to assess the usefulness of the classification technique. The higher the F-score, the better the prediction of the classifier. Figure 43.e shows that when using the EEG+ECG approach, the classifier detected 80 cases with F-score above 90%. This result shows the predictive power of the classifier due to the advantages of the proposed features. Moreover, within the F-score range above 90%, the classifier detected 67 cases using the proposed EEG method.

Kappa evaluates the performance of the classifier compared with the possibility of achieving this performance through random chance. The higher the Kappa value, the more accurate the classifier is. Figure 43.f illustrates that when using both the EEG+ECG and EEG methods, the classifier detects most of the cases with kappa range above 90%.



To summarize this section, the proposed approaches (EEG+ECG and EEG) achieved high results within the range of 90–100%, which proves the quality of the extracted features to detect the insider threats and supports the accuracy metric.

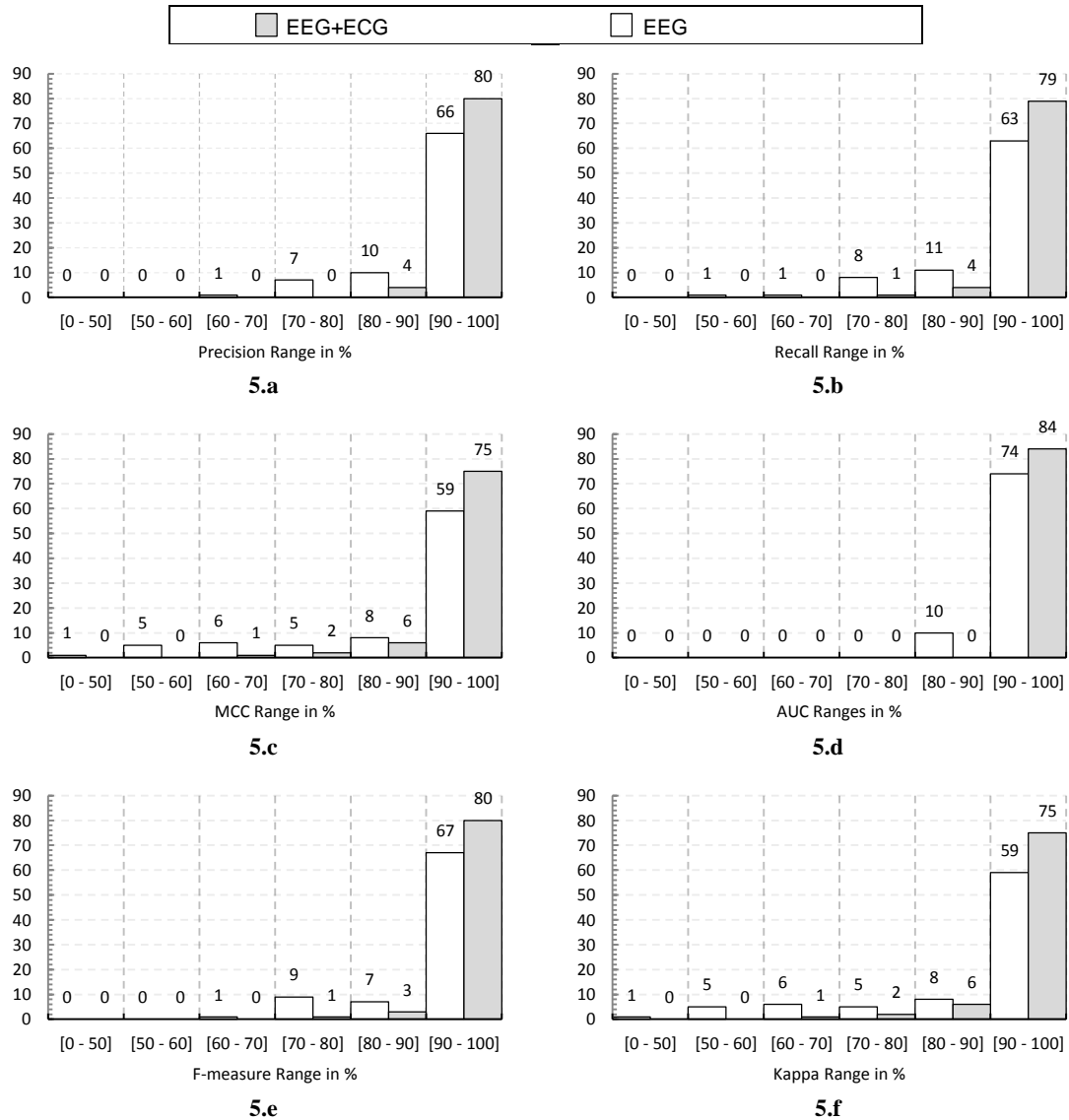


Figure 43: (a,b,c,d,e,f): Evaluating the Results using Several Metrics

## 6.9 Evaluation Using Three Classifiers

Since classification is a crucial phase for distinguishing between insider threats and normal activities, three classification algorithms have been used to assess their influences on the

results of the proposed approach: the random forest (RF) [158], support vector machine (SVM) [160], and back propagation neural network (NN) [162]. Each algorithm has its model and learning method, discussed earlier in Chapter 5. This section illustrates the effect of these classifiers on the results of the proposed EEG and EEG+ECG methods.

Table 12 compares the proposed approach with the raw brainwaves. The table summarize the results of the eighty four participants that are shown in Figure 40 and Figure 42 by illustrating the average accuracy, the confidence interval (C.I), false positive rate (FPR) and false negative rate (FNR). The table illustrates the comparison using three classifiers which are RF, SVM and NN. The table illustrates the improvement of the proposed approach for detecting internal threats.

**Table 12: Comparing Approaches using Confidence Interval, FPR and FNR**

Classifier	Random Forest			Support Vector Machine			Neural Network		
	Accuracy $\pm$ 95% C. I.	FPR	FNR	Accuracy $\pm$ 95% C. I.	FPR	FNR	Accuracy $\pm$ 95% C. I.	FPR	FNR
Raw BW	79.69 $\pm$ 2.74	13.5%	31.4%	75.51 $\pm$ 2.3	10.3%	48.6%	77.9 $\pm$ 2.72	13.5%	36.1%
EEG	94.8 $\pm$ 1.49	3.9%	7%	95 $\pm$ 1.6	3.4%	7.1%	95.33 $\pm$ 1.71	3.8%	4.2%
EEG+ECG	98.42 $\pm$ 0.62	1.1%	2.3%	98.26 $\pm$ 0.77	1.3%	2.4%	98.3 $\pm$ 0.82	1.4%	2.1%

Table 13 illustrates the performance of the proposed approaches using different classifiers and several evaluation metrics. Although the three classifiers achieved high average accuracy in identifying the insider threats when using the proposed EEG+ECG features, the random forest (RF) slightly outperformed, with an average accuracy up to 98.42%. The very close results of the three classifiers when using EEG+ECG indicate the quality of the extracted features that allow the three classifiers to identify the malicious activities with high accuracy. On the other hand, when using the EEG features, the back propagation neural network (NN) achieved an average accuracy of 95.33%. Moreover, the table

clarifies the improvement in the results when using the extracted EEG features over those when using the raw brainwaves.

**Table 13: Comparison of EEG+ECG with EEG Approach and the Raw Brainwaves**

Classifier	Random Forest			Support Vector Machine			Neural Network		
	EEG+ECG	EEG	Raw EEG	EEG+ECG	EEG	Raw EEG	EEG+ECG	EEG	Raw EEG
Accuracy	98.42	94.80	79.69	98.26	95.01	75.51	98.30	95.33	77.90
Precision	0.98	0.94	0.76	0.98	0.95	0.76	0.98	0.95	0.74
Recall	0.98	0.93	0.69	0.98	0.93	0.51	0.98	0.94	0.64
AUC	1.00	0.97	0.84	0.98	0.95	0.71	0.99	0.97	0.81
F-measure	0.98	0.94	0.72	0.98	0.94	0.57	0.98	0.94	0.67
Kappa	0.97	0.89	0.56	0.96	0.90	0.43	0.97	0.90	0.51
MCC	0.97	0.89	0.56	0.96	0.90	0.46	0.97	0.90	0.52

Furthermore, Table 14 shows that, for the three classifiers, the results of detecting insider threats from females are better than the results for males when using the extracted EEG features. This can be attributed to the fact that females report more levels of fear and anxiety than males due to different factors, including biological and cultural influences [179, 180]. On the other hand, the accuracy gap between males and females diminished when using the EEG+ECG features; this can be attributed to that the addition of the ECG features increases the distinction between the normal and suspicious activities in both genders, thus diminishing the gap between the results.

**Table 14: Comparing EEG+ECG with the Extracted EEG Based on Gender**

Measure	Classifier											
	Random Forest				Support Vector Machine				Neural Network			
	EEG+ECG		EEG		EEG+ECG		EEG		EEG+ECG		EEG	
	M	F	M	F	M	F	M	F	M	F	M	F
Accuracy	98.17	98.68	92.68	97.01	98.05	98.47	92.67	97.46	98.21	98.39	92.90	97.87
Precision	0.98	0.98	0.92	0.97	0.98	0.98	0.93	0.97	0.98	0.98	0.92	0.97
Recall	0.97	0.98	0.90	0.96	0.97	0.98	0.89	0.96	0.98	0.98	0.91	0.97
AUC	0.99	0.99	0.96	0.99	0.98	0.98	0.92	0.97	0.99	0.99	0.96	0.99
F-measure	0.98	0.98	0.91	0.96	0.98	0.98	0.91	0.97	0.98	0.98	0.92	0.97
Kappa	0.96	0.97	0.85	0.94	0.96	0.97	0.85	0.95	0.96	0.97	0.85	0.96
MCC	0.96	0.97	0.85	0.94	0.96	0.97	0.85	0.95	0.96	0.97	0.85	0.96

Figure 44 illustrates the classification accuracy of EEG+ECG features using three classifiers, i.e., RF, SVM, and NN. The RF classifier detects 81 cases with accuracy above 90%, which shows some improvement over SVM and NN. The SVM and NN classifiers correctly classified 80 and 79 cases with accuracy range above 90%, respectively. To simplify the comparison and discuss the classification results further, the percent difference PD is calculated as illustrated in Table 15.

Table 15 shows a comparison between the classifiers using the PD and a threshold level of 85%. The threshold level was selected to illustrate the accuracy of the three classifiers above this level. We can notice that when accuracy is above 85%, the RF classifier outperforms the remaining classifiers with around 1.2%, whereas the NN and SVM achieved the same number of detected cases. On the other hand, when accuracy is below the threshold level, the RF achieved accuracy 66.7% lower than the NN and SVM classifiers, where the RF detected one case compared with two cases identified by the other classifiers. Despite the small value of PD, the results of the three classifiers are almost the same.

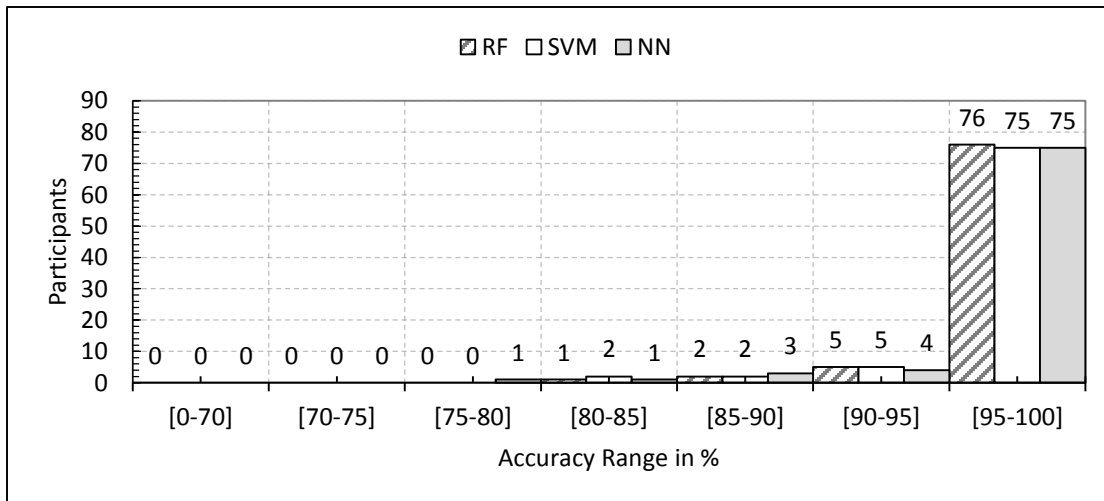
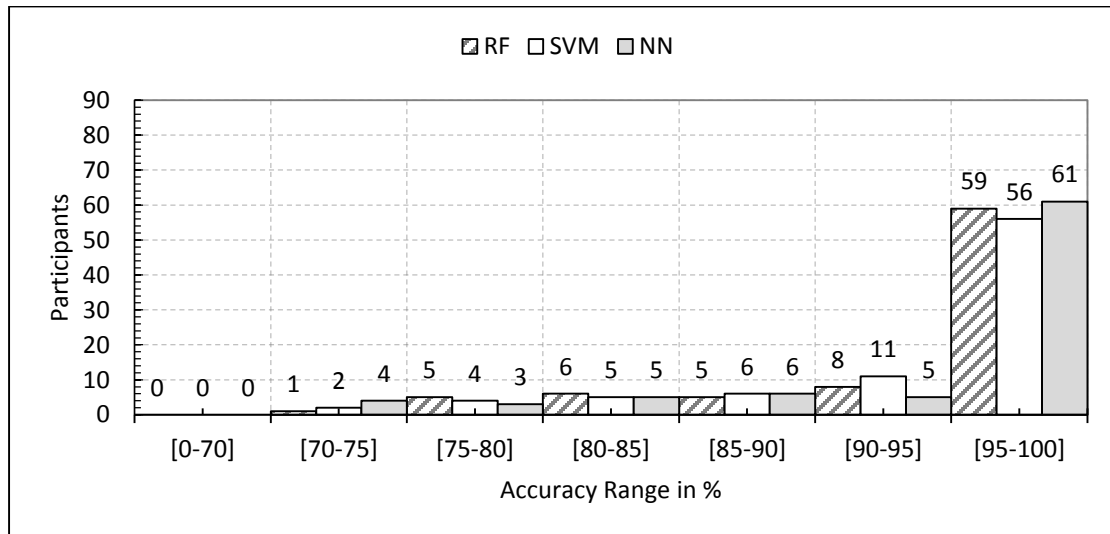


Figure 44: Accuracy of proposed EEG+ECG method using three classifiers

**Table 15: PD of The Classifiers' Accuracy Using EEG Features**

Accuracy Ranges in %	Volunteers			Percent Difference (PD)					
	RF	SVM	NN	RF&SVM		RF&NN		SVM&NN	
[75 - 80]	0	0	1	---	66.7%	200%	66.7%	200%	0%
[80 - 85]	1	2	1	66.7%		0%		66.7%	
[85 - 90]	2	2	3	0%	1.2%	40%	1.2%	40%	0%
[90 - 95]	5	5	4	0%		22.2%		22.2%	
[95 - 100]	76	75	75	1.3%		1.3%		0%	

Figure 45 demonstrates the influences of the classification algorithms on the results of EEG features. When accuracy is above 90%, the RF and SVM classifiers detect 67 cases compared to 66 cases detected by the NN classifier. Moreover, the three classifiers detected eleven cases within the accuracy range of 80–90%. The figure demonstrates that the classifiers achieved approximately the same accuracy, which proves the quality of proposed EEG features is not affected by the learning algorithms of the classifiers.



**Figure 45: Accuracy of proposed EEG method using three classifiers**

Furthermore, Table 16 illustrates the impact of classification algorithms on the EEG features. The table shows that when the 85% threshold level is used, RF and NN detected

the same number of cases. Thus, they do not have any PD. In contrast, SVM achieved 1.4% higher results within the accuracy range above 85%.

**Table 16: PD of Classifiers's Accuracy Using EEG+ECG**

Accuracy Ranges in %	Volunteers			Percent Difference (PD)					
	RF	SVM	NN	RF&SVM		RF&NN		SVM&NN	
[70 - 75]	1	2	4	66.7%		120%		33.3%	
[75 - 80]	5	4	3	22.2%	8.7%	50%	0%	13.3%	8.7%
[80 - 85]	6	5	5	18.2%		18.2%		0%	
[85 - 90]	5	6	6	18.2%		18.2%		0%	
[90 - 95]	8	11	5	31.6%	1.4%	46.2%	0%	75%	1.4%
[95 - 100]	59	56	61	5.2%		3.3%		8.6%	

### 6.10 Evaluation Using a Group of Frames

Using an average of five frames certainly reduces the data processing time compared to analyzing each frame of incoming data. However, the accuracy of the proposed EEG+ECG could be affected by compressing the data. To illustrate the effect of using a group of five frames on detecting malicious attacks, a comparison between the accuracy of the whole set of incoming data (1F) and the accuracy of a group of five frames was conducted, as shown in Figure 46. The comparison includes the average (AVG), median, and standard deviation (STD) of a group of five frames. From the figure, we can notice that, when accuracy is above 95%, the RF classifier classifies the data of only 18 volunteers using the standard deviation, which is very far from the classification results when using one frame (EEG+ECG) or the average and the median of five frames. However, using 1F, the classifier correctly classifies the data of 74 volunteers, with an accuracy above 95%. In the same range, the classification results of the average and the median of five frames are 75 and 76 volunteers, respectively.

The average and the median of a group of five frames do not significantly affect the accuracy of detecting malicious activities, which indicates the quality of extracted EEG+ECG features. The proposed EEG+ECG approach succeeded to differentiate between normal and malicious activities, and it divided the related signals of each activity into a separate group of similar signals. So, the average and the median of the similar five frames provide approximate value to the values in the five frames. On the contrary, the standard deviation of each five frames increased the dispersion of values and decreased the accuracy of detecting malicious activities.

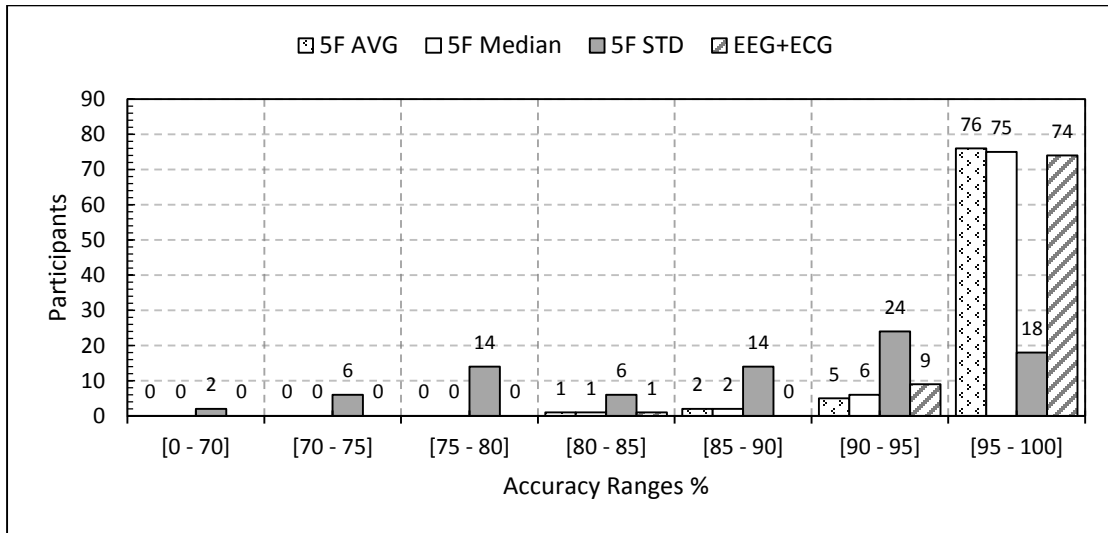


Figure 46: Comparing EEG+ECG Accuracy of AVG, Median, and STD

### 6.11 Evaluation Using Varied Amount of Malicious Data

It is very important that the proposed approach recognizes the malicious data, even if the amount of the incoming data is too small. Figure 47 shows the comparison of the learning curves of the three classification models using the error rates while varying the amount of incoming data. The figure illustrates the error rate for the classifiers while feeding the classifiers with 10% less data at a time. As illustrated from the figure, the NN classifier

with the proposed EEG approach can correctly detect the malicious activities with an error rate around 8% when using only 10% of the incoming data. On the other hand, the NN classifier has an error rate around 24.7% with the raw EEG data. Moreover, Figure 47 shows that when the incoming data are too small, the results of the proposed EEG approach using NN classifier are much better than RF and SVM classifiers. On the other hand, the results of the RF classifier outperformed the other classifiers when using the raw EEG data.

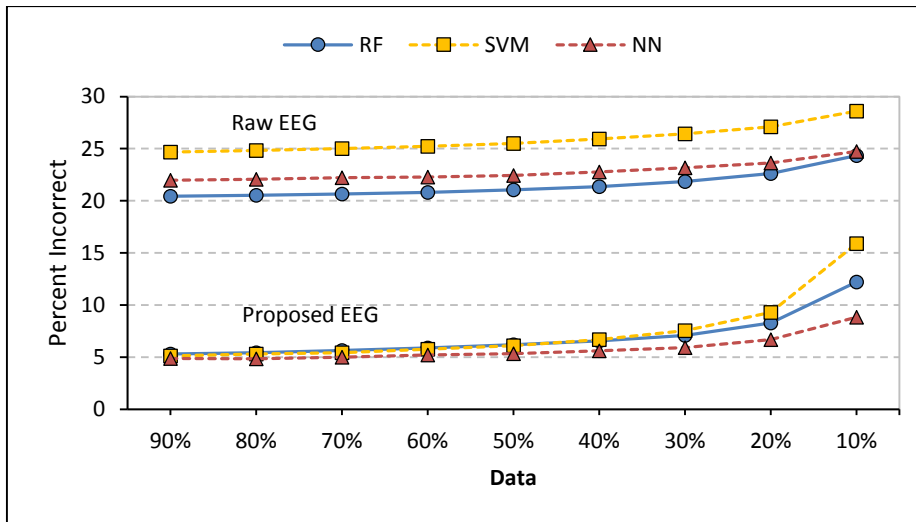


Figure 47: Percent incorrect with Different Size of Data

Figure 48 illustrates that the NN classifier with the proposed EEG+ECG features can correctly detect malicious activities with an error rate around 4.5% when using only 10% of the incoming data. On the other hand, the NN classifier has an error rate around 8% when using the proposed EEG features. Moreover, the figure illustrates that, when using less than 20% of the incoming data, the accuracy of the SVM classifier becomes the worst compared to the results of NN and RF classifiers. When using less than 15% of the incoming data, the NN classifier with the proposed EEG features achieved a lower error rate than the SVM classifier with the proposed EEG+ECG features. SVM achieved the worst results when the data is too small because SVM classifier tends to ignore the minority



class while building the model, especially when dealing with unbalanced data [104, 105]. This drawback of the SVM classifier appears clearly in the proposed approach only when we use less than 20% of the incoming data. Therefore, the SVM classifier is not recommended when there is too little data.

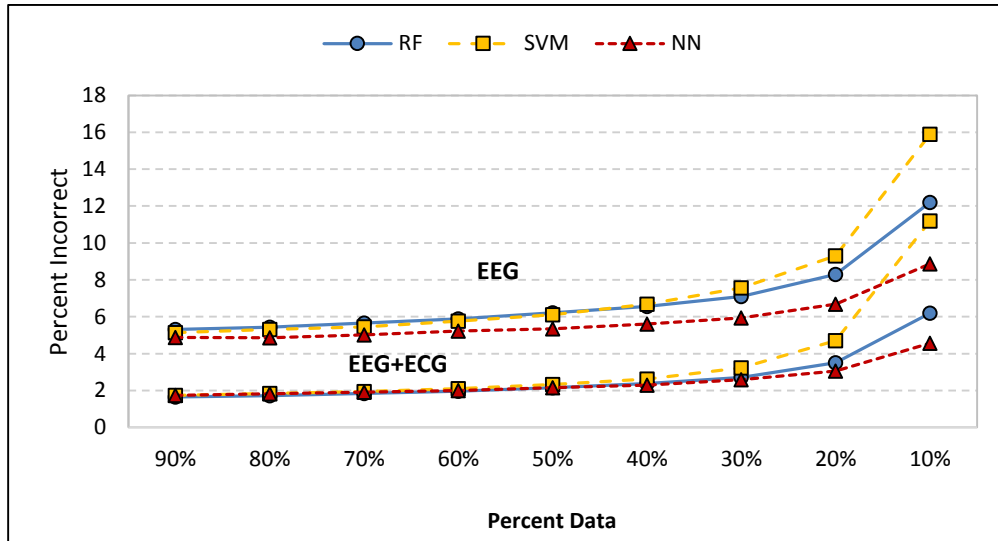


Figure 48: Incorrect Data of the Proposed EEG+ECG and the EEG Only

Time is an important factor in evaluating the performance of machine learning classifiers. The elapsed time to build a machine learning model for each classifier, known as the training time, varies. To show the elapsed training time versus the achieved results, the ROC curves and the average training time were drawn. The average training time is the time needed to build the model per person. Therefore, in this research, the total time per second to build the model for the classifier is the average training time multiplied by 84, which is the number of cases in our experiments.

Figures 49, 50, and 51 illustrate the average training time versus the ROC curves for each classifier when the classifiers are fed with 10% less data at a time. From the figures, using a group of five frames significantly reduced the time needed to build the classification

models. Additionally, the problem with the SVM results when there is 20% less incoming data affects the results of the SVM classifier when a group of five frames is used, but it does not affect the one-frame results (1F) of the EEG+ECG, as illustrated in Figure 50.

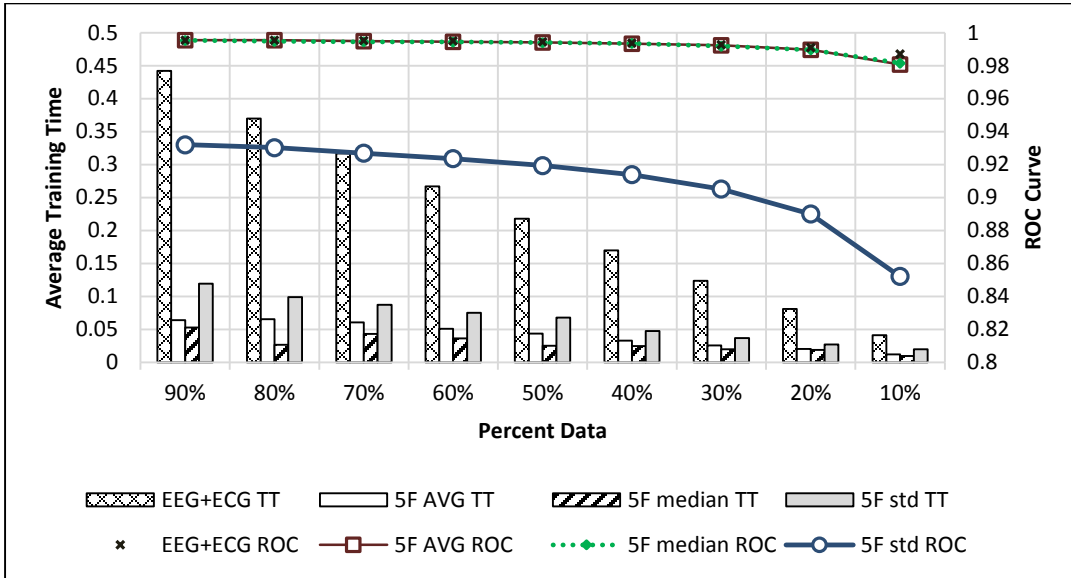


Figure 49: ROC Curves and the Training Time per Second Using RF

Moreover, there is no direct correlation between the results obtained and the training time. The average training time of the NN classifier is around 10 times more than that of the RF and the SVM classifiers. However, the ROC curve achieved by the NN classifier is slightly less than the ROC curve of the RF classifier, as illustrated in Figure 49 and Figure 51. Furthermore, although the standard deviation of a group of five frames (5F STD) achieved the worst accuracy results, it did not achieve the shortest training time: The average and median of five frames (5F AVG and 5F Median) achieved better results with an average training time less than that of 5F STD, as clearly shown in Figure 49.

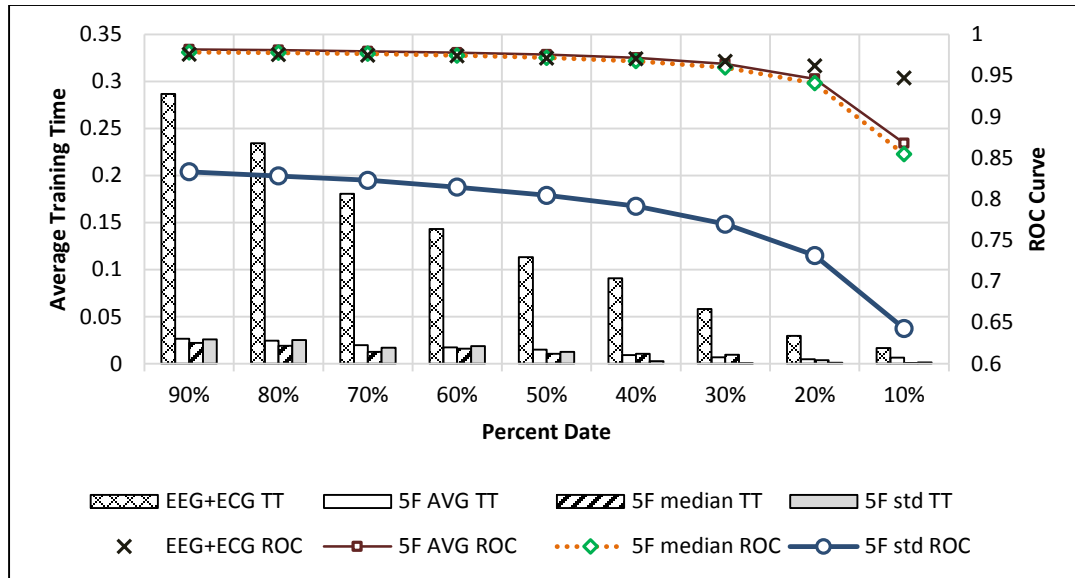


Figure 50: ROC Curves and the Training Time per Second Using SVM

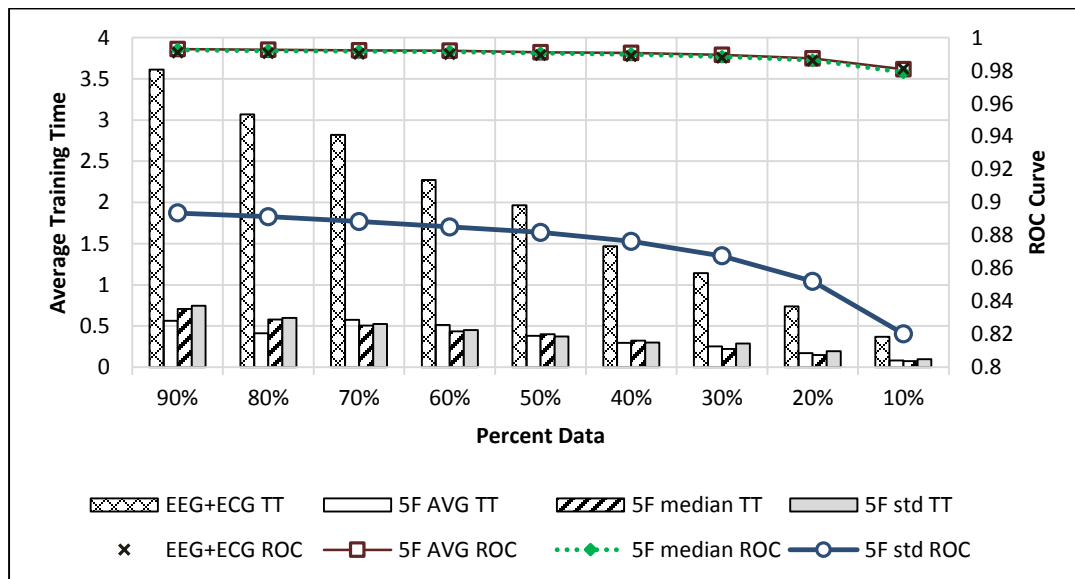


Figure 51: ROC Curves and the Training Time per Second Using NN

## 6.12 Evaluation for New Incoming Data

To evaluate the proposed features for the new incoming data, we tested the approach as a unit for the whole population. Two types of evaluation were used including 10 folds cross-validation of the whole participants' data and 70% Training – 30% Testing of the data. Moreover, to assess the detection quality when the number of frames is changed, the test includes the utilization of a different number of frames, where each session's frames were compressed as discussed in Section 6.5.

Table 17 illustrates the results of the proposed EEG+ECG approach that calculated over the whole population per session when each participant's session consists of whole frames (1 Frame), a group of (5, 10 and 15) Frames. The table shows the false positive (FP) and false negative (FN) rates, and the accuracy. From the table, we can notice that the proposed approach has no major difference in the accuracy when the number of frames is changed.

**Table 17: Confidence Intervale, FP and FN Rates of Proposed EEG+ECG Approach**

Frames	10 Folds Cross-Validation			70% Training, 30% Testing		
	FNR	FPR	Accuracy $\pm$ CI	FNR	FPR	Accuracy $\pm$ CI
1 Frame	4%	11%	93.38 $\pm$ 0.21	6.4%	14.4%	90.59 $\pm$ 0.14
5 Frames	10%	21%	85.75 $\pm$ 0.27	11.5%	22.5%	84.45 $\pm$ 0.18
10 Frames	9%	17.4%	87.88 $\pm$ 0.38	10.2%	21.4%	85.6 $\pm$ 0.34
15 Frames	7.7%	18.6%	88.26 $\pm$ 0.44	10.7%	19.6%	86 $\pm$ 0.40

Moreover, we evaluated the proposed features for the whole population per participants, using 10 folds cross-validation and 70% Training – 30% Testing of participants. In the cross-validation the participants were divided into 10 groups one of the groups used as a test whereas the other 9 groups used for training the model. The test repeated 10 times then

the average of the results is calculated. On the other test, the 70% of the participants used as Training – and 30% used as Testing. The test repeated 10 times where the 30% of participants in the testing selected randomly from the populations. Table 18 shows the results of the evaluation.

**Table 18: Results of EEG+ECG Approach for the Whole Population per Participant**

Folds	10 Folds Cross-Validation			70% Training, 30% Testing		
	FNR	FPR	Accuracy	FNR	FPR	Accuracy
1	22%	5%	87.87	24%	12%	83.0
2	21%	8%	86.37	22%	16%	81.10
3	19%	10%	86.04	31%	4%	85.74
4	35%	1%	85.91	24%	9%	84.93
5	17%	7%	88.60	26%	11%	83
6	18%	11%	85.75	30%	7%	84.4
7	31%	2%	87.87	26%	11%	82.82
8	15%	8%	88.91	23%	6%	87.1
9	23%	2%	90.46	20%	12%	84.6
10	16%	7%	89.40	36%	7%	82.12
<b>Avg</b>	22%	6%	87.72	26%	10%	83.88

Furthermore, to further evaluate the proposed approach for new incoming data, the approach evaluated for participant separately. To achieve this goal, only one participant is used as testing data whereas the others 83 participants used for training the model. Figure 52 illustrates the accuracy of each participant. The X axis illustrates the accuracy value whereas the Y axis illustrates the number of participants achieved this value. Moreover, Table 19 shows the average accuracy, FPR and FNR of the 84 participants where the detailed results of each participant is illustrated in Appendix B.

**Table 19: Average Accuracy, FPR and FNR of the 84 Participants**

Features	FNR	FPR	Accuracy	95% C.I.
EEG+ECG	19.7%	4.6%	90%	90 ± 1.26
EEG	19.3%	11.8%	85.6%	85.6 ± 2.27
Raw Brainwaves	62%	11.5%	69.88%	69.88 ± 1.97

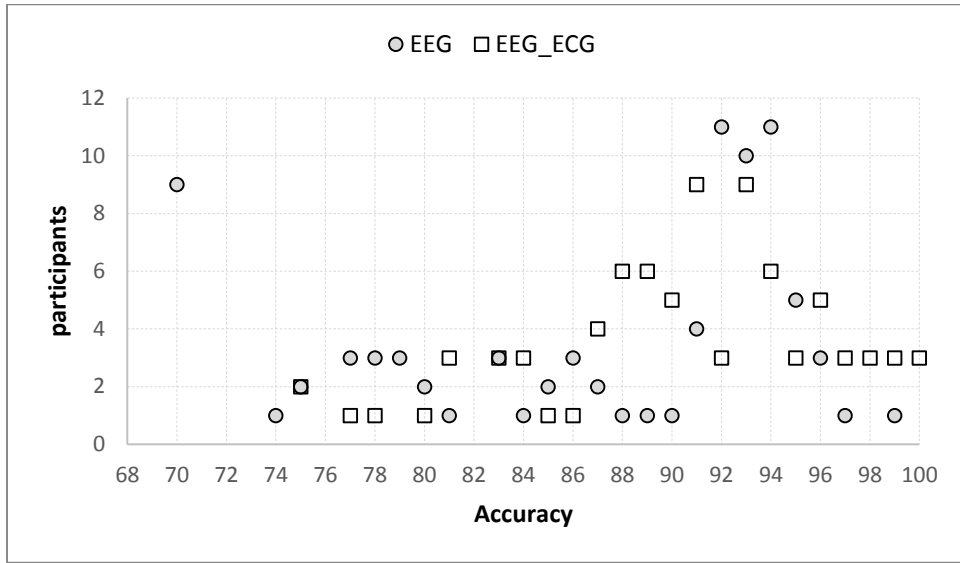


Figure 52: Accuracy per participant

To evaluate the efficacy of the proposed approach based on gender, Table 20 shows the average of accuracy, FPR and FNR for the males and the females. Where this average is calculated from the test of evaluated each participant separately (1 person test, 83 persons Train). From the table, we can notice that the results of detecting insider threats from females are better than the results for males when using the extracted EEG features. This can be attributed to the fact that females report more levels of fear and anxiety than males due to different factors, including biological and cultural influences [179, 180]. Moreover, this fact can be noticed when using a raw brainwaves. In contrast, when using the EEG+ECG features, the accuracy gap between genders diminished; this can be attributed to that the addition of the ECG features increases the distinction between the normal and suspicious activities in both genders, thus diminishing the gap between the results.

**Table 20: Comparing Approaches Based on Gender**

Method	EEG+ECG		EEG		Brainwaves	
	Males	Females	Males	Females	Males	Females
<b>FPR</b>	7.7%	1.4%	18%	4%	8%	14.3%
<b>FNR</b>	15%	23%	19%	19%	57%	68%
<b>Accuracy</b>	89.44	90.57	80.92	90.5	72.89%	66.7%
<b>95% C.I.</b>	89.44±1.7	90.57±1.8	80.92±3.3	90.5±2.3	72.89±2.65	66.7±2.7

### 6.13 Evaluating proposed method with Suh and Yim approach

To assess the performance of the proposed approach, we compared our approach with Suh and Yim approach [109] utilizing our dataset. Suh and Yim used a quantitative EEG analysis to develop two types of indicators which are the relative power of (alpha, beta, theta, the sum of alpha and theta, and gamma waves) and the ratio of brainwave-to-brainwave (i.e., gamma/alpha and beta/alpha). Figure 53 shows the feature frame of Suh and Yim approach.

Theta	Alpha	Beta	Gamma	Alpha + Theta	$\frac{\text{Beta}}{\text{Alpha}}$	$\frac{\text{Gamma}}{\text{Alpha}}$
-------	-------	------	-------	---------------	------------------------------------	-------------------------------------

**Figure 53: Feature Frame of Suh and Yim Approach**

Figure 54 and Figure 55 illustrate the comparison of the proposed approach with Suh's method using RF classifier. The comparison includes two techniques of classification which are the 10 folds cross-validation and 70% of the data for the training. From the figures, we can notice that the proposed EEG only and EEG+ECG features achieved better results than Suh's method in detecting insider threats. This can be attributed to the proposed feature-factors which are dividing the EEG waves into smaller ranges, Normalization and extract additional features from the influential EEG waves. Moreover, Figure 55 illustrates

that the proposed approach achieved better computation training time than Suh’s method, this can be attributed to the small values resulting from normalizing of extracted features compared to the un-normalized data in Suh's method.

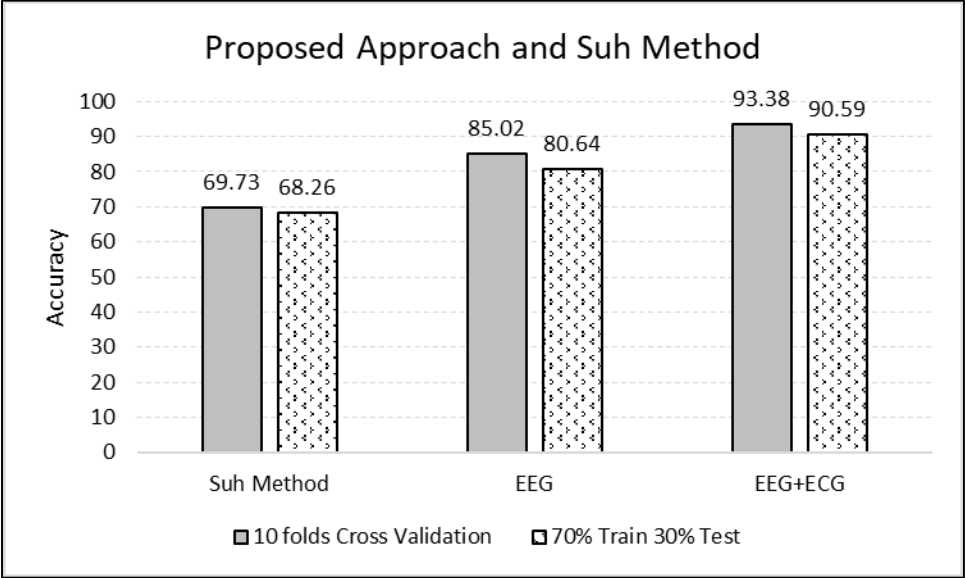


Figure 54: Comparing proposed approach with Suh’s Method using Accuracy

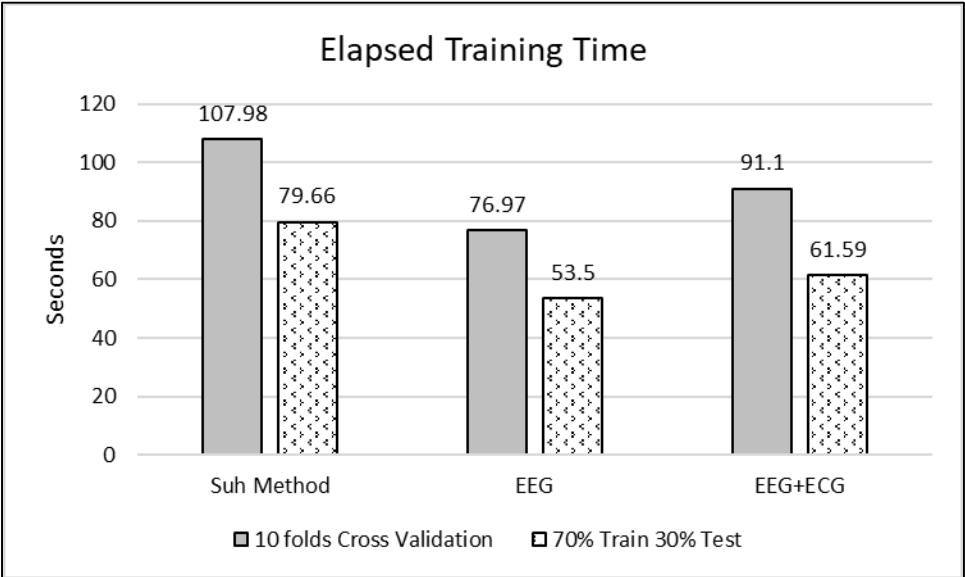


Figure 55: Comparing proposed approach with Suh’s Method using Training Time



## 6.14 Summary

Evaluating the proposed system is the final stage of the methodology of this research. To evaluate the proposed system, two null hypotheses were developed. Each hypothesis addressed the potential effect for a type of the bio-signals on distinguishing between the malicious and non-malicious activities. In this chapter, the null hypotheses have been rejected based on the accuracy of the classification results. Although the accuracy may give unreliable results when used with unbalanced classes of data, additional more reliable metrics were used to assess the correctness of the accuracy metric including precision, recall, area under the ROC curve, kappa, F-score, and Matthews correlation coefficients.

The classification results illustrate the ability of the proposed system to differentiate between malicious and normal activities with a considerable accuracy. The results show that the proposed approach can detect insider threats with an average accuracy up to 98.4%, which negates the validity of the null hypotheses and demonstrates the positive effect of the extracted features on detecting the insider threats.

Moreover, since the classification is a crucial phase in the proposed system, three classification algorithms were used to test and assess their influence on the results. The results illustrated that there is an insignificant difference between the accuracy of these classification algorithms in detecting the insider threats. Furthermore, the system has been evaluated when using a group of frames. Using a group of 5 frames showed some improvements in the classification results.

## **CHAPTER 7**

### **CONCLUSION, LIMITATIONS AND FUTURE**

#### **RESEARCH**

This research studies the feasibility of detecting insider threats by distinguishing between normal and malicious activities utilizing human bio-signals. Although the results of the proposed system are promising and worthy of future research, this research has some limitations which should be illustrated. This chapter concludes the dissertation, discusses the limitations and demonstrates the future research directions.

#### **7.1 Conclusion**

Insider threats are a considerable risk to organizations, more than external hackers, because insiders have more knowledge than outsiders do about the organization's system and its security mechanisms. Insiders are employees and trusted partners who have authorized access to the information and digital systems of the organization. Moreover, some insiders carefully plan out their intentions and deliberately take steps to put themselves in the best position to carry out these attacks. Therefore, the discovery of insider attacks is not easy, since it is hard to differentiate between these crimes and non-malicious activities.

Insider attacks cause extensive damage to organizations. Thinking about the added cost of the data breach, that is even more disconcerting and brings a bigger financial burden to an organization. The added costs come from a variety of sources: it is not just the financial

loss of that information, but it is everything from the cost of responding to that incident, to cleaning all damages, and installing preventative systems. There are also many tangible costs such as the loss of customer loyalty. Therefore, governments and organizations invest money and enact laws for mitigating the risk of such attacks.

This research aims to mitigate the risk of insider threats by proposing an insider attack detection system using human involuntary bio-signals. Human bio-signals are spontaneous signals done without will or self-control. Thus they are hard to imitate. Unfortunately, no available data set contains the bio-signals collected during malicious and normal activities of insiders, to the best of our knowledge. Therefore, this research provides a data set that contains sufficient samples of human bio-signals collected during real suspicious activities. To this end, two experiments were conducted to collect such signals. These experiments were based on two scenarios: normal and malicious activities.

In the first scenario, the human bio-signals were collected from participants while conducting normal work activities such as writing or thinking for solving problems. On the other hand, the second scenario was developed based on physiological rules, which state that intervention in decision-making and the sense of responsibility play an important role in influencing human bio-signals. So, in the second scenario, the appropriate environment and motivators have been provided for the participants to conduct malicious activities, leaving the decision-making to the participants to conduct such activities.

The bio-signals were assembled into a data set which is divided into two parts: normal activities and suspicious activities. Each part of the data set contains two types of bio-signals, which are: the electroencephalogram (EEG) and the electrocardiogram (ECG). In

order to support this search area, the data set was created to be public for conducting further researches.

To detect the insider threats using human bio-signals, thirteen features have been extracted from the collected EEG and ECG signals. These features represent the measurable characteristics of the bio-signals that can be used to distinguish between the normal and malicious behaviors.

Furthermore, an insider threat detection system was proposed in this study. The proposed system consists of eight units, which are: EEG sensors, ECG sensors, EEG interface, ECG interface, a feature extractor, attack assessment, comparative signals data set and attack evaluator. These units represent the different stages starting from the collecting of bio-signals to the stage of determining the attack. With the rapidity of technological advancement, new wearable techniques are developed to collect and analyze human bio-signals. It is worth mentioning that this research did not put any restriction on sensor type; any types of sensor could be used as long as they can collect bio-signals accurately and do not obstruct the employees' work.

The results show that the proposed EEG+ECG approach can detect insider threats with an average accuracy up to 98.4% when building a machine learning model for each participant, which means that less than 2% of all the incoming malicious signals were misclassified. Thus, classifiers can detect the insider threats even if the incoming data is too small. Furthermore, since the accuracy can be influenced by the small size of the malicious data class, the correctness of accuracy metric has been verified using several

reliable metrics. All in all, the results illustrate that the proposed system is effective for detecting insider threats.

Furthermore, we have tested the proposed EEG+ECG approach for new incoming data using several methods. When testing the whole data of participant using 10 folds cross-validation and 70% of the data for training, the proposed EEG+ECG achieved accuracy of 93.38%, and 90.59% respectively. Moreover, the proposed EEG+ECG approach has been tested for each participant individually by using the data for a single participant for testing whereas the data for the rest of 83 participants has been used for training the machine learning, the EEG+ECG features achieved an average accuracy of 90%.

In closing, this research study makes some contributions: developing an insider threat detection system that can accurately identify malicious activities, which will help organizations to detect insider attackers and to take the necessary actions to mitigate the risk of such attacks. Regardless of the insider attack mechanisms, the proposed system uses involuntary signals to detect such attacks. Also, this research work provides a data set which contains the physiological signals for 84 volunteers. The availability of the human physiological data set collected during normal and malicious activities for research use would provide an opportunity to develop and extend the research in this area.

## 7.2 Limitations

The results of this research study are promising and worthy of further research. However, this research has some limitations to be discussed. The following subsections clarify these limitations.

### 7.2.1 Sensors

Although the sensors used in this research to collect bio-signals have many features such as the small size, wearable, easy to connect to the computer, comfortable, cheap, and have been used in several types of research [131–133], but there are several devices that are more accurate and contain many sensors such as the 32-channel Biosemi headset [181], the 64-channel g.Nautilus [182], and the 14-channel Emotiv EPOC [100]. Using these devices to collect human bio-signals, more signals can be assembled with additional details about the signals collected, thereby enriching the data set and ensuring the reliability of the results. A useful comparison between several EEG devices was conducted by Nijboer *et al.* [183].

Furthermore, the devices used in this research to collect EEG and ECG signals are developed by different vendors. This led to the use of different interfaces to handle the collected bio-signals, which further complicated the analysis and storage of data in the data set. Utilizing devices created by the same company would facilitate the collection and analysis of the bio-signals. An example of a company that provides different devices for collecting human bio-signals is g.tec medical engineering [184].

### **7.2.2 Number of Used Devices**

In this research, only four devices were used to collect human bio-signals due to the limited funds. These are two devices for collecting EEG signals and two for ECG signals. In each experiment session, only two participants use the devices. However, in the second scenario, one of the participants is a true participant, as we discussed in Chapter 5. So, the bio-signals of only one participant were collected each time. The nature of the second scenario (suspicious activities) requires that the procedures of this scenario be confidential, so this scenario should be carried out during a short period. However, due to the limited number of devices, conducting this scenario took a longer time.

### **7.2.3 Hardware Limitations**

The EEG signals were collected using the NeuroSky device that runs on battery power and transmits the signals to the computer via Bluetooth. Such devices have lower data transmission range to minimize power consumption and called Class 2 Bluetooth devices [185]. Due to the range limitations, some EEG files in our data set contain very few bad signals which are not recorded in the file. In most cases, the bad signals between consecutive samples were five seconds at most and were symbolized in the file as NA (not available).

### **7.2.4 Unused Bio-signals**

This research focuses on using EEG and ECG signals to detect insider threats. It is worthy of mention that the skin conductance level was also assembled and stored in the data set to be used for further research. Moreover, the NeuroSky device which was utilized for collecting EEG signals produces additional data. These data are the meditation level, attention level, and eye blink. But the device's company did not reveal the algorithms that

have been used to produce these data. Therefore, these data were considered informal signals. These data were collected and not used or stored in the formal data set but will be available upon request.

### **7.2.5 Environment of collecting Data**

The experiments were conducted in the academic environment and the bio-signals were collected from university students and staffs. However, more bio-signals needs to be collected from industry environment to assess the implications of our research in the different environments.

### **7.2.6 Final Product Deployment**

Since the major aim of this research to propose a new approach to detect insider threats, we extracted features and achieved the classification results by (the batch mode) using two different software MATLAB and Java + WEKA APIs. Because the mentioned software are reliable and have great facilities. However, deploying the final product of the proposed approach should consist of integrated client and server applications.



### **7.3 Future Research**

Providing the human physiological data set collected during normal and malicious activities for research purposes would provide an opportunity to develop and extend the research in this area. Although the results of the proposed system offer excellent potential for detecting insider threats, the proposed system has not utilized all the collected bio-signals, because this study focused only on EEG and ECG signals. For instance, the skin conductivity level was collected during the experiments and stored in the bio-signals data set for further investigation. Moreover, the proposed system uses only three ECG features; more features can be extracted from the collected data, such as the heartbeat irregularity [186]. Further research is needed to refine the proposed system by extracting new features.

Furthermore, using devices that have many sensors will provide more accurate results and offer additional details on the collected human bio-signals. The more details on the bio-signals will provide additional features for detecting insider threats. Although the devices that have many sensors are expensive, we are planning to contact the vendors to denote the devices for a research purposes.

## References

1. Perlman, R., Kaufman, C., Speciner, M.: Network security: private communication in a public world. Pearson Education India (2016)
2. Malav, S., Avinash, M.S., Satish, N.S., Sandeep, S.C.: Network Security Using IDS, IPS & Honeypot. (2016)
3. Hu, H., Han, W., Ahn, G.-J., Zhao, Z.: FLOWGUARD: building robust firewalls for software-defined networks. In: Proceedings of the third workshop on Hot topics in software defined networking. pp. 97–102. ACM (2014)
4. Bishop, M., Nance, K., Claycomb, W.: Inside the Insider Threat (Introduction). (2016)
5. Nurse, J.R., Buckley, O., Legg, P.A., Goldsmith, M., Creese, S., Wright, G.R., Whitty, M.: Understanding insider threat: A framework for characterising attacks. In: Security and Privacy Workshops (SPW), 2014 IEEE. pp. 214–228. IEEE (2014)
6. SpectorSoft (2014) SpectorSoft 2014 insider threat survey, [https://media.scmagazine.com/documents/90/spectorsoft-2014-insider-threa\\_22395.pdf](https://media.scmagazine.com/documents/90/spectorsoft-2014-insider-threa_22395.pdf)
7. AlgoSec: The State of Network Security 2013: Attitudes and Opinions. (2013)
8. New Market Research - SolarWinds Survey Investi... |thwack, <https://thwack.solarwinds.com/thread/71368>
9. Greitzer, F.L., Kangas, L.J., Noonan, C.F., Dalton, A.C., Hohimer, R.E.: Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. In: System Science (HICSS), 2012 45th Hawaii International Conference on. pp. 2392–2401. IEEE (2012)
10. Schultz, E.E.: A framework for understanding and predicting insider attacks. *Computers & Security*. 21, 526–531 (2002)
11. Ambre, A., Shekokar, N.: Insider Threat Detection Using Log Analysis and Event Correlation. *Procedia Computer Science*. 45, 436–445 (2015)
12. D’mello, S.K., Kory, J.: A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)*. 47, 43 (2015)
13. Soleymani, M., Villaro-Dixon, F., Pun, T., Chanel, G.: Toolbox for Emotional feAture extraction from Physiological signals (TEAP). *Front. ICT*. 4, (2017). doi:10.3389/fict.2017.00001

14. Wioleta, S.: Using physiological signals for emotion recognition. In: 2013 6th International Conference on Human System Interactions (HSI). pp. 556–561 (2013)
15. Altop, D.K., Levi, A., Tuzcu, V.: Towards Using Physiological Signals As Cryptographic Keys in Body Area Networks. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. pp. 92–99. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium (2015)
16. Sanches, C.L., Augereau, O., Kise, K.: Manga Content Analysis Using Physiological Signals. In: Proceedings of the 1st International Workshop on coMics ANalysis, Processing and Understanding. pp. 6:1–6:6. ACM, New York, NY, USA (2016)
17. Sadoff, R.: The Evolution of Forensic Psychiatry: History, Current Developments, Future Directions. Oxford University Press (2015)
18. Suh, Y.-A., Yim, M.: An Investigation into the Applicability of Biodata, from Health Wearable Devices, to Insider Threat Detection in Nuclear Power Plants. In: 2016 annual INMM Conference, Atlanta, USA (2016)
19. Swan, M.: Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks*. 1, 217–253 (2012). doi:10.3390/jsan1030217
20. Arboleda Carpio, S.L., Sohail, S., Clark, K., Fagan, J.M.: Fitness Gadgets as a Form of Preventative Healthcare. (2016)
21. Hunker, J., Probst, C.W.: Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques. *JoWUA*. 2, 4–27 (2011)
22. Alpaydin, E.: Introduction to machine learning. MIT press (2014)
23. Witten, I.H., Frank, E., Hall, M.A., Pal, C.J.: Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann (2016)
24. Singer, P.W., Friedman, A.: Cybersecurity: What Everyone Needs to Know. Oxford University Press (2014)
25. Carter, W.A., Sofio, D.G., Alperen, M.J.: Cybersecurity legislation and critical infrastructure vulnerabilities. *Foundations of Homeland Security: Law and Policy*. 233–249 (2017)
26. Almehmadi, A.: On the potential of intent-based access control (IBAC) in preventing insider threats, (2015)
27. Keramati, H., Mirian-Hosseiniabadi, S.: Integrating software development security activities with agile methodologies. In: 2008 IEEE/ACS International Conference on Computer Systems and Applications. pp. 749–754 (2008)

28. Sahebjamnia, N., Torabi, S.A., Mansouri, S.A.: Integrated business continuity and disaster recovery planning: Towards organizational resilience. *European Journal of Operational Research*. 242, 261–273 (2015). doi:10.1016/j.ejor.2014.09.055
29. Chandra, S., Paira, S., Alam, S.S., Sanyal, G.: A comparative survey of Symmetric and Asymmetric Key Cryptography. In: *Electronics, Communication and Computational Engineering (ICECCE), 2014 International Conference on*. pp. 83–93. IEEE (2014)
30. Bodin, L.D., Gordon, L.A., Loeb, M.P.: Information Security and Risk Management. *Commun. ACM*. 51, 64–68 (2008). doi:10.1145/1330311.1330325
31. Owens, W.A., Dam, K.W., Lin, H.S.: *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*. National Academy Press, Washington, DC, USA (2009)
32. McCrie, R.: *Security Operations Management*. Butterworth-Heinemann, Newton, MA, USA (2016)
33. Fennelly, L.: *Effective Physical Security*. Butterworth-Heinemann, Newton, MA, USA (2012)
34. Ferraiolo, D., Cugini, J., Kuhn, D.R.: Role-based access control (RBAC): Features and motivations. In: *Proceedings of 11th annual computer security application conference*. pp. 241–48 (1995)
35. CMMI Institute - Home, <https://cmmiinstitute.com/>
36. Pieprzyk, J., Hardjono, T., Seberry, J.: *Fundamentals of computer security*. Springer Science & Business Media (2013)
37. Walton, R., Limited, W.-M.: Balancing the insider and outsider threat. *Computer Fraud & Security*. 2006, 8–11 (2006). doi:10.1016/S1361-3723(06)70440-7
38. Chang, S.-Y., Hu, Y.-C.: SecureMAC: Securing Wireless Medium Access Control Against Insider Denial-of-Service Attacks. *IEEE Transactions on Mobile Computing*. (2017)
39. Cappelli, D.M., Moore, A.P., Trzeciak, R.F.: *The CERT guide to insider threats: how to prevent, detect, and respond to information technology crimes (Theft, Sabotage, Fraud)*. Addison-Wesley (2012)
40. Lewis, J.A.: *The cyber war has not begun*. Center for Strategic and International Studies. (2010)
41. Cherkashin, V., Feifer, G.: *Spy handler: memoir of a KGB officer: the true story of the man who recruited Robert Hanssen and Aldrich Ames*. Basic Books (2008)

42. Shannon, E., Blackman, A.: *The spy next door: The extraordinary secret life of Robert Philip Hanssen, the most damaging FBI agent in US history.* Hachette UK (2008)
43. Rai, G.: *The Vulnerability to Fraud: Factors, Motivations, and Fraud Detection and Deterrence.* (2017)
44. Duncan, A., Creese, S., Goldsmith, M.: An overview of insider attacks in cloud computing. *Concurrency and Computation: Practice and Experience.* 27, 2964–2981 (2015)
45. Fisher, K.: *The Psychology of Fraud: What Motivates Fraudsters to Commit Crime?* (2015)
46. Carmichael, S.: *True believer: Inside the investigation and capture of Ana Montes, Cuba's master spy.* Naval Institute Press (2012)
47. Prevelakis, V., Spinellis, D.: The athens affair. *Spectrum, IEEE.* 44, 26–33 (2007)
48. Fenster, M.: Disclosure's effects: WikiLeaks and transparency. *Iowa L. Rev.* 97, 753 (2011)
49. Sifry, M.L.: *WikiLeaks and the Age of Transparency.* OR Books (2011)
50. Dingley, J.C.: The road to peace? Northern Ireland after the Belfast Agreement: causes of failure. *Democracy and Security.* 2, 263–286 (2006)
51. Neill, W.J.: Return to Titanic and Lost in the Maze: The search for Representation of 'Post-conflict' Belfast. *Space and Polity.* 10, 109–120 (2006)
52. Salvatore, S.: Psychological Evaluations for the US Army Biological Personnel Reliability Program. *Journal of Biosecurity, Biosafety, and Biodefense Law.* 8, 3–17 (2017)
53. Davies, S.E.: *Biosecurity dilemmas: dreaded diseases, ethical responses, and the health of nations.* Oxford University Press (2017)
54. Flower, R.J.: The outsider: The rogue scientist as terrorist. *Journal of medical ethics.* 40, 282–283 (2014)
55. Brackney, R.C., Anderson, R.H.: *Understanding the Insider Threat. Proceedings of a March 2004 Workshop.* RAND CORP SANTA MONICA CA (2004)
56. Heath, L.J.: *An Analysis of the Systemic Security Weaknesses of the US Navy Fleet Broadcasting System, 1967-1974, as Exploited by CWO John Walker.* Army Command and General Staff Coll Fort Leavenworth KS (2005)
57. 2018 Insider Threat Report, <https://www.securonix.com/2018-insider-threat-report/>

58. Randazzo, M.R., Keeney, M., Kowalski, E., Cappelli, D.M., Moore, A.P.: Insider threat study: Illicit cyber activity in the banking and finance sector. (2005)
59. Greitzer, F.L., Ferryman, T.A.: Methods and metrics for evaluating analytic insider threat tools. In: Security and Privacy Workshops (SPW), 2013 IEEE. pp. 90–97. IEEE (2013)
60. Cappos, J., Weiss, R.: Teaching the security mindset with reference monitors. In: Proceedings of the 45th ACM technical symposium on Computer science education. pp. 523–528. ACM (2014)
61. Ortiz, E., Reinerman-Jones, L., Matthews, G.: Developing an Insider Threat Training Environment. In: Advances in Human Factors in Cybersecurity. pp. 267–277. Springer, Cham (2016)
62. Chi, H., Allen, C., Rubio, D.A.: Design Insider Threat Hands-on Labs. (2015)
63. Thompson, M.F., Irvine, C.E.: Active Learning with the CyberCIEGE Video Game. In: CSET (2011)
64. Krutz, D.E., Meneely, A., Malachowsky, S.A.: An insider threat activity in a software security course. In: Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE. pp. 1–6. IEEE (2015)
65. Bhuyan, M.H., Bhattacharyya, D.K., Kalita, J.K.: Network anomaly detection: methods, systems and tools. IEEE communications surveys & tutorials. 16, 303–336 (2014)
66. Zargar, A., Nowroozi, A., Jalili, R.: XABA: A zero-knowledge anomaly-based behavioral analysis method to detect insider threats. In: Information Security and Cryptology (ISCISC), 2016 13th International Iranian Society of Cryptology Conference on. pp. 26–31. IEEE (2016)
67. Ring, M., Wunderlich, S., Grüdl, D., Landes, D., Hotho, A.: A Toolset for Intrusion and Insider Threat Detection. In: Data Analytics and Decision Support for Cybersecurity. pp. 3–31. Springer (2017)
68. Legg, P.A., Buckley, O., Goldsmith, M., Creese, S.: Automated insider threat detection system using user and role-based profile assessment. IEEE Systems Journal. 11, 503–512 (2017)
69. Young, W.T., Memory, A., Goldberg, H.G., Senator, T.E.: Detecting unknown insider threat scenarios. In: Security and Privacy Workshops (SPW), 2014 IEEE. pp. 277–288. IEEE (2014)
70. Defense Advanced Research Projects Agency 2010. Anomaly Detection at Multiple Scales (ADAMS) Broad Agency Announcement DARPA-BAA-11-04. Arlington VA. Presented at the

71. Han, W., Zhao, Z., Doupé, A., Ahn, G.-J.: Honeymix: Toward sdn-based intelligent honeynet. In: Proceedings of the 2016 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization. pp. 1–6. ACM (2016)
72. Stockman, M., Heile, R., Rein, A.: An open-source honeynet system to study system banner message effects on hackers. In: Proceedings of the 4th Annual ACM Conference on Research in Information Technology. pp. 19–22. ACM (2015)
73. Barbu, I.-D., Petrica, G., Axinte, S.-D., Bacivarov, I.: Analyzing Cyber Threat Actors of E-Learning Platforms by The Use of a Cloud Based Honeynet. *eLearning & Software for Education*. 3, (2017)
74. Hecker, C., Hay, B.: Automated honeynet deployment for dynamic network environment. In: System Sciences (HICSS), 2013 46th Hawaii International Conference on. pp. 4880–4889. IEEE (2013)
75. Memari, N., Hashim, S.J.B., Samsudin, K.B.: Towards virtual honeynet based on LXC virtualization. In: Region 10 Symposium, 2014 IEEE. pp. 496–501. IEEE (2014)
76. Fan, W., Fernández, D., Du, Z.: Adaptive and flexible virtual honeynet. In: International Conference on Mobile, Secure and Programmable Networking. pp. 1–17. Springer (2015)
77. Fan, W., Du, Z., Fernandez, D.: Taxonomy of honeynet solutions. In: SAI Intelligent Systems Conference (IntelliSys), 2015. pp. 1002–1009. IEEE (2015)
78. Sqalli, M., Al-Shaikh, R., Ahmed, E.: Towards Simulating a Virtual Distributed Honeynet at KFUPM: A Case Study. In: Computer Modeling and Simulation (EMS), 2010 Fourth UKSim European Symposium on. pp. 316–321. IEEE (2010)
79. Sqalli, M., AlShaikh, R., Ahmed, E.: A distributed honeynet at KFUPM: a case study. In: International Workshop on Recent Advances in Intrusion Detection. pp. 486–487. Springer (2010)
80. Virvilis, N., Serrano, O.S., Vanautgaerden, B.: Changing the game: The art of deceiving sophisticated attackers. In: Cyber Conflict (CyCon 2014), 2014 6th International Conference On. pp. 87–97. IEEE (2014)
81. Akoglu, L., Tong, H., Koutra, D.: Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*. 29, 626–688 (2015)
82. Kent, A.D., Liebrock, L.M., Neil, J.C.: Authentication graphs: Analyzing user behavior within an enterprise network. *Computers & Security*. 48, 150–166 (2015)
83. Zhou, D., Karthikeyan, A., Wang, K., Cao, N., He, J.: Discovering rare categories from graph streams. *Data Mining and Knowledge Discovery*. 31, 400–423 (2017)

84. Mongiovi, M., Bogdanov, P., Ranca, R., Papalexakis, E.E., Faloutsos, C., Singh, A.K.: Netspot: Spotting significant anomalous regions on dynamic networks. In: Proceedings of the 2013 SIAM International Conference on Data Mining. pp. 28–36. SIAM (2013)
85. Lamba, H., Glazier, T.J., Schmerl, B., Pfeffer, J., Garlan, D.: Detecting insider threats in software systems using graph models of behavioral paths. In: Proceedings of the 2015 Symposium and Bootcamp on the Science of Security. p. 20. ACM (2015)
86. Feng, X., Zheng, Z., Hu, P., Cansever, D., Mohapatra, P.: Stealthy attacks meets insider threats: a three-player game model. In: Military Communications Conference, MILCOM 2015-2015 IEEE. pp. 25–30. IEEE (2015)
87. Van Dijk, M., Juels, A., Oprea, A., Rivest, R.L.: FlipIt: The game of “stealthy takeover.” *Journal of Cryptology*. 26, 655–713 (2013)
88. Kim, K.-N., Yim, M.-S., Schneider, E.: A study of insider threat in nuclear security analysis using game theoretic modeling. *Annals of Nuclear Energy*. 108, 301–309 (2017)
89. Sherwood, L.: *Human physiology: from cells to systems*. Cengage learning (2015)
90. Hall, J.E.: *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences (2015)
91. Smith, E.A.: *Evolutionary ecology and human behavior*. Routledge (2017)
92. Zipf, G.K.: *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio Books (2016)
93. Greitzer, F.L., Imran, M., Purl, J., Axelrad, E.T., Leong, Y.M., Becker, D.E., Laskey, K.B., Sticha, P.J.: Developing an Ontology for Individual and Organizational Sociotechnical Indicators of Insider Threat Risk. In: STIDS. pp. 19–27 (2016)
94. Lumini, A., Nanni, L.: Overview of the combination of biometric matchers. *Information Fusion*. 33, 71–85 (2017)
95. Ngo, D.C.L., Teoh, A.B.J., Hu, J.: *Biometric security*. Cambridge Scholars Publishing (2015)
96. Matthews, G., Reinerman-Jones, L., Wohleber, R., Ortiz, E.: Eye Tracking Metrics for Insider Threat Detection in a Simulated Work Environment. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. pp. 202–206. SAGE Publications Sage CA: Los Angeles, CA (2017)



97. Babu, B.M., Bhanu, M.S.: Prevention of Insider Attacks by Integrating Behavior Analysis with Risk based Access Control Model to Protect Cloud. *Procedia Computer Science*. 54, 157–166 (2015)
98. Rudrapal, D., Das, S., Debbarma, N., Debbarma, S.: Internal attacker detection by analyzing user keystroke credential. *Lecture Notes on Software Engineering*. 1, 49 (2013)
99. Hashem, Y., Takabi, H., GhasemiGol, M., Dantu, R.: Towards Insider Threat Detection Using Psychophysiological Signals. In: *Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats*. pp. 71–74. ACM (2015)
100. Wearables for your brain | EEG, <https://emotiv.com/>
101. Abdi, H., Williams, L.J.: Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2, 433–459 (2010)
102. Bro, R., Smilde, A.K.: Principal component analysis. *Analytical Methods*. 6, 2812–2831 (2014)
103. Morse, J.M.: *Determining sample size*. Sage Publications Sage CA: Thousand Oaks, CA (2000)
104. Sun, Y., Kamel, M.S., Wong, A.K.C., Wang, Y.: Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*. 40, 3358–3378 (2007). doi:10.1016/j.patcog.2007.04.009
105. Cao, W., Dong, G., Xie, Y.-B., Peng, Z.: Prediction of wear trend of engines via on-line wear debris monitoring. *Tribology International*. 120, 510–519 (2018). doi:10.1016/j.triboint.2018.01.015
106. Almeahmadi, A., El-Khatib, K.: On the possibility of insider threat prevention using intent-based access control (IBAC). *IEEE Systems Journal*. 11, 373–384 (2017)
107. Almeahmadi, A., El-Khatib, K.: On the possibility of insider threat detection using physiological signal monitoring. In: *Proceedings of the 7th International Conference on Security of Information and Networks*. p. 223. ACM (2014)
108. Hashem, Y., Takabi, H., Dantu, R., Nielsen, R.: A multi-modal neuro-physiological study of malicious insider threats. In: *Proceedings of the 2017 International Workshop on Managing Insider Security Threats*. pp. 33–44. ACM (2017)
109. Suh, Y.A., Yim, M.-S.: “High risk non-initiating insider” identification based on EEG analysis for enhancing nuclear security. *Annals of Nuclear Energy*. 113, 308–318 (2018)

110. Borders, K.R.: Method, system and computer program product for detecting at least one of security threats and undesirable computer files, (2015)
111. Oberheide, J., Cooke, E., Jahanian, F.: Rethinking Antivirus: Executable Analysis in the Network Cloud. In: HotSec (2007)
112. Sachan, A., Panchagavi, R.: Honeypots: Sweet OR Sour spot in Network Security? (2016)
113. Pandya, S.S.: Active Defence System for Network Security— Honeypot. *Advances in Computer Science and Information Technology (ACSIT)*. 2, 383–386 (2015)
114. Danchenko, N.M., Prokofiev, A.O., Silnov, D.S.: Detecting suspicious activity on remote desktop protocols using Honeypot system. In: *Young Researchers in Electrical and Electronic Engineering (EIConRus), 2017 IEEE Conference of Russian*. pp. 127–128. IEEE (2017)
115. Liang, X., Xiao, Y.: Game theory for network security. *IEEE Communications Surveys & Tutorials*. 15, 472–486 (2013)
116. Manshaei, M.H., Zhu, Q., Alpcan, T., Başçar, T., Hubaux, J.-P.: Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*. 45, 25 (2013)
117. Chinchani, R., Iyer, A., Ngo, H.Q., Upadhyaya, S.: Towards a theory of insider threat assessment. In: *Dependable Systems and Networks, 2005. DSN 2005. Proceedings. International Conference on*. pp. 108–117. IEEE (2005)
118. Azaria, A., Richardson, A., Kraus, S., Subrahmanian, V.S.: Behavioral analysis of insider threat: A survey and bootstrapped prediction in imbalanced data. *IEEE Transactions on Computational Social Systems*. 1, 135–155 (2014)
119. Ofole, N.M.: Self-acceptance of students repeating classes in Ibadan Metropolis: relationship with parents' sense-of-competence, locus of control and quality of parents-child-relationship. *IFE PsycholIA: An International Journal*. 25, 133–150 (2017)
120. Jones, G.E., Kavanagh, M.J.: An Experimental Examination of the Effects of Individual and Situational Factors on Unethical Behavioral Intentions in the Workplace. In: *Citation Classics from the Journal of Business Ethics*. pp. 657–674. Springer (2013)
121. Frith, C.D.: Action, agency and responsibility. *Neuropsychologia*. 55, 137–142 (2014)
122. Deci, E.L., Ryan, R.M.: Intrinsic motivation and self-determination in human behavior. *Perspectives in social psychology*. (1985)

123. Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., Sirigu, A.: The involvement of the orbitofrontal cortex in the experience of regret. *Science*. 304, 1167–1170 (2004)
124. Webster, J.G., Eren, H.: *Measurement, instrumentation, and sensors handbook: spatial, mechanical, thermal, and radiation measurement*. CRC press (2017)
125. Thakur, M., Dofe, R., Jadhav, S.: *Flexible electronic skin*. (2014)
126. NeuroSky - Home Page Support, <http://support.neurosky.com/kb/science/thinkgear-measurements-mindset-protgem>
127. NeuroExperimenter, <https://store.neurosky.com/products/neuroexperimenter>
128. Weka 3 - Data Mining with Open Source Machine Learning Software in Java, <https://www.cs.waikato.ac.nz/ml/weka/index.html>
129. Weka API, <http://weka.sourceforge.net/doc.stable/>
130. Net Monitor for Employees, <http://networklookout.com>
131. Robbins, R., Stonehill, M.: *Investigating the NeuroSky MindWave™ EEG Headset*. Transport Research Foundation. 1, (2014)
132. Abo-Zahhad, M., Ahmed, S.M., Abbas, S.N.: A novel biometric approach for human identification and verification using eye blinking signal. *IEEE Signal Processing Letters*. 22, 876–880 (2015)
133. Lim, C.-K.A., Chia, W.C.: Analysis of single-electrode EEG rhythms using MATLAB to elicit correlation with cognitive stress. *International Journal of Computer Theory and Engineering*. 7, 149 (2015)
134. Wild Divine Support, <http://support.wilddivine.com/>
135. Teplan, M.: Fundamentals of EEG measurement. *Measurement science review*. 2, 1–11 (2002)
136. Bittner, K.: *Spectral Analysis of Pre-and Post-Ictal Periods using Intracranial EEG*. (2017)
137. De, A., Konar, A., Samanta, A., Biswas, S., Basak, P.: Seizure prediction using low frequency EEG waves from WAG/Rij rats. In: *Convergence in Technology (I2CT), 2017 2nd International Conference for*. pp. 244–249. IEEE (2017)
138. Kleen, J.K., Lowenstein, D.H.: Progress in Epilepsy: Latest Waves of Discovery. *Jama neurology*. 74, 139–140 (2017)
139. Nunez, P.L., Srinivasan, R.: *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA (2006)

140. Herrmann, C.S., Strüber, D., Helfrich, R.F., Engel, A.K.: EEG oscillations: From correlation to causality. *International Journal of Psychophysiology*. 103, 12–21 (2016). doi:10.1016/j.ijpsycho.2015.02.003
141. Campisi, P., La Rocca, D.: Brain waves for automatic biometric-based user recognition. *IEEE transactions on information forensics and security*. 9, 782–800 (2014)
142. Ismail, W.W., Hanif, M., Mohamed, S.B., Hamzah, N., Rizman, Z.I.: Human emotion detection via brain waves study by using electroencephalogram (EEG). *International Journal on Advanced Science, Engineering and Information Technology*. 6, 1005–1011 (2016)
143. Buzsaki, G.: *Rhythms of the Brain*. Oxford University Press (2006)
144. Rangayyan, R.M.: *Biomedical signal analysis*. John Wiley & Sons (2015)
145. Lewis, M.C., Maiya, M., Sampathila, N.: A Novel Method for the Conversion of Scanned Electrocardiogram (ECG) Image to Digital Signal. In: Dash, S.S., Das, S., and Panigrahi, B.K. (eds.) *International Conference on Intelligent Computing and Applications*. pp. 363–373. Springer Singapore (2018)
146. Bauer, A., Camm, A.J., Cerutti, S., Guzik, P., Huikuri, H., Lombardi, F., Malik, M., Peng, C.-K., Porta, A., Sassi, R.: Reference values of heart rate variability. *Heart rhythm*. 14, 302–303 (2017)
147. What is Heart Rate Variability (HRV)? And how it can enhance your training, <https://www.myithlete.com/what-is-hrv/>
148. Ernst, G.: *Heart rate variability*. Springer (2016)
149. SweetWater Health™ Getting Started, <http://www.sweetwaterhrv.com/getstarted.shtml>
150. Goel, S., Kaur, G., Tomar, P.: Performance analysis of Welch and Blackman Nuttall window for noise reduction of ECG. In: *Signal Processing, Computing and Control (ISPC), 2015 International Conference on*. pp. 87–91. IEEE (2015)
151. Abburi, R., Chandrasekhara Sastry, A.S.: Fpga based fetal ecg feature extraction for prenatal monitoring using hybrid method. *J. Adv. Res. Dyn. Control Syst*. 9, 69–90 (2012)
152. Sammut, C., Webb, G.I.: *Encyclopedia of machine learning*. Springer Science & Business Media (2011)
153. Zhou, Y., Mason, S.G., Birch, G.E.: Impact of an energy normalization transform on the performance of the LF-ASD brain computer interface. In: *Advances in Neural Information Processing Systems*. pp. 725–732 (2004)

154. Ulrich, G.: The Theoretical Interpretation of Electroencephalography (EEG): The Important Role of Spontaneous Resting EEG and Vigilance: Electroencephalography for Neuropsychiatrists, Neurologists, Clinical Psychologists, Neuropsychologists, Psychophysicologists, and Neuroscientists. Bmed Press Llc (2013)
155. Mitchell, T.M.: Machine learning. McGraw-Hill Boston, MA: (1997)
156. Kannan, R., Vasanthi, V.: Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease. In: Muppalaneni, N.B., Ma, M., and Gurumoorthy, S. (eds.) *Soft Computing and Medical Bioinformatics*. pp. 63–72. Springer Singapore, Singapore (2019)
157. Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D.: Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971. (2015)
158. Belgiu, M., Drăguț, L.: Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*. 114, 24–31 (2016)
159. Resende, P.A.A., Drummond, A.C.: A Survey of Random Forest Based Methods for Intrusion Detection Systems. *ACM Comput. Surv.* 51, 48:1–48:36 (2018). doi:10.1145/3178582
160. Ben-Hur, A., Weston, J.: A user’s guide to support vector machines. In: *Data mining techniques for the life sciences*. pp. 223–239. Springer (2010)
161. Bloodgood, M.: Support Vector Machine Active Learning Algorithms with Query-by-Committee Versus Closest-to-Hyperplane Selection. In: *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. pp. 148–155 (2018)
162. Wang, L., Zeng, Y., Chen, T.: Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications*. 42, 855–863 (2015)
163. Ranganathan, V., Natarajan, S.: A New Backpropagation Algorithm without Gradient Descent. arXiv:1802.00027 [cs]. (2018)
164. Panchal, G., Ganatra, A., Shah, P., Panchal, D.: Determination of over-learning and over-fitting problem in back propagation neural network. *International Journal on Soft Computing*. 2, 40–51 (2011)
165. Liu, M., Wang, M., Wang, J., Li, D.: Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical*. 177, 970–980 (2013)

166. Encyclopedia of Machine Learning | SpringerLink, <https://link.springer.com/referencework/10.1007%2F978-0-387-30164-8>
167. Montgomery, D.C.: Design and analysis of experiments. John Wiley & Sons (2017)
168. Li, X., Dvornek, N.C., Papademetris, X., Zhuang, J., Staib, L.H., Ventola, P., Duncan, J.S.: 2-Channel convolutional 3D deep neural network (2CC3D) for fMRI analysis: ASD classification and feature learning. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 1252–1255 (2018)
169. A Robust profit measure for binary classification model evaluation - ScienceDirect, <https://www.sciencedirect.com/science/article/pii/S0957417417306498>
170. Pontius Jr, R.G., Millones, M.: Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*. 32, 4407–4429 (2011)
171. Viera, A.J., Garrett, J.M.: Understanding interobserver agreement: the kappa statistic. *Fam Med*. 37, 360–363 (2005)
172. Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*. 405, 442–451 (1975)
173. Sperschneider, J., Dodds, P.N., Singh, K.B., Taylor, J.M.: ApoplastP: prediction of effectors and plant proteins in the apoplast using machine learning. *New Phytologist*. 217, 1764–1778 (2017). doi:10.1111/nph.14946
174. Shi, L., Campbell, G., Jones, W.D., Campagne, F., Wen, Z., Walker, S.J., Su, Z., Chu, T.-M., Goodsaid, F.M., Puzstai, L.: The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature biotechnology*. 28, 827 (2010)
175. Törnqvist, L., Vartia, P., Vartia, Y.O.: How should relative changes be measured? *The American Statistician*. 39, 43–46 (1985)
176. GreedyStepwise, <http://weka.sourceforge.net/doc.stable/>
177. WrapperSubsetEval, <http://weka.sourceforge.net/doc.stable/>
178. Kohavi, R., John, G.H.: Wrappers for feature subset selection. *Artificial intelligence*. 97, 273–324 (1997)
179. McLean, C.P., Anderson, E.R.: Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review*. 29, 496–505 (2009). doi:10.1016/j.cpr.2009.05.003

180. Fischer, A.H., Rodriguez Mosquera, P.M., Van Vianen, A.E., Manstead, A.S.: Gender and culture differences in emotion. *Emotion*. 4, 87 (2004)
181. Biosemi EEG ECG EMG BSPM NEURO amplifier electrodes, <https://www.biosemi.com/products.htm>
182. g.tec: g.tec medical engineering, <http://www.gtec.at/Products/Hardware-and-Accessories/g.Nautilus-Specs-Features>
183. Nijboer, F., Van De Laar, B., Gerritsen, S., Nijholt, A., Poel, M.: Usability of three electroencephalogram headsets for brain–computer interfaces: a within subject comparison. *Interacting with computers*. 27, 500–511 (2015)
184. g.tec: g.tec medical engineering, <http://www.gtec.at/Products>
185. Kellerman, A., Gurusinghe, N., Ariyaratna, T., Gouws, R.: Smart whiteboard for interactive learning. In: 2018 IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES). pp. 515–520 (2018)
186. Kanani, P., Padole, M.: Recognizing Real Time ECG Anomalies Using Arduino, AD8232 and Java. In: Singh, M., Gupta, P.K., Tyagi, V., Flusser, J., and Ören, T. (eds.) *Advances in Computing and Data Sciences*. pp. 54–64. Springer Singapore (2018)





```

pk=1; % counter for peak
na=0; % counter for directories names
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Modify file names so they are handled easily
for sub_dir_index=3:length(sub_directories)
    csvs=dir(fullfile(Root_directory,sub_directories(sub_d
ir_index).name));
    na=na+1;
    sub_name{na}=sub_directories(sub_dir_index).name;
    n=sub_name{na};
    % change any space in the filename with _
    h='';
    for f=1:length(n)
        if(n(f)==' ')
            h(f)='_';
        elseif (n(f)=='.')
            break
        else
            h(f)=n(f);
        end
    end
    display(h)
    sub_name{na}=h;
%Read the data from files
    for j=3:length(csvs)
        filename = fullfile(csvs(j,1).name);
%Get The ECG Data
        if strcmp(filename,'sample.csv')
            s=strcat(Root_directory,'\ ',sub_directories(sub_d
ir_index).name,'\ ',csvs(j,1).name);
            fileID = fopen(s);
            C = textscan(fileID,'%*f %*f %*f %*f %s %f
%f','Delimiter',';', 'TreatAsEmpty',{'NA','na'}, 'C
ommentStyle','#');
            xsample{sa}=C;
            sa=sa+1;
            fclose(fileID);
        end

        if strcmp(filename,'sdsn.csv')
            s=strcat(Root_directory,'\ ',sub_directories(sub_d
ir_index).name,'\ ',csvs(j,1).name);
            fileID = fopen(s);
            Q = textscan(fileID,'%*f %*f %*f
%f','Delimiter',';', 'TreatAsEmpty',{'NA','na'}, 'C
ommentStyle','#');
            xsdsn{sd}=Q;

```

```

sd=sd+1;
fclose(fileID);
end

if strcmp(filename, 'peak.csv')
s=strcat(Root_directory, '\\', sub_directories(sub_d
ir_index).name, '\\', csvs(j,1).name);
fileID = fopen(s);
p= textscan(fileID, '%*f %*f %*f %s %f
%f', 'Delimiter', ';', 'TreatAsEmpty', {'NA', 'na'}, 'C
ommentStyle', '#');
xpeak{pk}=p;
pk=pk+1;
fclose(fileID);
end
end
end
end
%%% Save data to file
save('Full_WD_English.mat', 'sub_name', 'xsample', 'xsdnn', 'xp
eak');

```

### 3. Average Mean and standard Deviation of Five frames

```

clear all
clc
load ('brain_waves_Ar_EN.mat')

Frame_size=5;

for i=1:length(brain_waves_Nurosky)
count=0; % count represents the number of train signals
count_test=0; % represents the number of test signals
var_nero_train=0;
var_nero_test=0;
train_feature=[];
test_feature= [];
% Seperate the train from the test data for a person
var=brain_waves_Nurosky{1,i};% var contains neurosky and WD
(train and test )data for a single person
var_nero=var;% var_nero contains the EEG+ECG signal
(train+test) data with the class value
for j=1:length(var_nero)
if (var_nero(j,17)==0)
count=count+1;
end
end
end
end

```

```

count_test=length(var_nero)-count;
var_nero_train=var_nero(1:count,:);
var_nero_test=var_nero(count+1:length(var_nero),:);
%% mean of train data
c=1;
for k=1:Frame_size:count- mod(count,Frame_size)
    train_frame=var_nero_train(k:k+(Frame_size-1),:);
    train_feature(c,:)=mean(train_frame);
    c=c+1;
end
if(mod(count,Frame_size)~=0)
s1=floor(count/Frame_size)*Frame_size+1;
    if(count-s1~=0)
        train_last=mean(var_nero_train(s1:count,:));
        train_feature(c,:)=train_last;
    end
end
train_feature(:,17)=0;
%% mean of test data
c2=1;
for k=1:Frame_size:count_test-
mod(count_test,Frame_size)
    test_frame=var_nero_test(k:k+(Frame_size-1),:);
    test_feature(c2,:)=mean(test_frame);
    c2=c2+1;
end
if(mod(count_test,Frame_size)~=0)
s2=floor(count_test/Frame_size)*Frame_size+1;
    if(count_test-s2 ~=0)
        test_last=mean(var_nero_test(s2:count_test,:));
        test_feature(c2,:)=test_last;
    end
end
test_feature(:,17)=1;

%% merge train and test data then generate the arff file

features=[train_feature;test_feature];
sname=strcat(all_fname{1,i}, '_5F_Brainwaves_mean');
generatefile2(features, sname, 'arff',1);
end

```

## JAVA with Weka API

### 1. Evaluation with 10 folds cross validation

```
package weka.api;
//import classes
import weka.core.Instances;
import java.util.Random;
import weka.core.converters.ConverterUtils.DataSource;
import weka.classifiers.trees.J48;
import weka.classifiers.functions.MultilayerPerceptron;
import weka.classifiers.functions.SMO;
import weka.classifiers.Evaluation;
import weka.classifiers.functions.supportVector.Puk;
import java.io.File;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.PrintWriter;
import java.io.File;

public class Evaluation_CrossValidation2_CSV {

    public static void main(String args[]) throws
Exception{
        // set the path for the data set
        String path="path for the dataset";
        // Read the list of files from the folder
        File folder = new File(path);
        File[] listOfFiles = folder.listFiles();
        int no_of_files=0;

        for (int i = 0; i < listOfFiles.length; i++)
        {
            if (listOfFiles[i].isFile())
            {
                no_of_files=no_of_files+1;
                System.out.println("File " +
listOfFiles[i].getName());
            }
        }
        //load data sets
        DataSource source = new
DataSource(path+
listOfFiles[i].getName());
        Instances dataset =
source.getDataSet(); //set class index to the last
attribute
        dataset.setClassIndex(dataset.numAttributes()-1);
        //create and build the classifier!
```

```

        Evaluation eval = new Evaluation(dataset);
        Random rand = new Random(1);
        int folds = 10;

// the evaluation using Neural Network Classifier
    MultilayerPerceptron nn = new MultilayerPerceptron();
    nn.buildClassifier(dataset);
    eval.crossValidateModel(nn, dataset, folds, rand);

// Evaluation using Random Forest Classifier
    RandomForest RF = new RandomForest();
    RF.setMaxDepth(0);
    RF.setNumExecutionSlots(1);
    RF.setNumDecimalPlaces(2);
    RF.setNumFeatures(0);
    RF.buildClassifier(dataset);
    eval.evaluateModel(tree, testDataset);
    eval.crossValidateModel(RF, dataset, folds, rand);

// the evaluation using SVM Classifier
    SMO svm = new SMO();
    Puk rbf = new Puk();
    svm.setKernel(rbf);
    svm.buildClassifier(dataset);
    eval.crossValidateModel(svm, dataset, folds, rand);

// Print the evaluation results in a file
try{
PrintWriter writer = new PrintWriter(new
FileOutputStream(new File("destination path with file
name"), true));
writer.append(listOfFiles[i].getName()+" ");
writer.append("Correct % = "+eval.pctCorrect()+" ");
writer.append("Correct_NO = "+eval.correct()+" ");
writer.append("Incorrect % = "+eval.pctIncorrect()+" ");
writer.append("Incorrect NO = "+eval.incorrect()+" ");
writer.append("AUC = "+eval.areaUnderROC(1)+" ");
writer.append("kappa = "+eval.kappa()+" ");
writer.append("MAE = "+eval.meanAbsoluteError()+" ");
writer.append("RMSE = "+eval.rootMeanSquaredError()+" ");
writer.append("RAE = "+eval.relativeAbsoluteError()+" ");
writer.append("RRSE =
 "+eval.rootRelativeSquaredError()+" ");
writer.append("Precision = "+eval.precision(1)+" ");
writer.append("Recall = "+eval.recall(1)+" ");
writer.append("fMeasure = "+eval.fMeasure(1)+" ");

```

```

writer.append("Error Rate =,"+eval.errorRate()+"");
writer.append("No of instances
=", "+eval.numInstances()+"");
writer.append("FN =,"+eval.numFalseNegatives(1)+"");
writer.append("FP =,"+eval.numFalsePositives(1)+"");
writer.append("TN =,"+eval.numTrueNegatives(1)+"");
writer.append("TP =,"+eval.numTruePositives(1)+"","\n");
//the confusion matrix
//writer.append(eval.toMatrixString("=== Overall Confusion
Matrix ===\n"));
writer.close();
    }
    catch (IOException e) {
        // do something
    }
}
}
}

```

## 2. Attribute Selection

```

package weka.api;
import weka.attributeSelection.*;
import weka.core.*;
import weka.core.converters.ConverterUtils.*;
import weka.classifiers.*;
import weka.classifiers.meta.*;
import weka.classifiers.trees.*;
import weka.filters.*;
import weka.attributeSelection WrapperSubsetEval;
import java.io.File;
import java.io.FileOutputStream;
import java.io.IOException;
import java.io.PrintWriter;
import java.util.*;
import weka.classifiers.functions.MultilayerPerceptron;
import weka.classifiers.functions.SMO;
import weka.classifiers.functions.supportVector.Puk;

/**
 * performs attribute selection using CfsSubsetEval and
 * GreedyStepwise
 * (backwards) and trains J48 with that. Needs 3.5.5 or
 * higher to compile.
 */

```

```

* @author FracPete (fracpete at waikato dot ac dot nz)
*/
public class AttributeSelectionTest2 {

/**uses the low level approach */
    protected static void useLowLevel(Instances data,
        String name) throws Exception {
        System.out.println("\n3. Low-level");
        AttributeSelection attsel = new AttributeSelection();
        WrapperSubsetEval eval = new WrapperSubsetEval();
        //J48 base = new J48();
        SMO base = new SMO();
        Puk rbf = new Puk();
        base.setKernel(rbf);
        //MultilayerPerceptron base = new
MultilayerPerceptron();
        eval.setClassifier(base);
        GreedyStepwise search = new GreedyStepwise();
        search.setSearchBackwards(false);
        search.setThreshold(-1.7976931348623157E308);
        attsel.setEvaluator(eval);
        attsel.setSearch(search);
        attsel.SelectAttributes(data);
        int[] indices = attsel.selectedAttributes();
        for (int i = 0; i < indices.length; i++) {
            int w=indices[i];
            w=w+1;
            indices[i]=w;
        }
        System.out.println("selected attribute indices
(starting with 0):\n" + Utils.arrayToString(indices));
        try{
            PrintWriter writer = new PrintWriter(new
FileOutputStream(new File("D:\\Exprements Data
final\\Results\\SVM\\Final_extracted_EEG_Selected_features_
SVM.txt"), true));
            writer.append(name+", ");
            writer.append("features,"+
Utils.arrayToString(indices)+"\n");
            writer.close();
        }
        catch (IOException e) {
            // do something
        }
    }
}

```

```

public static void main(String[] args) throws Exception {

    // Load Dataset

    String path="Path to the dataset\\";
    // Read the list of files from the folder
    File folder = new File(path);
    File[] listOfFiles = folder.listFiles();
    int no_of_files=0;

    for (int i = 0; i < listOfFiles.length; i++) {
        if (listOfFiles[i].isFile()) {
            no_of_files=no_of_files+1;
            System.out.println("File " +
                listOfFiles[i].getName());
            String name=listOfFiles[i].getName();

            //Read Each File from the dataset path
            DataSource source = new DataSource(path+
                listOfFiles[i].getName());
            Instances dataset = source.getDataSet();
            //set class index to the last attribute
            dataset.setClassIndex(dataset.numAttributes()-1);

            useLowLevel(dataset,name);
        }
    }
}
}

```



## Appendix B

Participant	Accuracy	FPR%	FNR%	Participant	Accuracy	FPR%	FNR%
1	95.22	0.00	12.21	43	89.13	8.61	14.83
2	90.36	9.17	10.34	44	89.20	0.00	29.72
3	90.48	8.48	11.30	45	95.06	0.00	13.04
4	97.02	2.32	3.81	46	87.72	8.28	18.76
5	76.81	0.12	64.42	47	95.62	1.91	8.43
6	87.37	8.75	19.33	48	91.79	0.00	22.17
7	87.96	0.00	19.69	49	90.70	0.00	27.51
8	94.50	0.12	15.35	50	88.95	0.00	30.60
9	90.51	10.55	7.64	51	94.35	0.00	20.30
10	92.24	8.15	7.10	52	92.68	7.38	7.23
11	93.87	0.00	17.28	53	83.07	0.00	55.96
12	88.32	0.00	33.19	54	74.14	0.00	72.90
13	85.63	0.00	40.60	55	99.91	0.00	0.32
14	91.89	8.19	8.00	56	83.95	5.37	34.51
15	80.67	12.82	26.26	57	86.72	0.00	34.10
16	93.28	6.52	7.02	58	96.96	0.00	9.29
17	92.15	11.72	1.56	59	87.77	0.00	33.40
18	88.75	16.27	5.14	60	97.09	0.00	7.80
19	92.08	0.00	20.86	61	86.70	12.23	14.73
20	74.05	32.66	16.45	62	77.00	38.13	1.04
21	90.89	4.51	19.20	63	94.29	7.87	1.98
22	88.45	12.84	9.73	64	92.57	8.68	5.46
23	95.76	0.00	11.76	65	99.57	0.12	1.35
24	92.26	0.00	21.00	66	93.41	5.21	8.54
25	82.12	0.00	49.58	67	92.71	0.23	21.15
26	80.44	0.00	53.66	68	93.35	0.00	28.27
27	91.04	9.36	8.41	69	84.97	0.00	42.60
28	90.77	0.00	25.48	70	83.48	3.43	35.20
29	89.96	0.36	28.98	71	90.03	0.00	27.93
30	96.22	0.18	9.82	72	92.81	8.75	4.69
31	87.56	0.00	34.39	73	90.46	10.22	8.53
32	86.55	0.00	37.31	74	79.24	0.12	59.09
33	88.07	0.00	34.00	75	89.34	10.65	10.69
34	86.66	0.00	33.78	76	99.01	0.00	3.03
35	89.97	10.51	9.23	77	92.82	5.96	9.29
36	98.50	0.12	4.02	78	90.06	11.25	7.48
37	98.91	0.12	2.72	79	95.04	1.25	9.63
38	82.03	15.57	21.84	80	93.88	0.00	18.27
39	87.06	20.74	3.00	81	82.64	0.00	50.23
40	98.66	0.00	5.38	82	93.06	0.00	19.87
41	97.81	0.00	5.87	83	96.94	0.00	12.59
42	80.02	26.78	3.06	84	88.31	0.00	32.91



## Vitae

Name	Azzat Ahmed Ali AL-Sadi
Nationality	Yemen
Date of Birth	11/3/1979
Email	azzat.sadi@gmail.com
Address	Mukalla - Hadhramout - Yemen - zip code 1914

### Academic Background

- PhD in Computer Science and Engineering (KFUPM, Saudi Arabia) 2018.
  - MS in Computer Networks (KFUPM, Saudi Arabia) 2012.
  - BSc in Information Engineering (Baghdad University, Iraq) 2005.
1. Mohammed H. Sqalli, Raed AlShaikh, and Azzat Ahmed, "A Distributed Honeynet at KFUPM: A Case Study," The 13th International Symposium on Recent Advances in Intrusion Detection (RAID'2010), Ottawa, Ontario, Canada, September 15-17, 2010.
  2. Mohammed H. Sqalli, Raed AlShaikh, and Azzat Ahmed, "Towards Simulating a Virtual Distributed Honeynet at KFUPM: A Case Study," The IEEE UKSim 4th European Modelling Symposium on Mathematical Modelling and Computer Simulation (EMS), Pisa, Italy, November 17-19, 2010.
  3. El-Alfy, E.-S. M. and Al-Sadi, A. "A More Effective Steganographic Approach for Color Images by Combining Simple Methods," The 7th International Computing Conference in Arabic, ICCA 2011, Riyadh, Saudi Arabia, May 2011. (in Arabic)
  4. El-Alfy, E.-S. M. and Al-Sadi, A. "A Comparative Study of PVD-Based Schemes for Data Hiding in Digital Images," The 9th ACS/IEEE International Conference on

Computer Systems and Applications, (AICCSA 2011), Sharm El-Sheikh, Egypt, June 27-30, 2011.

5. Al-Sadi, A. A. and El-Alfy, E.-S. M., "An Adaptive Steganographic Method for Color Images Based on LSB Substitution and Pixel Value Differencing," The International Conference on Advances in Computing and Communications (AC 2011), Kochi Kerala, India, Jul 22-24, 2011.
6. El-Alfy, E.-S. M. and Al-Sadi, A. "Pixel-Value Differencing Steganography: Attacks and Improvements," in Proceedings of the First Taibah University International Conference on Computing and Information Technology, (ICCIT2012), Al-Madinah Al-Munawwarah, Saudi Arabia, March 12-14, 2012.
7. El-Alfy, E.-S. M. and Al-Sadi, A. "Pixel Improved Pixel Value Differencing Steganography Using Logistic Chaotic Maps," in Proceedings of the 8th International Conference on Innovations in Information Technology, (IIT2012), Al Ain, UAE, March 18-20, 2012.
8. El-Alfy, E.-S. M. and Al-Sadi, A. "High-Capacity Image Steganography Based on Overlapped Pixel Differences and Modulus Function," in Proceedings of the Fourth International Conference on Networked Digital Technologies, (NDT2012), Dubai, UAE, April 24-26, 2012.
9. Arshad, Shoieb, Azzat Al-Sadi, and Abdulaziz Barnawi. "Z-MAC: Performance Evaluation and Enhancements." *Procedia Computer Science*, The 4th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2013) and the 3rd International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH), 21 (2013): 485–90.
10. Azzat Al-Sadi, Manaf Bin Yahya, Ahmad Almulhem, "Identification of image fragments for file carving", World Congress on Internet Security (WorldCIS-2013), London, UK, December 9-12, 2013.