NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

ARTIFICIAL INTELLIGENT BASED HYBRID MODELS FOR BODY FAT PERCENTAGE PREDICTION

PhD THESIS

Solaf ALI HUSSAIN

Nicosia
January, 2022

SOLAF ALI HUSSAIN

ARTIFICIAL INTELLIGENT
BASED HYBRID MODELS FOR
BBBODY FAT PREDICTION

PhD THESIS

2022

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

ARTIFICIAL INTELLIGENT BASED HYBRID MODELS FOR BODY FAT PERCENTAGE PREDICTION

PhD THESIS

Solaf ALI HUSSAIN

Supervisor

Prof. Dr. Nadire CAVUS

Nicosia

January, 2022

Abstract

Artificial Intelligent Based Hybrid Models for Body fat Percentage Prediction

Hussain, Solaf Ali

Prof. Dr. Nadire Cavus

PhD, Department of Computer Information Systems

January, 2022, 98 pages

Obesity is a global health problem that affect human health and lifestyle. Obesity determine actual fat mass in clinical studies is described as body fat percentage (BFP). Estimation of BFP, frequently connected with expensive procedures that require specialized equipment. Machine learning (ML) studies applied various models for BFP prediction from different types of data such as anthropometric measurements. The prediction model of BFP may consider large number of features depending on the way the dataset collected and the main objective of the study. Although, single stage ML model can work efficiently with all types of datasets, they could never identify the significant features if they used as standalone model in large datasets with multiple features. In the other hand, studies with hybrid models of BFP prediction registered better performance, especially when they employed a feature selection model to criticize important features to eliminate feature dimensionality. Hybrid models in the employed small size dataset b gender. Ased male gender and specific anthropometry. However, the need for more hybrid ML studies in BFP prediction with intelligent feature selection models and large datasets with gender variety is crucial to full gaps in previous hybrid models. This study aims to propose two hybrid models combined with the base of the emotional artificial neural network (EANN) for BFP prediction. The first model proposed support vector regression (SVR) as sensitive feature selection model, followed by the EANN training algorithm. The second model proposed a hybrid genetic algorithm emotional neural network (GAENN) for the estimation of BFP with a novel approach for feature selection. A special and relatively large dataset released for this study included seven anthropometric parameters. This study is the first to apply EANN for BFP prediction. Results showed that both models (SVR-EANN and GAENN) can achieve higher performance accuracy for BFP prediction in compare with benchmark ML models employed in the same dataset. The study validates a cost-effective and accurate BFP prediction model that could be applicable in obesity control programs and data mining programs to analyze obesity levels in societies.

Key Words: Body fat percentage, prediction, artificial intelligence, machine learning, hybrid models

Özet

Vücut Yağ Yüzdesi Tahmini için Yapay Zeka Tabanlı Hibrit Modeller

Hussain, Solaf Ali
Prof. Dr. Nadire Cavus
PhD, Bilgisayar Enformatik Anabilim Dalı
Ocak, 2022, 98 sayfa

Obezite, insan sağlığını ve yaşam biçimini etkileyen küresel bir sağlık sorunudur. Obezite, klinik çalışmalarda gerçek yağ kütlesini belirleyen vücut yağ yüzdesi (VYK) olarak tanımlanır. VYK tahmini, sıklıkla özel ekipman gerektiren pahalı prosedürlerle bağlantılıdır. Makine öğrenimi (MÖ) çalışmaları, antropometrik ölçümler gibi farklı veri türlerinden BFP tahmini için çeşitli modeller uygular. VYK'nin tahmin modeli, veri setinin toplanma şekline ve çalışmanın temel amacına bağlı olarak çok sayıda özelliği dikkate alabilir. Tek aşamalı makine öğrenimi modeli her tür veri kümesiyle verimli bir şekilde çalışabilmesine rağmen, birden çok özelliğe sahip büyük veri kümelerinde bağımsız model olarak kullanıldıklarında önemli özellikleri hiçbir zaman tanımlayamazlar. Öte yandan, VYK tahmininin hibrit modelleri ile yapılan çalışmalar, özellikle özellik boyutluluğunu ortadan kaldırmak için önemli özellikleri eleştirmek için bir özellik seçim modeli kullandıklarında daha iyi performans kaydedilir. Kullanılan küçük boyutlu veri kümesindeki hibrit modeller b cinsiyet. Ased erkek cinsiyeti ve spesifik antropometri. Bununla birlikte, akıllı özellik seçim modelleri ve cinsiyet çeşitliliğine sahip büyük veri kümeleri ile VYK tahmininde daha fazla hibrit MÖ çalışmasına duyulan ihtiyaç, önceki hibrit modellerdeki boşlukların doldurulması için çok önemlidir.Bu çalışma, VYK tahmini için duygusal yapay sinir ağının (DYSA) temeli ile birleştirilmiş iki hibrit model önermeyi amaçlamaktadır. İlk model, hassas öznitelik seçim modeli olarak destek vektörü regresyonu (DVR)'yi önerdi ve ardından DYSA eğitim algoritması geldi. İkinci model, özellik seçimi için yeni bir yaklaşımla VYK'nin tahmini için bir melez genetik algoritma duygusal sinir ağı (GADSA) önerdi. Bu çalışma için yayınlanan özel ve nispeten büyük bir veri seti, yedi antropometrik parametreyi içeriyordu. Bu çalışma, VYK tahmini için DYSA'ı uygulayan ilk çalışmadır. Sonuçlar, her iki modelin de (DVR-DYSA ve GADESNA), aynı veri setinde kullanılan kıyaslama MÖ modelleriyle karşılaştırıldığında VYK tahmini için daha yüksek performans doğruluğu sağlayabildiğini gösterdi. Çalışma, toplumlardaki obezite düzeylerini analiz etmek için obezite kontrol programlarında ve veri madenciliği programlarında uygulanabilecek uygun maliyetli ve doğru bir VYK tahmin modelini doğrulamaktadır.

***Anahtar Kelimeler*:** Vücut yağ yüzdesi, tahmin, yapay zeka, makine öğrenmesi, hibrit modelleme

SOLAF ALI HUSSAIN

ARTIFICIAL INTELLIGENT
BASED HYBRID MODELS FOR
BBBODY FAT PREDICTION

PhD THESIS

2022

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

ARTIFICIAL INTELLIGENT BASED HYBRID MODELS FOR BODY FAT PERCENTAGE PREDICTION

PhD THESIS

Solaf ALI HUSSAIN

Nicosia

January, 2022

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF COMPUTER INFORMATION SYSTEMS

ARTIFICIAL INTELLIGENT BASED HYBRID MODELS FOR BODY FAT PERCENTAGE PREDICTION

PhD THESIS

Solaf ALI HUSSAIN

Supervisor

Prof. Dr. Nadire CAVUS

Nicosia

January, 2022

Approval

We certify that we have read the thesis submitted by Solaf Ali Hussain titled "**Artificial Intelligent Based Hybrid Models for Body Fat Percentage Prediction**" and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of PhD in Computer Information Systems.

Examining Committee	**Name-Surname**	**Signature**
Head of the Committee:	Prof. Dr. Sadik Ülker	…..……………
Committee Member*:	Prof. Dr. Mehmet Toycan	…..……………
Committee Member*:	Assoc. Prof. Dr. Boran Şekeroğlu	…………………
Committee Member*:	Assist. Prof. Dr. Seren Başaran	…………………
Supervisor:	Prof. Dr. Nadire Çavuş	..……………….

Approved by the Head of the Department

23/02/2022

Prof. Dr. Nadire Çavuş
Head of Department

Approved by the Institute of Graduate Studies

…../…../2022

Prof. Dr. Kemal Hüsnü Can Başer
Head of the Institute

Declaration

I hereby declare that all information, documents, analysis, and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Solaf Ali Hussain

12/12/2021

Acknowledgments

Firstly, I would like to express my thanks to my supervisor Prof. Dr. Nadire Cavus for her precious support and valuable information during the study period.

I would like to forward my great thanks and appreciate to Assoc. Prof. Dr. Boran Sekeroglu for his valuable assistant.

My appreciate and thanks to Thesis defence jury members for their valuable comments.

I want to appreciate my family particularly my parents, for their immerse support throughout the period of this study. All brothers and sisters for their spirit guidance and support in difficult times.

My friends for their important support. sincere heart of gratitude goes to the entire instructors in the department and my supervisor, for the valuable impact in my study.

Solaf Ali Hussain

Abstract

Artificial Intelligent Based Hybrid Models for Body fat Percentage Prediction

Hussain, Solaf Ali
Prof. Dr. Nadire Cavus
PhD, Department of Computer Information Systems
January, 2022, 98 pages

Obesity is a global health problem that affect human health and lifestyle. Obesity determine actual fat mass in clinical studies is described as body fat percentage (BFP). Estimation of BFP, frequently connected with expensive procedures that require specialized equipment. Machine learning (ML) studies applied various models for BFP prediction from different types of data such as anthropometric measurements. The prediction model of BFP may consider large number of features depending on the way the dataset collected and the main objective of the study. Although, single stage ML model can work efficiently with all types of datasets, they could never identify the significant features if they used as standalone model in large datasets with multiple features. In the other hand, studies with hybrid models of BFP prediction registered better performance, especially when they employed a feature selection model to criticize important features to eliminate feature dimensionality. Hybrid models in the employed small size dataset b gender. Ased male gender and specific anthropometry. However, the need for more hybrid ML studies in BFP prediction with intelligent feature selection models and large datasets with gender variety is crucial to full gaps in previous hybrid models. This study aims to propose two hybrid models combined with the base of the emotional artificial neural network (EANN) for BFP prediction. The first model proposed support vector regression (SVR) as sensitive feature selection model, followed by the EANN training algorithm. The second model proposed a hybrid genetic algorithm emotional neural network (GAENN) for the estimation of BFP with a novel approach for feature selection. A special and relatively large dataset released for this study included seven anthropometric parameters. This study is the first to apply EANN for BFP prediction. Results showed that both models (SVR-EANN and GAENN) can achieve higher performance accuracy for BFP prediction in compare with benchmark ML models employed in the same dataset. The study validates a cost-effective and accurate BFP prediction model that could be applicable in obesity control programs and data mining programs to analyze obesity levels in societies.

Key Words: Body fat percentage, prediction, artificial intelligence, machine learning, hybrid models

Özet

Vücut Yağ Yüzdesi Tahmini için Yapay Zeka Tabanlı Hibrit Modeller

Hussain, Solaf Ali
Prof. Dr. Nadire Cavus
PhD, Bilgisayar Enformatik Anabilim Dalı
Ocak, 2022, 98 sayfa

Obezite, insan sağlığını ve yaşam biçimini etkileyen küresel bir sağlık sorunudur. Obezite, klinik çalışmalarda gerçek yağ kütlesini belirleyen vücut yağ yüzdesi (VYK) olarak tanımlanır. VYK tahmini, sıklıkla özel ekipman gerektiren pahalı prosedürlerle bağlantılıdır. Makine öğrenimi (MÖ) çalışmaları, antropometrik ölçümler gibi farklı veri türlerinden BFP tahmini için çeşitli modeller uygular. VYK'nin tahmin modeli, veri setinin toplanma şekline ve çalışmanın temel amacına bağlı olarak çok sayıda özelliği dikkate alabilir. Tek aşamalı makine öğrenimi modeli her tür veri kümesiyle verimli bir şekilde çalışabilmesine rağmen, birden çok özelliğe sahip büyük veri kümelerinde bağımsız model olarak kullanıldıklarında önemli özellikleri hiçbir zaman tanımlayamazlar. Öte yandan, VYK tahmininin hibrit modelleri ile yapılan çalışmalar, özellikle özellik boyutluluğunu ortadan kaldırmak için önemli özellikleri eleştirmek için bir özellik seçim modeli kullandıklarında daha iyi performans kaydedilir. Kullanılan küçük boyutlu veri kümesindeki hibrit modeller b cinsiyet. Ased erkek cinsiyeti ve spesifik antropometri. Bununla birlikte, akıllı özellik seçim modelleri ve cinsiyet çeşitliliğine sahip büyük veri kümeleri ile VYK tahmininde daha fazla hibrit MÖ çalışmasına duyulan ihtiyaç, önceki hibrit modellerdeki boşlukların doldurulması için çok önemlidir.Bu çalışma, VYK tahmini için duygusal yapay sinir ağının (DYSA) temeli ile birleştirilmiş iki hibrit model önermeyi amaçlamaktadır. İlk model, hassas öznitelik seçim modeli olarak destek vektörü regresyonu (DVR)'yi önerdi ve ardından DYSA eğitim algoritması geldi. İkinci model, özellik seçimi için yeni bir yaklaşımla VYK'nin tahmini için bir melez genetik algoritma duygusal sinir ağı (GADSA) önerdi. Bu çalışma için yayınlanan özel ve nispeten büyük bir veri seti, yedi antropometrik parametreyi içeriyordu. Bu çalışma, VYK tahmini için DYSA'ı uygulayan ilk çalışmadır. Sonuçlar, her iki modelin de (DVR-DYSA ve GADESNA), aynı veri setinde kullanılan kıyaslama MÖ

modelleriyle karşılaştırıldığında VYK tahmini için daha yüksek performans doğruluğu sağlayabildiğini gösterdi. Çalışma, toplumlardaki obezite düzeylerini analiz etmek için obezite kontrol programlarında ve veri madenciliği programlarında uygulanabilecek uygun maliyetli ve doğru bir VYK tahmin modelini doğrulamaktadır.

***Anahtar Kelimeler*: Vücut yağ yüzdesi, tahmin, yapay zeka, makine öğrenmesi, hibrit modelleme**

Table of Contents

Approval.. i
Declaration.. ii
Acknowledgments…………………………………………………..…………. iii
Abstract.. iv
Özet.. vi
Table of Contents.. viii
List of Tables.. x
List of Figures.. xi
List of abreviations.. xii

CHAPTER I

Introduction... 1
Body Fat Percentage... 1
The Problem Statements... 2
Objectives of Study... 4
Significance of Study.. 4
Study Limitations.. 5
Overview of the Thesis... 5

CHAPTER II

Literature Review... 6
Systematic Literature Review... 6
Literature Review Process.. 6
Planning... 6
Data Collecting.. 6
Results and Discussion of the SLR.. 11
Summary of SLR Results... 22

CHAPTER III

Methodology.. 24
Data Collection and Analysis... 24
Dataset... 24

Design of Research Hybrid Model 1 (SVR-EANN).. 26
Input Variable Selection... 28
Design of Research Hybrid Model 2 (GA-EANN).. 29
Dominant Input Variables Identification.. 30
Artificial Neural Network.. 31
Emotional Artificial Neural Networks.. 33
Regression Using Support Vectors.. 36
Hybrid Genetic Algorithm Emotional Artificial Neural Networks.................. 37
Data Normalization and Performance Evaluation Criteria.............................. 39
The Validation of the Model.. 40

CHAPTER IV

Results.. 42
Results of Model1 (SVR-EANN)... 42
Results of Model2 (GAEANN)... 48

CHAPTER V

Discussion.. 56
Discussion of model1:SVR-EANN.. 56
Discussion of Model2:GAEANN.. 59

CHAPTER VI

Conclusion and Recommendations... 64
Conclusion.. 64
Recommendations for Future Works.. 65

REFERENCES.. 67
APPENDICES... 75
APPENDIX A: Ethical Committee Approval letter... 75
APPENDIX B: Permissions Regarding the Use of Dataset.............................. 76
APPENDIX C: Turnitin Similarity Report.. 77
APPENDIX D: Author CV.. 85

List of Tables

Page

Table 2.1 Inclusion-Exclusion Criteria... 7

Table 2.2 Description of Included Articles……………………….......…….. 9

Table 2.3 BODY FAT Dataset Features……………………...………....... 12

Table 3.1 The Dataset's Descriptive Statistics……………………….….... 25

Table 3.2 Correlation between Variables……………………………….…… 26

Table 4.1 Sensitivity Analysis Output……………………………....…….. 43

Table 4.2 Modelling Results for Model1 ………………………………..... 44

Table 4.3 BFP Prediction Results with Bench Marks Models for SVR-EANN.. 47

Table 4.4 Results for the Modeling and Performance Analysis …................... 49

Table 4.5 Results for the Proposed Models and Benchmark Algorithms for GAENN.. 53

List of Figures

Page

Figure 2.1 Research Methodology to Collect Data………………………… 8

Figure 3.1 The Correlation matrix between the Variables………………….. 26

Figure 3.2 Schematic of the SVR-EANN Hybrid Model…………………... 27

Figure 3.3 Schematic Representation of the SVR Omitted Sensitivity Analysis.. 29

Figure 3.4 Schematic of the GAEANN Hybrid Model…………………....... 31

Figure 3.5 Structure of the Three-layer FFNN with Single neuron in Output Layer 32

Figure 3.6 Neuron Architecture in FF-BPNN &EANN …………………… 36

Figure 3.7 Conceptual Structure of SVR Model ……………….………….... 37

Figure 4.1 Error Plots for the Models (FFNN, EANN, SVR-EANN)…..…... 45

Figure 4.2 Scatter Plots for Predicted BFP in Training ……………….……. 46

Figure 4.3 Scatter Plots for Predicted BFP in Verification ………..……….. 46

Figure 4.4 Error plots for the models (ANN, EANN, GAEANN)………… 50

Figure 4.5 Scatter Plots for GAEANN Model2……………………….….... 51

Figure 5.1 Visualization of the Results Obtained for Model1……………… 56

Figure 5.2 Visualization of the Results for M1……………………………... 60

Figure 5.3 Visualization of the Results for M2……………………………… 61

Figure 5.4 Visualization of the Results for M3……………………………… 62

List of Abbreviations

AI:	Artificial Intelligence
ANFIS:	Adaptive Neuro Fuzzy Inference System
ANN:	Artificial Neural Network
BFP:	Body Fat Percentage
BIA:	Bioelectrical Impedance Analysis
BMI:	Body Mass Index
CT:	Computed Tomography
DEXA:	Dual-Energy X-Ray Absorptiometry
DNN:	Deep Neural Network
DT:	Decision Tree
EANN:	Emotional Artificial Neural Network
FFNN:	Feed Forward Neural Network
GA:	Genetic Algorithm
GRADBOOST:	Gradient Boosting
MARS:	Multi Variate Adaptive Regression Spline
ML:	Machine learning
MLP:	Multiple Layer Perceptron
MSE:	Mean Square Error
MR:	Multiple Regression
MRI:	Magnetic Resonance Imaging
NHANES:	National Health and Nutrition Examination Survey
NIR:	Near-Infrared Interactance
NN:	Neural Network
PCA	Principal Component Analysis
PPG:	Photoplethysmography Signal
RBF:	Radial Basis Function
RE:	Relative Error
RMSE:	Root Mean Square Error
RF:	Random Forest
SLR:	Systematic Literature Review
PC:	Pearson Correlation
SVM:	Support Vector Machine
SVR:	Support Vector Regression
TSVR:	Twin Support Vector Regression
UWW:	Underwater Weighing
WHR:	Waist-Hip-Ratio
XGBOOST:	External Gradient Boosting

CHAPTER I
Introduction

This chapter presents a brief introduction to Body fat percentage estimation and proposes the problem statement, the main objectives of the study, the study significance, hypothesis, the limitations, and an overview of other thesis sections.

Body Fat Percentage Estimation

Obesity is a serious global health problem (Kupusinac et al., 2014). The human body is made up of three types of tissues: fat, muscle, and lean. Having an excess of fat tissue in the body is what causes obesity (Uçar et al., 2021a). The actual fat tissue represented by the body fat percentage BFP. This percentage indicates the measure of human fitness level and basic relative body composition measurement. Obesity which is associated with excess body fat is attributed to several health issues and deadly sickness such as high blood pressure, diabetes and heart diseases (Fan et al., 2017). On the other hand, low body fat below certain threshold is also unwanted, because brain functions required convinced amount of body fat (Maughan, 1993 as cited in Shao, 2014).

The presence of essential body fat is required for the maintenance of healthy life. Due to the requirements of pregnancy and other hormonal actions, women gain a higher percentage of dynamic body fat than men. Men have a 3%-5% necessary fat content, while women have an 8%-12% of essential fat (Freedman et al., 2009). Internal body fat is made up of fat that has accumulated in adipose tissue and serves to protect internal organs in the chest and belly (Freedman et al., 2009).

There are several methods for BFP estimation. Accurate but expensive methods, depend clinical devices named body analysis instruments for instance (“dual-energy X-ray absorptiometry” (DEXA), “bioelectrical impedance analysis” (BIA), “underwater weighing” (UWW), “magnetic resonance imaging ” (MRI),) “near-infrared interactance “(NIR), and computed tomography (CT) (Wang et al., 2010).

Despite the confusion caused by internal fat levels, calculation-based anthropometric measurements of fat mass, such as body mass index (BMI), waist circumference, and waist-to-hip ratio (WHR), and skinfold estimation) are more commonly used (Swainson et al., 2017). Some of these measurements depended on special equations to determine obesity or body fat (e.g., the BMI which is simple anthropometric measure of weight in relation to height that is frequently used to identify humans as overweight or obese). It is calculated by dividing an individual's weight in kilograms by the square on his height in meters (kg/m2) (DeGregory et al., 2018). However, using BMI has significant disadvantages, as it does not adequately represent actual body fat content. WHR which represent the ratio between waist circumference to hip circumference depends in two regions of the body and cannot compute the total body fat precisely.

Another anthropometric measurement model that relies heavily on statistical computations is skinfold estimate, often known as the pinch test, where a pinch of skin is measured by callipers at numerous standardized sites on the body to calculate the thickness of the subcutaneous fat layer and results indicated in previous statistic samples related to general population samples (Sarría et al., 1998).

The significance of prediction abilities of machine learning approaches could make impact revolution in BFP prediction studies to enhance the poor solutions provided by traditional and clinical approaches. These models enrich the BFP estimation problem domain with more precise and cost-effective options. The increased learning abilities with high performance prediction algorithms encouraged researchers to employ the capabilities of artificial intelligence (AI) and machine learning (ML) models to address the difficult non-linear problem for BFP prediction model. For this end, this study tries to produce a hybrid ML solution with simple anthropometric features and new generation of neural networks (NNs).

The Problem Statement

Specialized body analysis devices for BFP estimation are expensive, while anthropometric measurements of BFP are inaccurate that makes accurate and reliable measurement of body fat a challenging task. This led researchers to devote several machine learning studies to predict obesity disease and BFP from simple anthropometric measurements in order to monitor and keep under control. In regard to improve prediction accuracy of some algorithms in BFP prediction, Some studies employed single stage algorithms, for example, neural networks (NNs) and/or support vector machines (SVM)s because of their ability to handle the nonlinearity in the dataset (Ferenci & Kovács, 2018; Kupusinac et al., 2014; Akman et al., 2021; Chiong et al., 2021). Gao et al. (2020) employed the optimization abilities of genetic algorithm in BFP prediction model through a modified adaptive genetic algorithm within novel parameter selection models. However, because of the shortcomings associated with these individual models they often provide unsatisfactory result. For instance, ANN cannot handle data uncertainties as well as it is difficult to identify their structure (Abba et al., 2020; Muhammad et al., 2018) and Even though impressive prediction results were obtained, the uninterpretable relationship among “input-output” data of ANNs prohibited the degree of relationship and importance of appropriate input features in BFP prediction. In order to improve prediction ability and find key features in the BFP prediction model, researchers investigated the development of ML-based hybrid models with the combination of feature selection and prediction algorithms. The combination of comprehensive and primary datasets as well as feature selection methods can improve the prediction model. Multilayer feed-forward neural network (MLFFNN), decision tree (DT), SVM, and linear regression (LR) employed by Shao (2014) and Uçar et al. (2021a) in hybrid ML models to improve feature selection for BFP prediction model. They used the bodyfat dataset generated by (Johnson, 1996) that included 14 parameters of 252 males. outcomes of the studies proved that anthropometric measurement may be successfully utilized to predict BFP, and advance feature selection methodologies will enhance BFP prediction performance. However, both researches applied modelling approaches to “small” datasets with gender based only male subjects, and validation of the proposed models in larger datasets were never tested.

Aforementioned literature registered attempts to combine neural networks and genetic algorithms in hybrid models for solving prediction problems, but they are both optimization and learning techniques, and each has its own advantages and disadvantages. Some Researches employed A hybrid model consisting of ANN and a GA procedure for medical sector and disease prediction, they successfully Combined the relative advantages of GA to update network structure of the NN model used (Karegowda et al., 2011; Mantzaris et al., 2011; Amirov et al., 2014).

Evidence in literature showed the importance and effectiveness of EANN in modelling nonlinear problems. Hence, implementation of Newley emergent hybrid models based EANN to increase the prediction ability for BFP model is contributed in this study. We proposed two hybrid models with relatively large dataset include basic information (gender, age) and some anthropometric parameters (height, weight, BMI, WHR, abdominal or waist circumference). First hybrid model (SVR-EANN) employs support vector regression (SVR) for feature selection and depends

EANN as prediction algorithm. While, the second hybrid model combines GA with EANN for BFP prediction with novel parameter selection.

Objectives of Study

The main objectives of this thesis are as follows:

i Build and estimate the performance of innovative machine learning ML hybrid models based emotional artificial neural network EANN for adult (age ≥ 18) body fat percentage prediction.

ii The exploration of significant parameters for BFP prediction through the employment of different feature selection models varied between intelligent and traditional novel selection models and estimate best BFP prediction accordingly.

iii Obtaining a large dataset with a diverse gender mix for better simulation analysis.

iv Develop and assessment the effectiveness of the two suggested hybrid-ML models in the prediction model of BFP named, support vector regression and emotional artificial neural network SVR-EANN, and genetic algorithm -emotional neural network GA-EANN model.

Significance of the Study

This study focused multiple significant factors expressed as follows:

- Gathering of primary data for BFP prediction within variables of anthropometric measurements for both genders, compromising a greater number of observations than previously involved in machine learning-based-hybrid model studies.
- For the first time, the WHR was used in a hybrid prediction model for BFP.
- The first use of EANN in this work will show a new generation of modelling methodologies for the BFP problem.
- Developing hybrid models depends the EANN for precise BFP prediction.

- Suggesting two variant methodologies for implementing the feature selection paradigm for the modelling algorithms, the first model used SVR as base of left-out sensitivity analysis for estimating influential factors, and the second model used a novel parameter selection approach.
- In this study, SVREANN and GAEANN were proposed as two separate hybrid models for BFP prediction.

Study Limitations

The thesis illustrated significant methodologies but some obstacles related to the data collection procedure for the thesis, resulted in limiting the age of subjects between (18-29) years old. The primary dataset and models applied are validated to the scope of this study.

Overview of the Thesis

The aim of the study included, collection of primary data for BFP prediction within anthropometric variables for both genders, compromising a greater number of observations than previously investigated in machine learning-based-hybrid model studies. For this end, the thesis organized into six chapters. A summary of the chapters contents is provided below:

- Chapter I: Provides an overview of the body fat percentage estimation problem, the problem statement, the aims of the study, the importance of the study, the limitations, and the general overview.
- Chapter II: Discuss relevant studies on AI based BFP prediction models. A Systematic literature review is conducted.
- Chapter III: Include the study methodology, data collecting description, and theoretical underpinning of the modelling methodologies employed.
- Chapter IV: Provides a detailed discussion of the proposed modelling approaches’ design and implementation.
- Chapter V: Provides a conclusion, recommendations, and proposals for future research.
- Chapter VI: Discuss results with previous studies.

CHAPTER II
Literature Review

In this chapter a systematic review of literature will be conducted to evaluate the impact of AI based models in BFP prediction, and find gaps in previous studies to relate them to the study.

Systematic Literature Review

A systematic literature review (SLR) research method was used to understand what other researchers do in the literature and then identify what missing gaps are found in the literature on the chosen thesis topic.

Literature Review Process

A systematic literature review (SLR) that follows Kitchenham and Charters (Kitchenham as cited in Safaei et al., 2021) strategy to make a brief idea about researches made in literature. The SLE methodology followed three major processes: Planning, data collecting, and discussion of SLR outcomes. The SLR methodology followed to implement the 3-process SLR is described in the following sections.

Planning

Identifying SLR Questions: Research questions for the SLR formulated according to basic objectives of the thesis. "Research questions" (RQs) for this SLR are:

- RQ1: What are the types of datasets used in history for BFP prediction models?
- RQ2: What is the rule of Machine learning models applied in the problem of BFP prediction?

Data Collecting

a) Search Strategy: To collect data in this study, a search strategy was used at the primary search level to assure the optimal search strategy to include the most similar studies. Google Scholar search engine was utilized to extract all of the papers that were indicated by the search keywords. Then, to combine relevant studies, search the most efficient databases and libraries. Because of the potential of this search approach to extract data from specific articles and assist scholars in locating the most relevant papers. In order to conduct the search step for this study, the following keywords utilized:

- (“artificial intelligence” OR” machine learning” OR” artificial” OR” intelligent”) AND (“body fat” OR” body composition”) AND (“prediction” OR” estimation”)

A manual search made limited to the range of years (from 2010 to October 2021) in the databases such as Science Direct, Springer Link, IEEE Explorer, Web of Science, Scopus to collect the relevant research articles.

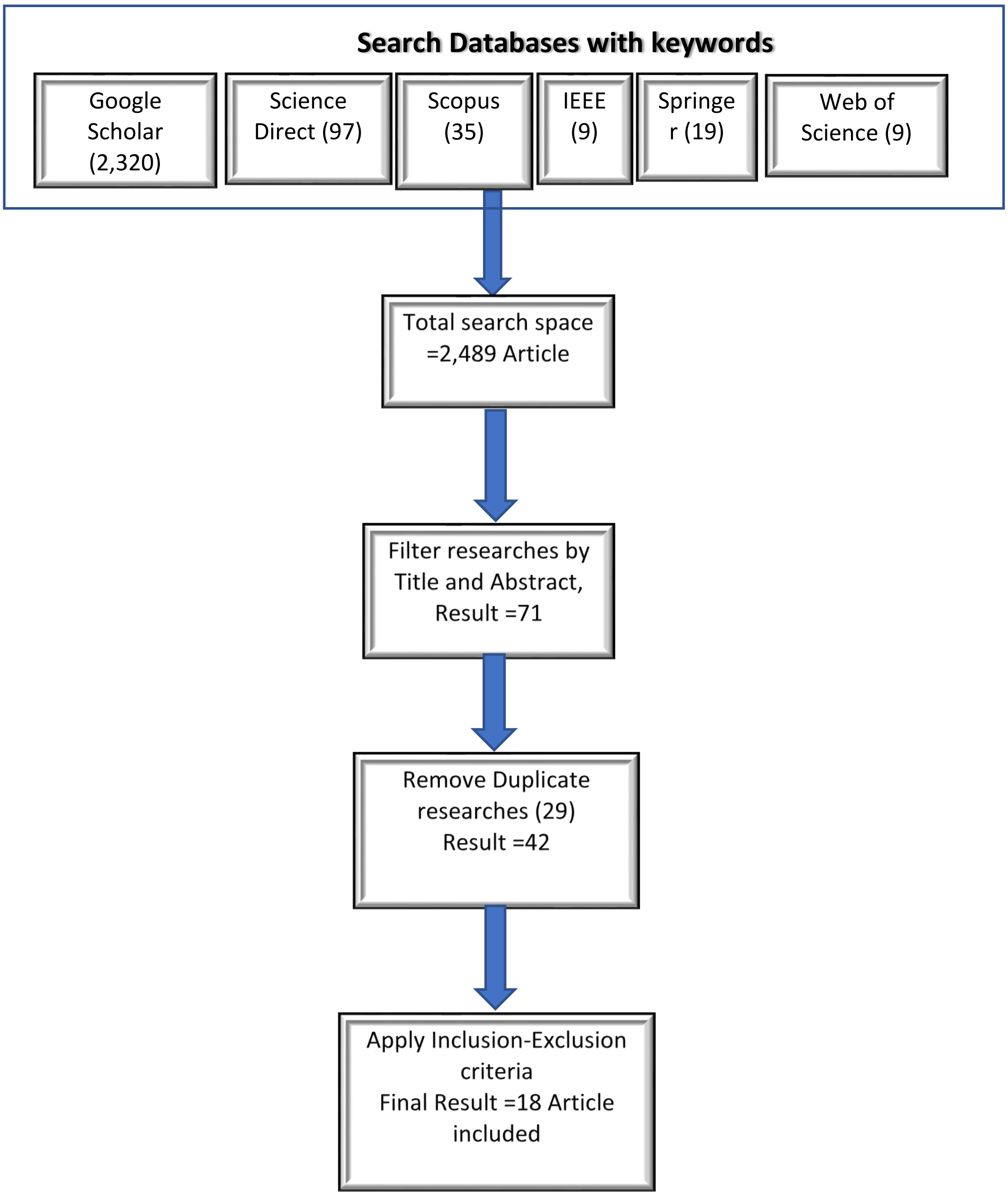
b) The Inclusion / Exclusion Criteria: “Inclusion” and “exclusion” criteria were devised for the verification of relevant publications to AI-based BFP prediction problem. This process started with a study of each article's title and abstract to see if it addressed the current SLR's focus. In a subsequent stage, each article was examined using predefined inclusion and exclusion criteria to determine whether it should be retained in the database. This review covered only BFP forecast papers published in English between 2010 and October 2021. Table 2.1 summarizes the detailed inclusion and exclusion criteria.

Table 2.1
Inclusion-Exclusion Criteria

Criteria(s)	Principle
Inclusions	English language-based studies Studies related to adult human estimation of BFP, age≥18 Prediction models compromised in this Study included regression models of BFP prediction. Full text researches of Relevant researches to defined RQs
Exclusions	Duplicated studies Animal based studies ML classification studies Irrelevant studies to the research questions

c) Data Selection Process: The primary goal of the selection procedure is to locate relevant articles. The overall search space includes 2,489 articles returned by the automatic search based on the above-mentioned keywords. The publications were then filtered by title, abstract, and keywords, yielding 71 studies. After removing 29 duplicate articles, the results were reduced to only 42 articles. For each study, inclusion and exclusion criteria were meticulously applied. Any papers that were inappropriate to the research topic were deleted. Total results, captured 18 articles to be included for discussion in this SLR study. Figure 2.1 depicts the research approach for data collection.

Figure 2.1
Research Methodology of the SLR to Collect Data



d) Extraction and Synthesis of Data: Following the data gathering technique, the papers were grouped for analysis. Data is structured in a table to provide a comprehensive description of the research articles comprised in the SLR. The retrieved data were assigned to the newly formed columns, which included the title, authors, dataset used, and AI-based algorithm(s) or approach(es) of each study. The full data extraction form utilized for 19 obtained research is shown in Table 2.2.

Table 2.2

Description of Included Articles

Authors	Article Title	Data Set Description	Algorithms
Farquad et al., 2010	"Support Vector Regression Based Hybrid Rule Extraction Methods for Forecasting"	BODY FAT	Hybrid models: SVR+CART, SVR+DENFIS , SVR+ANFIS
Balasundaram & Kapil, 2010	"On Lagrangian Support Vector Regression"	BODY FAT	Lagrangian SVR
Xu & Wang, 2012	"A weighted Twin Support Vector Regression"	BODY FAT	Weighted TSVR with different kernel functions
Kupusinac et al., 2014	"Predicting body fat percentage based on gender, age and BMI using artificial neural network"	Dataset of Tanita device for 2755 adult from Serbia (1332 male and 1423 men) aged (18 -88).	ANN
Shao, 2014	"Body Fat Percentage Prediction Using intelligent hybrid Approaches"	BODY FAT	Hybrid model MRANN, MRMARS MRSVR, MRRSMR MARSANN, MARSSVR
Lu et al., 2018	"3D Shape-Based Body Composition Prediction Model Using Machine Learning"	Data set included 50 males with DXA model for BFP	SVM, K-NEAREST neighbor KNN, DT
Ferenci & Kovács, 2018	"Predicting Body Fat Percentage from Anthropometric a& Laboratory Measurements using ANN"	Dataset of (n=862) males, aged>18 years, 39 variables of anthropometric and laboratory types.	FFNN, SVM

Table 2.2 -Continue

Shin et al., 2019	"Dry Electrode-Based Body Fat Estimation System with Anthropometric Data for Use in a Wearable Device"	Dataset with 163 subjects, features collected with sensor technology	DNN
Lu et al., 2020	"3D Shape-Based Body Composition Inference using a Bayesian Network"	Custom Dataset with 100 females with DXA radiation method BFP estimated by siri EQ.	Bayesian Network
Harty et al., 2020	"Novel Body Fat Estimation Using Machine Learning and 3-Dimensional Optical Imaging"	Dataset of 179 participants (103 female,76 male) completed five full-body scans with 3-dimentional optical (3DO) scanner	DT
Gao et al., 2020	"Predicting human body composition using a modified adaptive genetic algorithm with a novel selection operator"	Custom dataset with 220 subjects (124 male and 96 female), based BIA model for BFP estimation,	GA
Naik, 2021	"A novel sensitivity-based method for feature selection"	BODY FAT	DNN
Uçar et al., 2021a	"Estimation of body fat percentage using hybrid machine learning algorithms"	BODY FAT	MLFFNN, SVM, DT MLFFNN+DT ,MLFFNN+S VMMLFFNN +DT+SVM
Chiong et al., 2021	"Using an improved relative error support vector machine for body fat prediction"	BODY FAT &NHANES	IRE-SVM, MLP, SVM, RF,XGBoost, RE-SVM
Fan et al., 2021	"A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat prediction"	BODY FAT & NHANES	Fuzzy rules, SVM
Akman et al., 2021	"Determination of body fat percentage by photoplethysmography signal with gender based artificial intelligence"	dataset of 331 person,PPG signals and BFP from BIOPAC MP36 device	FFNN DT
Wang et al., 2021	"Pixel-wise body composition prediction with a multi-task conditional generative adversarial network"	CT devise data for total of 270 patients	Multi-task DNN
Uçar et al., 2021b	"Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence"	Dataset with 266 participant,138 male,128 women Features based ECG signals	FFNN DT

Note: Table data sorted by year of publication in ascending order.

Results and Discussion of the SLR

a) Description of Datasets used to Predict BFP

Body composition analysis compromised multiple techniques for estimation the general body fitness score represented by the BFP. These techniques varied between accurate and expensive direct techniques, which includes "clinical-based special devices". And indirect models that , include cheap but inaccurate techniques called anthropometry models (Kupusinac et al., 2014).

Body composition studies constitute a strong correlation between BFP and body consistence or named (body density). The Siri equation, has been universally used to shape body density to BFP (Johnson, 1996). According to the research, multiple models have been proposed to estimate body volume of human. The most popular and precise approach used was hydrostatic counting, which computes the body density estimated by under- water weighing UWW grounded on Archimedes-Principle. Practically, this manner is hard to apply in clinical tests (Lu et al., 2018b).

Literature witnessed many studies for BFP estimation. Statistic-based studies tried to find solutions to the model with anthropometric measurements as input variables for BFP estimation by establishing some linear and nonlinear equations. Johnson (1996), estimated multiple regression (MR) models from mapping simple metric dimensions included age, weight, height, and 10 circumference measurements with BFP which determined accurately by UWW technique (siris equation) for 252 males. As described in Table2.3. This dataset is mentioned in many statistical and machine learning studies. In this study, we used the name BODY FAT to mention Johnson's dataset. however, application of the anthropometrical variables to predict BFP with (MR) models or ML algorithms investigated in many studies, but still the relationship between body measurements and BFP is difficult to be identified. As described in (table 2.3), 8 research articles included in this systematic review used the "BODY FAT" dataset with different machine learning approaches. Although, it represents well-structured real dataset for BFP prediction models, but still considered as small size dataset and restricted to male gender.

Table 2.3

Body Fat Dataset Features

Variable	Meaning
D	"Density" determined from UUW
y	BFP=(495/D)-450
X_1	Age
X_2	Body Height(cm)
X_3	Body Weight(kg)
X_4	"Neck"-circumference (cm)
X_5	"Chest"- circumference (cm)
X_6	"Abdomen 2"- circumference (cm)
X_7	"Hip"- circumference (cm)
X_8	"Thigh"- circumference (cm)
X_9	"Knee" -circumference (cm)
X_{10}	"Ankle" -circumference (cm)
X_{11}	"Biceps"- circumference (cm)
X_{12}	"Forearm" -circumference (cm)
X_{13}	"Wrist" -circumference (cm)

Note: BFP represented by (y) variable calculated by Siri equation.

The National Health and Nutrition Examination Survey (NHANES) is a series of studies aiming to measure the health and nutritional status of adults and children in the United States. This is an ongoing initiative containing a large amount of data, including (dietary, demographic, laboratory questionnaire) data. This dataset invested in some studies for BFP prediction.

Recently, machine learning studies for BFP prediction employed datasets with BFP estimated by BIA device ,which described to be less accurate but cheaper than Other direct models (Uçar et al., 2021a). Ferenci and Kovács (2018) research used a data from a typical US health survey for (n = 862) adult males. Their study investigated the ability of predicting BFP from some easily estimated data: age, gender, height, weight, waist circumference and diverse laboratory test results. In addition, the simple anthropometric measurement represented by Body mass index (BMI) has been widely accepted as a simple and the most practical measure of obesity in clinical and epidemiological studies, even though it does not distinguish fat from lean body mass. BMI is calculated as the ratio of Body weight and the square of height. Kupusinac et al. (2014) predicted BFP from gender, age, BMI and BFP assessed by Tanita BIA model for a dataset include 2755 subjects (1332 women) and (1423 male).

Growth of 3-dimensional (3D) body scan technologies has smoothed significant anthropometric data collection models for biomedical studies and applications. A number of studies in clinical nutrition have explored the correlation between 3D body shapes and body compositions (Lu et al., 2018). Three-dimension object (3DO) devices exploit visible and infrared radiations to generate a 3-D image of the body, to capture visual automatic calculation of anthropometric features to be used in multiple anthropometry-based BFP prediction calculations (Harty et al., 2020). Lu et al. (2018) evaluated 3D-body-figures for an innovative model for BFP prediction. They presented the perception of “visual cue” by investigating the “second-order body shape signifiers”. Their regression model employed for feature extraction of the “zeroth-order body shape signifiers”. Then, the “visual cue” used as a shape identifier to expand the baseline prediction. Lu et al. (2020) presents a novel supervised induction model to analyze pixel-level body composition and level of muscle- to- fat ratio (i.e., BFP) utilizing 3D geometry optical shapes and body density. Their research employed (100 female) participants with each individual body density computed by DXA assessment to display a fat appropriation base prediction. Later utilized a “Bayesian network” to derive the possibility of the base prediction bias with “3D geometry” body features.

Harty et al. (2020) developed a unique BFP prediction technique. The study used 3DO measurement and a 4Component (4C) model to capture the Anthropometrical physical characteristics and body content analysis of 179 people (103 females, 76 males). Different specific assessment of five “full-body” scans made, all achieved with 3DO scanner technology. The body volume determined by air displacement plethysmography (BOD POD), bone mineral content measured by DXA, total body water evaluated by bioimpedance analysis spectroscopy, and body mass index measured by a standardized electronic measure were among the 4C of body. After a night of fasting from food and fluid, all of the participants' component measures were taken. The Texas Tech University Institutional Review Board approved the extensive efforts put into the data collection process. Despite the novelty of ML-based optical data, Complexity of test procedures and expensiveness of using DXA techniques and 3DO scan for body fat prediction, caused the small size of datasets involved in related machine learning researches. Besides, the anthropometrical assessments extracted from different 3DO structures may differ, even for analogous body positions, demonstrating that existing studies may likely specific to the appraised scanner.

A substantial association between ECG characteristics and obesity has been discovered in the literature, and several studies have recently relied on this new technology in data collection methods for the BFP prediction model. Akman et al. (2021) used the photoplethysmography signal (PPG) which is “A variant of the ECG signal”. The PPG signal which is a non-invasive signal that delivers information on the blood volume moving in the body around the skin. Data was acquired from 331 people (172 males and 159 females), and the BFP and PPG signals obtained from the BIOPAC MP36 device were computed using the BIA device. However, they supposed that gathering this type of data will reduce the high cost and risk associated with the DEXa based models. Uçar et al. (2021b) estimated BFP from a dataset depended ECG signal parameters and some basic and anthropometric information.

b) Machine Learning algorithms in BFP prediction

This section discusses the analyzation of the overall list of selected articles to answer the SLR(RQ2) which implies the major ML models employed in the BFP model. According to the implementation methodology of the ML models used the articles classified to single stage and hybrid models.

i) Single Stage ML based Algorithms for BFP Prediction

The AI algorithm used to solve any model is the most essential issue in any machine learning modeling problem. Several research have made use of data mining techniques such as ANN, SVR, and so on. ANN models widely used to handle real world complicated models, mostly when the problem is related to non-linear prototype or when the description of fundamental machineries is lacking. From previously described models, ANN may learn associations between input-output data pairs. For this capability, ANN is commonly used in the PBF prediction model.

Kupusinac et al. (2014) used a feed-forward ANN with a back-propagation training approach to estimate BFP using gender, AGE, and BMI. The investigated group comprises of 2755 respondents (1332 women and 1423 males) of (18-88) years age, all of them are from North Serbia. Data included (BMI, age, body height, and weight). Tanita BIA analyzer was used to determine BFP. The dataset was randomly divided into three sets for the ANN model, with the proportions 70:15:15 to create a training set with 1929 subjects, a validation set with 413 subjects, and a testing set with 413 subjects. To determine the best number of neurons in the hidden layer for the task, multiple single hidden layer feed-forward ANN topologies were explored. The best network topology was produced using a single hidden layer and a large number of neurons (n=31). The prediction accuracy was dependent on Mean Predictive Accuracy, which was attained at 80% with the provided model. The results were compared to various computational formulas for estimation BMI. The study underscored the significance of BMI in both genders' BFP prediction models. However, the study's findings were limited because they compared anticipated values using statistical methods of BFP computations.

Ferenci and Kovács (2018) present their research on how well BFP can be predicted using readily available data included basic information (age, gender), some basic anthropometrics (weight, height, waist circumference), and various laboratory petameters obtained from routine blood tests, as well as BFP estimated using a BIA device. In total, 39 variables were used from a representative US health survey (n = 862) of adult males (age > 18). They evaluated the predictive performance of three machine learning models: linear regression-based ordinary least squares, FFNN with various hidden neuron models, and SVM with gaussian radial basis function. a's data. Grid search is used to find the optimal parameters, and bootstrap validation is used to obtain realistic error estimates. According to the study's findings, "no approach accurately predicts body fat %," however support vector machines surpassed feedforward neural networks and linear regression (root mean square error 0.0988 0.00288,0.108 0.00928 and 0.107 0.012, respectively). Even with this highest performance, soft computing methods had an R2 of 44%, however this minor advantage is offset by the clinical interpretability of regression models.

In the Akman et al. (2021) approach for BFP assessment based on gender-based PPG signal data, the PPG signal is split into three lower frequency bands. Each frequency band and PPG signal has twenty-five statistical features extracted in total, including height, age, weight, and BMI for 331 participants (172 males and 159 females) surveyed for information. Finally, the dataset employed a total of 104 attributes. After collecting characteristics in five groups, BFP was calculated using ML models. Spearman feature selection strategy was employed for each group to capture significant characteristics for the machine learning algorithms MLFFNN and DT. The entire operation was carried out for each gender, with separate models constructed for men and women. The best performance of men was examined (RMSE=0.35, R=1), while the best performance of women was evaluated (RMSE =0.87, R=1). Another ECG- related model, proposed by Uçar et al. (2021b), BFP prediction model based (ANN and DT) with dataset of 225 ECG signal segments gathered from different QRS intervals, as well as age, height, and weight data. In terms of performance accuracy, R = 0.94 for men, and R = 0.91 for women. The study findings encourage the use of inspired and accurate models for BFP estimate within ESG data as an alternative for the costly DEXA model.

The purpose of a regression-base model is to notice the underlying mathematical association among a set of input observations and their corresponding output values. Support vector machine with €-intensive error loss function motivated by Vapnik (Vapnik,1998) is a viable technique in regression problems because to their interesting advantages in classification difficulties. Multiple studies based on different models of SVR algorithm were encountered for this advanced BFP prediction challenge. Xu and Wang (2012) developed a new SVR model known as the Twin SVR. Twin support vector regression (TSVR) is a novel regression algorithm that seeks €-insensitive up- and down-bound functions for training points. In order to accomplish so, smaller-sized “quadratic programming problems” must be solved as an alternative of a only huge one in a traditional SVR. TSVR, on the other hand, assigns the same penalty to the samples. At fact, samples in various positions have varying impacts on the bound function. Then, in this study, we present a weighted TSVR, where samples at different positions are offered to give varying penalties. To some extent, the final regressor can avoid the over-fitting problem while still providing excellent generalization ability. The body fat dataset was employed in the model's numerical experiments, and the results showed that the proposed methodology of TSVR with different parameter selection and multiple kernel functions outperformed the classical SVR. However, with SVR-based models, operation complexity remains a concern.

Chiong et al. (2021) proposed an “improved Relative error” SVM model for BFP estimation. The model exploited relative error constraints by adding extra bias error to explain a set of linear equations. The proposed model tested using two publicly available datasets in simulation tests (BODY FAT & NIHAN). The tests also feature five machine learning techniques for comparison: SVM, Relative error SVM, Multi-Layer Perceptron, Random Forest (RF), and Extreme Gradient Boosting (XGBoost). The proposed model with the additional “bias error” term, outperforms other models being examined.

For BFP forecasting model, Balasundaram and Kapil (2010) suggested a simple and linearly convergent Lagrange support vector machine algorithm for the dual of the twin support vector regression (TSVR). Though the technique initially requires the inverse of matrices, it has been demonstrated that they may be acquired by conducting matrix subtraction of the identity matrix by a scalar multiple of the inverse of a positive semi-definite matrix that arises in the original formulation of TSVR. The BODY FAT dataset used with the technique is simple to construct and does not require any optimization software. The suggested method's similar or superior generalization performance in less training time when compared to the standard and twin support vector regression methods clearly demonstrates its suitability and application. For performance estimation, their approach used the relative error function.

Fan et al. (2021) proposed recent research in BFP prediction and SVR, and their paper contributed the first use of fuzzy weighted gaussian RE-SVM for BFP prediction. By giving fuzzy weights to the proposed model's error constraints, the system creates a fuzzy-weighted operation to reduce the influence of noise input. To reduce the impact of noise data even more, they applied fuzzy weights to the dot product of the Gaussian kernel. In this scenario, sample importance is taken into account for the Gaussian kernel, and the dot product can be thought of as the similarity between each pair of samples. The suggested model's simulation results beat other ML-based models (ANN, SVR, RF, XGBoost) used for comparison. Testing the models with BODY FAT and NHANES datasets, the results revealed that the suggested model is successful for body fat prediction (BODY FAT, NHANES). The performance evaluation in terms of RMSE registered (4.3506) for BODY FAT dataset testing results and (5.7874) for NHANES dataset testing model. However, no SVR research has ever been able to discover the crucial aspects in BFP prediction, instead focusing on prediction accuracy by tweaking the algorithm's parameters.

Anthropometric variables are widely utilized in studies to evaluate obesity and predict the precise amount of BFP; nevertheless, researchers studied the "pixel level fat" prediction model. Several research that has been proposed in the sector have enforced an incidental mapping between "shape description" and "pixel level fat". An earlier study in body composition and 3D pictures proposed by (Harty et al., 2020) created a novel BFP prediction algorithm for 179 individuals utilizing anthropometrical information from 3D optical images and a 4C model. The model Constructed on (200) anthropometrical measurement locations, discovered by the 3DO scanning program, "stepwise" and "lasso" regression analyses were used to build a BFP formulation. The performance was assessed using cross validation, which yielded a fair performance of (R2=0.78). When compared to testing with Dexa-based Bland-Altman analysis, ML methods outperformed testing with Dexa-based Bland-Altman analysis in terms of estimated equivalency performance.

Lu et al. (2020), on the other hand, proposed a unique supervised inference model to predict "pixel-level body fat" using 3D geometry characteristics and body density. The dataset includes (100 female) DEXA body composition assessment individuals, with characteristics such as a bone-fat-lean tissue mass map, pixel-level fat percentage map, fat threshold map, and associated fat percentage threshold histogram. Their inference model was built in three parts. First, they used body density to build a base prediction. Second, they moderate the base prediction bias using a Bayesian network to analyze 3D geometry features. Third, if a sample is assessed as having a high chance of bias in base prediction, the prediction bias should be corrected. The simulation results revealed that the Bayesian network used for bias prediction had higher accuracy than the logistic regression, and the general model outperformed the BOD POD by 23.28 percent.

Deep convolutional networks have recently seen numerous applications in a variety of fields. Wang et al. (2021) proposed the first application of deep NN in a pixel-wise prediction model. They offer a multi-task deep neural network strategy that uses a conditional generative adversarial network to estimate pixel level body composition using only 3D body surfaces. The proposed method predicts 2D subcutaneous and visceral fat maps with good accuracy in a single network. They also included an interpreted patch discriminator, which improves the texture accuracy of 2D fat maps. The novel method's validity and efficacy are shown experimentally on datasets collected from computed tomography with a total of 270 patients. To estimate performance, simulation results were based on the mean square error. The suggested model beat competitive methods by at least 41.3 percent for whole-body fat percentage predictions, 33.1 percent for subcutaneous and visceral fat percentage forecasts, and 4.1 percent for regional fat predictions.

New technologies, such as sensor technology and communication technology, have emerged in the health care business (Shin et al., 2019). This breakthrough has numerous applications in a variety of medical fields. The requirement for ML algorithms to manipulate massive amounts of data, on the other hand, is critical. Shin et al. (2019) created a hardware device and software model for easy and accurate BFP prediction in the field of body composition assessment using sensor technologies. The developed hardware model incorporates an electrode-based wearable device for assessing body composition based on upper-body impedance data, while the software model employs deep NN for BFP prediction. Anthropometric and impedance data were collected from 163 university athletes and volunteers with varying levels of experience. The primary information involved in the first step of the study technique (age, gender, weight, height, waist, hip, BMI, WHR, BFP) was estimated predominantly utilizing BIA devise named (inbody-720). Second, the novel device was tested to gather upper-body impedance data in order to estimate the BFP using rules-based-linear regression approaches between the calibrated impedance value and the true BFP. The suggested software-based DNN was employed in various network architectures for BFP prediction to measure the performance of the new device in contrast to other traditional models and to determine the correlation between different BFP values estimated using devices. The correlation coefficient value was enhanced by roughly 9%, and the standard error of estimate was lowered by 28%, according to the results.

Gao et al. (2020) devised an extensible genetic algorithm for BFP prediction. Due to the enormous number of involved parameters, they suggest a novel parameter selection methodology for human body composition models. Following this phase, a body composition model was created using less input factors. In order to improve the calculation accuracy and weight of parameters, they proposed utilizing an improved adaptive genetic algorithm that integrates both roulette and optimum reservation techniques. Finally, they used BFP to compare the performance of the proposed model to that of other approaches using a dataset of 220 subjects. The output results of the simulations demonstrate that the suggested model's adaptability (0.9921) has a mean relative error of 0.05 percent, a mean square error of 1.3, and a correlation coefficient of 0.982. When compared to other models, our model's indices are the most adaptive and error-free. Furthermore, with a training time of 28.58s and a running time of 2.84s, the suggested model is faster than existing models.

In general, ANN and SVR was among the most important models involved for BFP prediction in Single stage models with different types of datasets, However, the black box concept related to ANN models, the difficulties accompany determining ideal network structure, and the vague in demonstrating the relationship between inputs and outputs for the prediction model, makes the ANN-based models for BFP prediction is not the optimal solution.

ii) Hybrid Machine Learning Models for BFP Prediction

While several studies employed single stage algorithms to predict PBF, others employed hybrid models with feature selection techniques to reduce feature dimensionality of the problem. Feature selection is the process of choosing a collection of characteristics that influence the prediction accuracy of the target variables/class labels in a specific machine learning model(s) (Naik, 2021).

Shao (2014) suggested a hybrid model in BFP with the goal of proposing single and hybrid forecasting models to predict BFP. Two modeling components are combined in hybrid models. The model's first component employs its own feature to capture the essential but few explanatory variables. The hybrid schemes' second component generates predictions based on the explanatory factors. MR, ANN, MARS, and SVR techniques included in the single-stage forecasting models. Because MR and MARS have a high ability to choose relevant explanatory variables, the hybrid forecasting models in this work are combinations of (MR & ANN), (MR & MARS), (MR & SVR), (MARS & MR), (MARS & ANN), and (MARS & SVR). The explanatory factors that are used to predict BFP using the MR and MARS models are determined utilizing Parson correlation coefficients between variables to identify those with strong collinearity. After removing variables with excessive collinearity, eight variables, X1, X2, X4, X6, X10, X11, X12, and X13, were utilized in the MR and MARS models to identify the most essential aspects for the prediction algorithms. For MAR and MARS, two sets with four and six parameters are indicated. All models could predict BFP with fewer explanatory variables.

Uçar et al. (2021a) used Multilayer Feedforward Neural Networks (MLFFNN), Support Vector Machine Regression Model (SVMs), and Decision Tree Regression (DT) as standalone models, and "MLFFNN & DT", "MLFFNN & SVMs", "DT & SVM", "MLFFNN & DT & SVM" structures as ML-based hybrid models. BFP and the Body Fat dataset with 13 features were used. The feature selection/sorting method sorts features based on their level of significance. The data set is divided into thirteen (13) distinct datasets. Each model was trained on a separate subset of the data. Performance evaluation criteria were calculated after each model. A statistical link between anthropometric characteristics and BFP was also studied in addition to these analyses. To reduce the data, PCA analysis was employed, and regression analysis was repeated. The prediction performance varies with the degree of feature selection. PCA analysis was also employed in order to reduce the number of features initially. MLFFNN was determined to have the best performance when compared to the four primary components. (DT) give significant performance over the investigated models in case of BFP prediction with single feature (abdomen 2 circumference which is X6 from BODY FAT dataset).

In a feature selection model for BFP prediction, Naik (2021) created the CS-FS approach, which adds a complex-step perception of the input feature to compute the feature sensitivity metric and find the essential features. It assesses the analytical quality of first-order derivatives without requiring additional calculations in neural networks. The implementation of complex-step perturbation as a feature selection approach in the framework of deep neural networks is described. Its effectiveness in determining essential features for Body Fat datasets is also proved. According to table2.2, the CS-FS model recognized the essential features in the sequences X6, X3,

X13, X4, X8, X2, X2, X1, X7, X5, X12, X11, X10, and X9. Furthermore, filter-based feature selection approaches are used, and the results of the suggested method are compared.

Farquad et al. (2010) introduced a hybrid rule extraction technique for solving regression problems. Their model consisted of two stages. In the first stage, SVR is used to extract support vectors from the training set in order to build two training sets. In the second phase, the collected sets were used separately to generate rules using CART, ANFIS, and DENFIS. We use the first training set to limit the number of patterns in the input space (since we just used support vectors), and the rules produced in phase 2 are not retrieved from SVR. However, by employing the second training set, we ensure that the rules developed in phase 2 are indeed derived from SVR. The Body Fat dataset was employed, and the average results of 10-fold cross validation, as well as the results on the validation set for the Body Fat dataset, are reported. The linear kernel generated the best prediction accuracy for this dataset, with an RMSE of 0.0189, and the collection of support vectors created from it is used in the rule development phase. With an RMSE of 0.0048 on validation data, our recommended hybrid “SVR & DENFIS” achieved the highest prediction accuracy.

Summary of SLR Results

Summarizing the results and outcomes from the review could be listed as follows:

- Best ML models in BFP prediction depended ANN, SVR, and DT.
- Single stage ML models depended custom datasets with different sizes (small, large).
- Hybrid models for BFP prediction, depended Body fat dataset which indicate a relatively small-sized set. Hence, No ML-based hybrid model proposed with a big sample dataset.
- The body fat data set is male restricted sample. Hence, all ML models for BFP prediction eliminated gender factor.
- Most feature selection models involved in hybrid models for BFP prediction based statistical models as indicator of relevant features.
- The literature did not register any study employed SVR model-based sensitivity analysis for feature selection in BFP prediction model.

- Genetic algorithm is optimization model that could be emerged with ANN to optimize the network structure and enhance the overall performance of the prediction model. However, no study in the literature invested these characteristics in BFP prediction model.
- In spite of all the importance shown by the performance of ANN model in BFP model, some newly emergent technologies that developed to enhance the learning ability of the traditional ANN, such as, EANN was not investigated in any study.

For best of our knowledge, according to the literature review outputs and analysis, there are missing gaps that we considered in the design of this study, first the (Body fat) dataset used with hybrid models considered small-sized dataset with 252 male subjects and analyzation of results inspected females' data. Second, the impact of some regressions depended in BFP estimation, such as, BMI&WHR, never discussed in these models. third, most hybrid models depended statistical models to identify the important features in the first step.

The proposed models in this study can solve many problems related to ML based hybrid models for BFP prediction with newly emergent models based EANN and a large dataset with gender variation.

CHAPTER III
Methodology

This chapter provides information about the research methodology and design, data collection and analysis procedures, Modelling algorithms, data normalization, and performance criteria.

Data Collection and Analysis

Dataset

With the primary data collected for the current investigation, the researchers attempted to improve a body fat percentage (BFP) prediction model that would be able to predict BFP with high accuracy rates based on relevant anthropometric parameters, as well as identify the most essential factors that influence the model. The dataset includes the following eight variables: Gender, height, age, weight, abdominal circumference, WHR, BMI, and estimated BFP. The BFP was determined using the BIA ACCUNIQ BC380 body composition analysis device located in the Department of Food and Nutrition at "Baxshin Hospital" in Iraq and overseen by qualified physicians. The abdominal C in this study referred to waist circumference, which is a "constant measure of abdominal obesity" (De Koning et al., 2007). A variety of methodologies were used to estimate the metric parameters and body composition of 2000 observations, which included observations from 387 men and 1613 women who were under the supervision of professional personnel. A total of 18 to 29-year-old people took part in this study, and the parameters were chosen based on the efficacy of anthropometric models in previous BFP estimation studies (Hussain et al, 2021).

In this research, the data collected comprises of a single observing per participant and did not provide timeseries data that may be used to forecast future BFP based on current anthropometric and laboratory measures. As a consequence, the study focused on properly forecasting recent BFP to minimize BFP prediction time and cost. Table3.1 summarizes the data descriptive statistics.

Table 3.1

The Dataset's Descriptive Statistics

Parameter	*"BMI" kg/m²*	*"Abdominal C" cm*	*"Weight" kg*	*"Gender"*	*"Height" cm*	*"WHR" cm*	*"Age"*	*"BFP"*
Mean	26.2180	86.63450	67.050	1.810	160.3010	0.8250	23.050	33.420
Standard Deviation	10.330	18.0310	21.7610	0.400	9.46380	0.0554	3.430	11.830
Min	12.90	60.70	15.400	1.000	63.000	0.660	18.00	3.00
Max	237.3	182.7	183.70	2.000	199.80	2.10	29.00	87.70

Note: Primary dataset used with the permission of Ethical Comette with application code (NEU/AS/2021/134)

The Pearson Correlation matrix, as shown in Table 3.2 and Figure 3.1, indicates how well a linear function may describe the relationship between the BFP variables. The direction or sign has no effect on the association's strength. In other words, a positive correlation suggests that a raise in one parameter leads to a raise in the other, whereas a negative coefficient shows an inverse connection, in which one parameter rises while the other falls (Eisinga et al., 2013; Ghali et al., 2020; Selin & Abba, 2020; Usman et al., 2020).

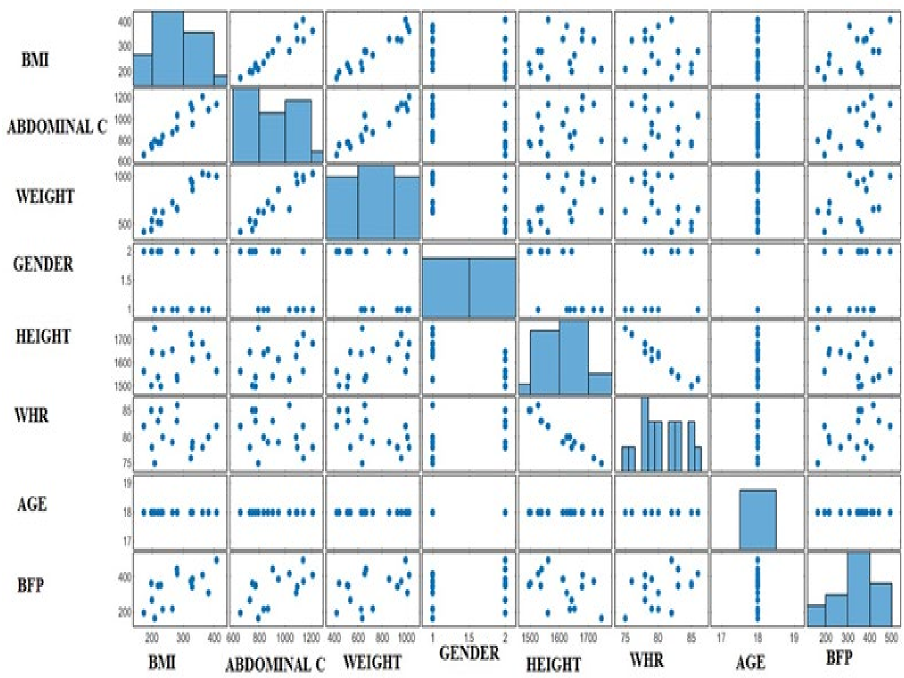
Table 3.2

Variable Correlation Coefficients Calculated from Correlation Matrix

	BMI	*Abdominal C*	*Weight*	*Gender*	*Height*	*WHR*	*Age*	*BFP*
BMI	1							
Abdominal C	0.9531	1						
Weight	0.9279	0.9589	1					
Gender	0.0118	-0.155	-0.2183	1				
Height	-0.0036	0.1203	0.3288	-0.6017	1			
WHR	0.0047	-0.0698	-0.2683	0.4822	-0.9358	1		
Age	0.2479	0.2042	0.2216	0.0289	-0.0059	0.0490	1	
BFP	0.8086	0.7583	0.6277	0.4411	-0.3979	0.3643	0.2097	1

Figure 3.1

The Correlation Matrix between the Variables



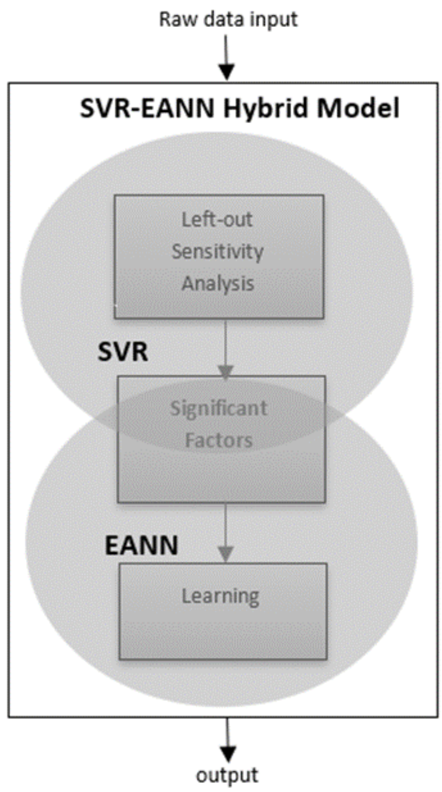
Design of Research Hybrid Model1 (SVR-EANN)

Hybrid models combine the varied capabilities of several models into a single system, resulting in much improved outcomes in many cases (Nourani et al., 2020a).

This study built SVR-EANN as a hybrid model for prediction of BFP by combining a novel production of neural networks (EANN) with SVR. The hybrid model proposed in this work is divided into three stages, as seen in Figure 3.2. The left-out sensitivity analysis utilizing the SVR model, which was carried out in order to choose the most salient features from the dataset, served as the initial stage in the creation of the proposed system, and in the second stage, an EANN model was built to train the dataset to predict accurate BFP values base on the significant parameters which had been nominated in first stage. Lastly, the performance of the EANN model was evaluated and then contrasted with that of the classic Feed Forward Neural Network (FFNN).

Figure 3.2

Schematic of the SVR-EANN Hybrid Model



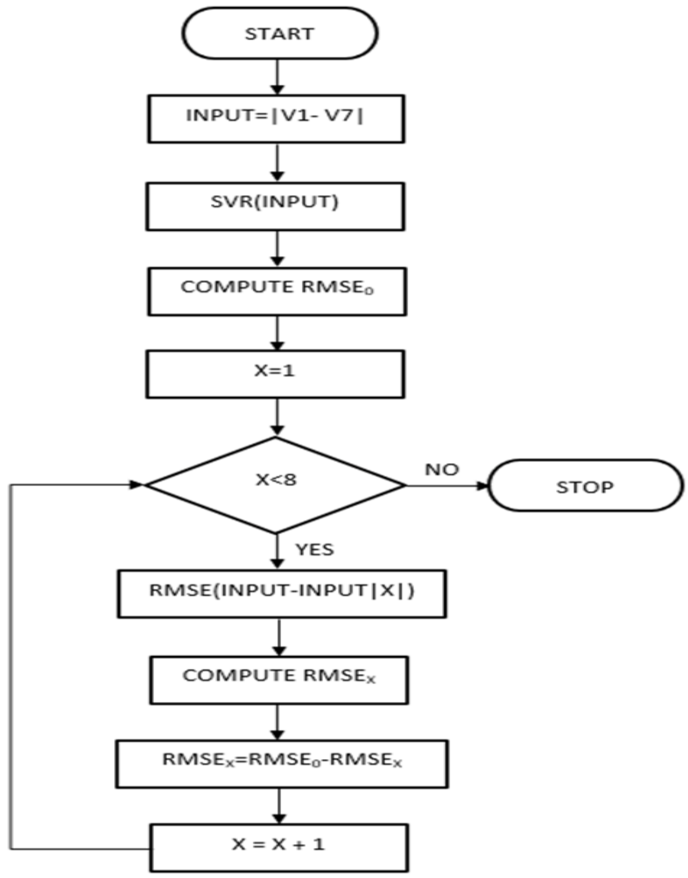
Note: SVR and EANN Models applied separately

Input variable Selection

One of the main requirements for constructing a data-driven model is having the ability to choose the varying input parameters (Nourani et al., 2020b). In this case, the left-out approach (SVR) is based on artificial intelligence (AI) (Nourani et al., 2019), In a sequential procedure to test variable affect in the prediction model, a variable was omitted from the training phase in order to allow the SVR model to be trained utilizing all of the available variables. The training procedure was then finished using the remaining variables. This was the point at which the omitted input was re-applied to all other inputs. We were able to discover which elements were the most responsible for the model's inaccuracy by employing this method, and as a consequence, we discovered that the model accuracy affected with the loss of the related variable (larger error estimated with the loss of efficient feature). When the major input is obtained and then replaced with a less crucial variable, the model's performance suffers a significant decrease in efficiency (here left-out variable). The grid search strategy using cross-validation was used to fine-tune the SVR parameters while training, which yields the best validation accuracy picked in the base of multiple parameters pooled in the domain, and several distinct parameter values were used. This method is tedious and naive, but it is still a better option than several advanced methods. Although it is time-consuming, if you have the resources, the process can be streamlined by first selecting the better regions with a coarse grid and then refining those regions with a fine grid (Hsu et al., 2003). Figure 3.3 depicts SVR left-out sensitivity analysis block diagram.

Figure 3.3

Schematic Representation of the SVR Omitted Sensitivity Analysis



Note: feature selection based SVR

In the first step of the hybrid model SVR-EANN design, the SVR model was used to decide the most fundamental parameters linked with BFP prediction. The second stage involved training the EANN model to predict BFP employing the stated variables. The backpropagation algorithm was used to carry out the training process. The EANN models' structure was established through several testing, and it was between (2-30) hidden neurons and five hormone neurons, based on the findings. Before deploying the EANN, it was trained for 50 epochs using the TanSig activation function to get better results.

Design of Research Hybrid Model2 (GAEANN)

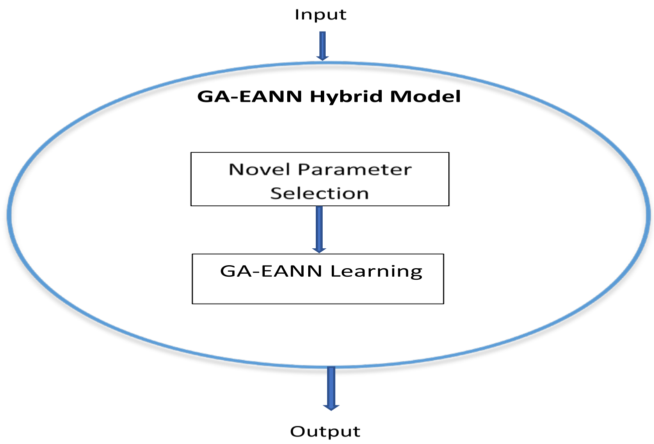
Evidence in literature showed the importance and effectiveness of EANN in modeling nonlinear problems even with leak in data, beside the importance of GA for optimization the network structure (Leung et al., 2003), hence we produce a hybrid model that combines GA with EANN for the first time in the problem of BFP prediction. Figure 3.4 views the GAEANN model design. The model employs a novel parameter selection depends subsets of parameters from original row data. Then the hybrid GAENN run separately with each subset to predict BFP.

Dominant Input Variables Identification

The dataset involved seven input features (gender, age, height, weight, abdominal circumference (AC), WHR, and BMI). Feature selection scenarios for the proposed prediction model depended the set of parameters (age, gender, height, weight, abdominal C) as primary simple independent parameters. While, most existing models for BFP prediction relays on only the primary anthropometric parameters, there is need to explore the potentials of some anthropometric regressions like BMI and WHR as secondary dependent parameters computed base in the primary parameters. This could be better understood by considering the newly adopted visualized correlation matrix between the parameters as indicated in Figure 3.1. For the purpose of this study, a novel selection model for parameter selection depended. Three subsets (M1, M2, M3) of parameters were derived from the row data as dominant input variables for the prediction model of the BFP. The first model (M1) uses the secondary parameters (BMI, WHR), second group (M2) named (age, gender, BMI, WHR) as predictors, while the third group (M3) consider the primary parameters (age, gender, height, weight, abdominal C). The data is divided in to training (calibration) and testing (validation) with 70% used for training and 30% for testing.

Figure 3.4

Schematic of the GAEANN Hybrid Model



Note: parameter selection differs from that used in model1

Artificial Neural Networks

An artificial neural network (ANN) is a computer simulation of the nervous system that mimics the behavior of a human. Additionally, because they are capable of learning associations from examples, in circumstances where the input and output data are not clearly linked, they are ideal. Every aspect of our life and scientific domains, such as economics (Kashman, 2010) and neurology (Ozsahin et al., 2019), has its own set of ANN research and applications that are unique and varied.

In order to convert and process information, the ANN's structure is built up of computational units (i.e., neurons) that are connected together by mathematical connections. It has become widely employed in recent decades because to its superior ability to model nonlinear issues without prior knowledge of the system. For its simplicity and capacity to perform a wide range of regression and classification tasks, the multilayer perceptron MLP feed-forward neural network is one of the most extensively used artificial neural networks (ANNs). Three-layer MLP is widely used. It has three layers: input, concealed, and output. Weights connect each layer's neurons to the next layer's neurons (s). Weights change as you study. A perceptron learns by translating a linear combination of weighted inputs into a nonlinear output (Liu et al., 2019). The output's equation is:

$$y = \varphi(\sum_{k=1}^{m} w_k x_k + z) \quad (1)$$

Where x k (k=1,..., n) denotes the input vector, wj (j = 1,..., m) denotes the related weights, z denotes a bias, and $\varphi(x)$ denotes the activation function. The sigmoidal active function is utilized in this investigation (Eq. 2). It can convert any integer between [0 and 1] to a value between positive and negative infinity.

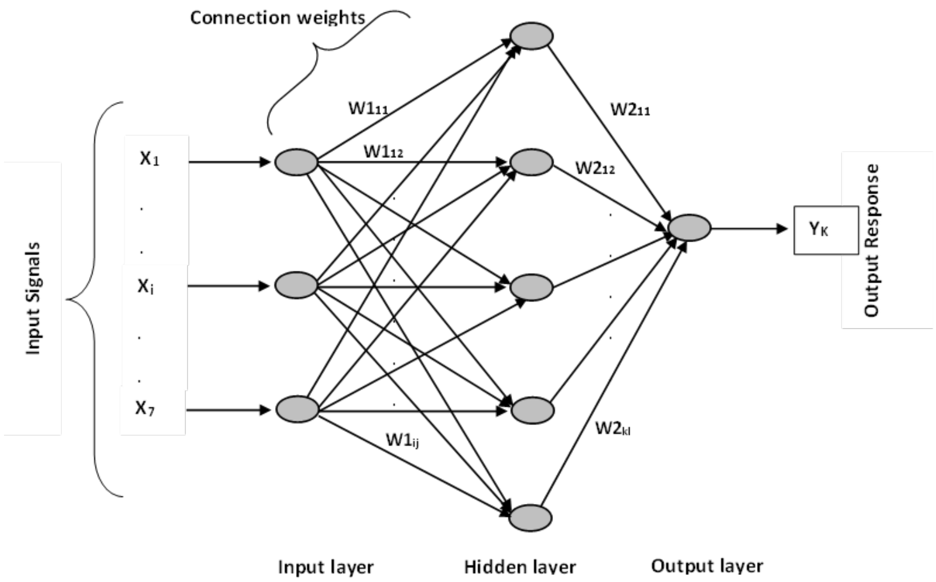
$$\varphi(x) = \frac{1}{1+e^{-x}} \tag{2}$$

Back-propagation training is used to train the MLP. The process uses a gradient descent technique, in which all the weights are treated as arguments. In order to reduce the difference between the forecast values and the measurements, the weights are updated.

In this study we employed the FFNN in the proposed models for comparison, to prove the efficiency of the new derived hybrid models based EANN structure for BFP prediction. Figure 3.5 shows the typical architecture of FFNN employed in the research.

Figure 3.5

Structure of the Three-Layer FFNN with Single Neuron in Output Layer



Note: The 3-layered FFNN architecture depended in all models of this study with different input variables.

Emotional Artificial Neural Networks

The building of an Emotional Artificial Neural Network (EANN) customs an enhancement in traditional machine learning techniques. With EANNs, the emotional system is built into the network, allowing neurons to produce hormones that boost cognitive, physical, and emotional capacities (Nourani, 2017; Nourani et al., 2020a).

The EANN model incorporates emotional feelings into the feed forward neural network (FFNN). The emotional mechanism used to the EANN makes the network's performance improve. One major difference between the (Figure 3.6b) feedback mechanism and FFNN is that feedback from the input and output neurons influences hormonal parameters, whereas the flow of information is in one direction only in FFNN (from input to output). In order to calculate the dynamic hormonal coefficients in EANN, such as Ha, Hb, and Hc, hormone glands such as these were proposed as formulas that can be initiated and modified over training epochs depending on the input-output (sample) patterns. In the hormone field, one research study found that some hormones might change network components such as activation functions, network functions, and the weights. Therefore, hormones could potentially affect network dynamics as well. The ith output neuron can be defined mathematically as stated by Gholami et al. (2019) as:

$$Y_i = \underbrace{\left(\rho_i + \sum_h \theta_{i,h} H_h\right)}_{I} \times f\left(\sum_j \left[\underbrace{\left(\alpha_i + \sum_h \psi_{i,h} H_h\right)}_{II} \times \underbrace{\left(\beta_{i,j} + \sum_h \varphi_{i,j,k} H_h\right)}_{III} X_{i,j} + \underbrace{\left(\lambda_i + \sum_h \phi_{i,h} H_h\right)}_{IV}\right]\right) \quad (3)$$

Where i, h, j indicates the input, hidden, and output neurons of the ith layer, correspondingly, but f () specifies the activation function of the ith layer. H h, which is defined for three hormones H a, H b, and H c as, represents the overall hormone value.

$$H_h = \sum_i H_{i,h} \qquad (h = a, b, c) \quad (4)$$

The term I in Eq. (3) includes the hormonal factor of $\sum_h \theta_{i,h} H_h$, refers to the connective weight used for the activation function. Whereas (II) denotes the net function's weight (summation), The third term (III) denotes the related weight for input X (i,j), which is determined by the jth neuron in the preceding layer; the term (IV) denotes the bias in the net function, which includes the hormone coefficient $\sum_h \phi_{i,h} H_h$within a fixed neural weight λ_i. The impact of the system's hormonal level (H_h) is managed by by $\theta_{i,h}$, $\psi_{i,h}$ and $\varphi_{i,j,k}$. Additionally, the neural output (Y_i) supplies the system with hormonal feedback $H_{i,h}$ in a closed loop fashion, as below:

$$H_{i,h} = glandity_{i,h} \times Y_i \tag{5}$$

The calibration parameter $glandity_{i,h}$ specifies the hormonal level for each gland Hi. The initialization of the value of each hormone (H h) can be done in a variety of ways based on the input patterns. For instance, an initialization method may be based on the mean value of the input parameters for each sample (input vector). As a result, during training rounds, the values of the hormones are updated based on the network output Y I as well as Eqs. (4) and (5), until an acceptable level of agreement between the anticipated and observed values is achieved.

Numerous studies presented a reduced version of EANN by omitting certain elements and modules from the overall EANN network (Khashman, 2008; Lotfi & Akbarzadeh-T., 2014), as well as references therein. Khashman (2008) established the Emotional Backpropagation (EmBP) method in this field by modernizing the normal Backpropagation algorithm (BP). Two extra parameters (emotional anxiety (μ) and confidence (k), (0<μ,k<1)) are utilized in addition to the standard BP momentum rate (∝) and learning factor (η) to further decrease inaccuracy in EmBP. A relationship is established between the anxiety scores throughout each training cycle and the output error and input pattern. At first, there is a lot of anxiety and a lack of confidence. But as training develops, the latter increases in importance while the former declines, culminating in the highest degree of confidence and the lowest amount of fear towards the conclusion of the training session. It is possible to overtrain the conventional FFNN when it is trained using the standard BP approach, however this may be minimized by including EANN into the training process. This is because throughout the training phase, the hormonal parameters in EANN allow the network to adapt to a range of system situations. Also, in EANN models, the output and hidden layers are identical. The "hidden" layer's input is also obtained from the potential of all input values rather than the individual values themselves. The output error is communicated backward for each iteration of EmBP training in order to control the hidden layer weights (wjh) and bias (wjb):

$$w_{jh}(new) = w_{jh}(old) + \mu.\Delta.YH_h + \alpha.\left[\delta w_{jh}(old)\right] \quad (6)$$

$$w_{jb}(new) = w_{jb}(old) + \mu.\Delta + \alpha.\left[\delta w_{jb}(old)\right] \quad (7)$$

As well as, w_{jm} indicates emotional 'weights are attuned by:

$$w_{jm}(new) = w_{jm}(old) + \mu.\Delta.Y_{avg} + k.\left[\delta w_{jm}(old)\right] \quad (8)$$

Where $\delta w_{jh}(old)$, $\delta w_{jb}(old)$ and $\delta w_{jm}(old)$ correspond to the previously substituted hidden layer weight and bias, as well as emotional weights. YHh and Y_{avg} represent the hth hidden neuron's output and the input pattern's average value for each iteration. The following are anxiety and confidence indicators marked by μ and k:

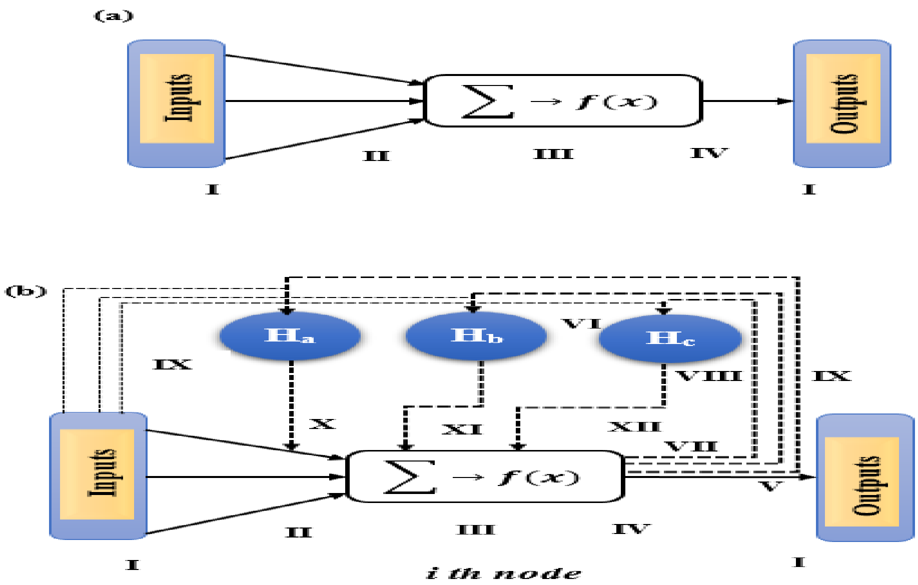
$$\mu = Y_{avg} + \Delta \quad (9)$$

$$k = \mu_0 - \mu \quad (10)$$

Whereas μ_0 is the anxiety factor's value after the first iteration Similarly, the bias and weights of the input-hidden layer are changed. The EANN network is regularly subjected to the imposition of normalized data in order to improve accuracy. The implementation of the EmBP algorithm in EANN networks is explained in further detail in (Khashman, 2008; Lotfi & Akbarzadeh-T., 2014; Nourani ,2020a,; Roshni, 2020). The EmBP learning algorithm (Khashman, 2008) is employed in this study to construct the EANN model.

Figure 3.6

Neuron Architecture in FF-BPNN &EANN



Note: Neurons in (a) FF-BPNN and (b) EANN (combined with emotional units) (Gholami et al., 2019).

Regression Using Support Vectors

Initially suggested for classification problems, the support vector machines have been effectively used in subsequent research areas (Cavus et al., 2021), but the SVR learning model has shown to be capable of providing an acceptable and realistic answer to the problem of prediction, classification, pattern recognition, and regression. An additional function of the SVR models is that it minimizes structural risk, which differentiates it from other machine learning approaches like ANNs. SVR differs from other machine learning models in that it examines data points rather than data itself. Data points are used to choose support vectors for efficiency and structural safety. To evaluate the model's effectiveness for both linear and nonlinear difficulties, utilize this hyperplane.

Figure 3.7 depicts the SVR model's overall structure, while Equation provides the SVR Equation 11 (Wang et al., 2013):

$$(\boldsymbol{x}, \boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i^*) = \sum_{i=1}^{N} (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^*) \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}_i) + \boldsymbol{b} \ ) \quad (11)$$

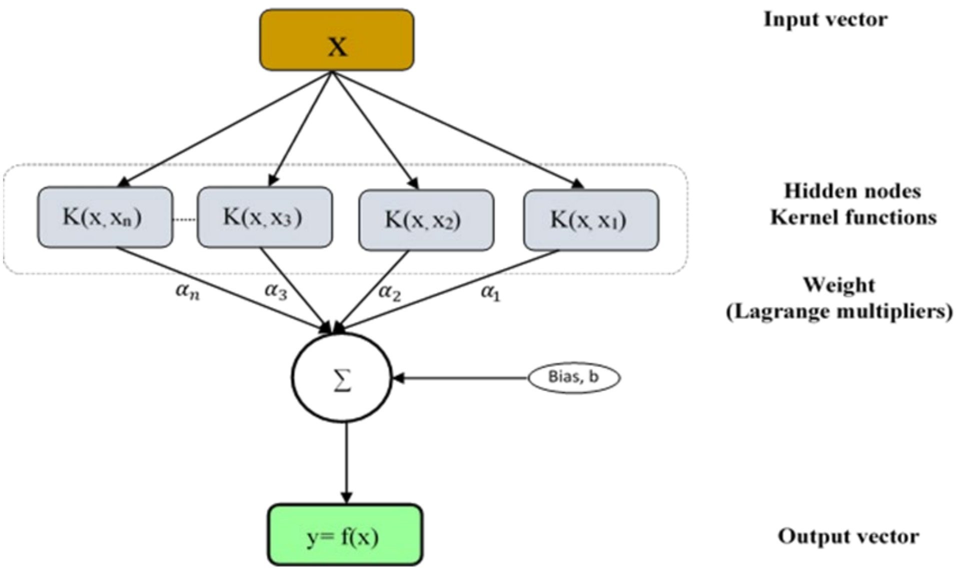
Where x and b denote the input vector and bias term, correspondingly, αi, αi*, K denote the Lagrange multipliers and kernel function, respectively.

The regression line errors are minimized by employing the kernel function K(x,xi) to examine projected support vectors, which are the closest data points. This error minimization and function optimization approach uses Lagrange multipliers to find and optimize local maxima and minima. The radial basis function (RBF) kernel was employed in this analysis, as it is the most often used kernel in the SVR (Sekeroglu, B.& Tuncal, K. ,2021).

The SVR in this study used as feature selection methodology with sensitivity analysis based left-out strategy in the hybrid model of SVR-EANN.

Figure 3.7

Conceptual Structure of SVR Model (Oytun et al., 2020)



Note: the kernel function used radial based kernel

Hybrid Genetic Algorithm- Emotional Artificial Neural Network

Genetic algorithm GA is an optimization technique that seeks to minimize a cost function (fitness) in order to determine the optimal solution to a problem (best chromosome) (Lotfi et al., 2014). In this study, we increase EANN's crisp numerical weights using a genetic algorithm (GA). The first step is to initialize the chromosomes. A chromosome is constructed and represented in the following manner:

Choromi = [v1 , v2 ,…, vn , vn+1 , w1 , w2 ,…, wn , b] (12)

So, if there are n input characteristics, there are 2n+2 learning weights, for example, the number of genes per chromosome. Assume Yk is the model's output with Choromi's weights in response to Pk.

Yk = E (Pk ; Choromi) (13)

E (Pk; Choromi) is the pattern output of the model for the kth pattern with the specified weights in Choromi. The following formula is used to determine this chromosome's fitness function:

$$fitness(choromi) = \frac{1}{m}\left[\sum_{k=1}^{m}(Y^k - T^k)^2\right]^{0.5} \quad (14)$$

There are a total of m pattern-targets in this example. Tk is the target associated with the kth input pattern (Pk), and m is the total number of pattern-targets in this case. As a result, lowering the fitness function also means lowering the overall error across all training samples. The next part describes supervised GA-based learning for the EANN architecture (Lotfi et al., 2014):

GA-based Learning algorithm of ENN

Input: Pc×n is a matrix containing c patterns with n characteristics, and Tc×1 is an array containing the targets of c patterns.

Output: For a single output, the array Wm(2n+2) includes(2n+2) weights from the array Wm.

Create a population of actual 2n+2 genes for each chromosome at the start of the experiment.

Equation (14) may be used to calculate the fitness of each individual.

Choose the people you want to work with.

Mating

Mutations

Replace the impoverished with powerful children in order to produce the next generation of people.

If stop criterion is not satisfied goto step 3 else

W(i) = best chromosome

End

The GA is used to find the network's optimum parameter. The input-hidden layers and hidden-output layers employ "Tansig" and "Logsig" transfer functions to convey data. The EANN model is also optimized using a Genetic Algorithm (GA). GA got the best EANN prediction model with 1500 population and 3500 generations.

Data Normalization and Performance Evaluation Criteria

Grouping all predictor and regressor variables in the same numeric range before feeding them into the intelligent models prevents data from the lower numeric range from being dominated by data from the higher numeric ranges (Nourani et al., 2019a). To increase model accuracy and speed up convergence, data normalization can reduce the number of numerical computations necessary in the system. Using Equation 15, the data in this inquiry was normalized between 0 and 1.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (15)$$

The normalized data value is xnorm, while the observed, maximum, and minimum data values are x, x_{max}, and x_{min}, respectively. The normalized data set is separated into two parts for model development: 75 percent for calibration and 25 percent for verification.

The effectiveness of the modeling methods was assessed through some functions that estimate efficiency and efficacy of the proposed intelligent models. Both the R^2 and the RMSE were used. R^2, which has values ranging from - ∞ to 1, is used to assess the model's fit. The model's performance increases as the R^2 approaches 1. As the R^2 moves closer to 1, the model's performance improves (Elkiran et al., 2018). The root mean squared error (RMSE) formula can be used to determine the variance between observed and calculated data. The optimal model is the one with the smallest root mean square error (RMSE). The best model is one with the lowest R^2 and RMSE values.

Additionally, it was utilized to calculate the remaining root mean square error RRMSE. The definition of remaining RRMSE states that excellent results are obtained if the values are less than 10 percent, whereas good results are obtained if the values are between 10 percent and 20 percent, but Fair results are obtained if the values are between 20 percent and 30 percent, and unacceptable results are obtained if the values are greater than 30 percent (Rabehi et al., 2020). Additionally, Equations 16-19 are utilized for performance evaluations.

$$R^2 = \mathbf{1} - \frac{\sum_{i=1}^{n} \left(N_{obs_i} - N_{pre_i}\right)^2}{\sum_{i=1}^{n} \left(N_{obs_i} - \overline{N_{obs}}\right)^2} \quad . \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(N_{obs_i} - N_{pre_i}\right)^2}{n}} \tag{17}$$

$$\text{rRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n} \left(N_{obs_i} - N_{pre_i}\right)^2}{n}}}{\frac{1}{N}\sum_{i=1}^{n} \left(N_{obs_i}\right)} \text{ x } 100 \tag{18}$$

$$R = \frac{\sum_{i=1}^{n}\left(N_{obs_i} - \overline{N_{obs_i}}\right)\left(N_{pre_i} - \overline{N_{pre_i}}\right)}{\sqrt{\sum_{i=1}^{n}\left(N_{obs_i} - N_{obs_i}\right)^2 \sum_{i=1}^{n}\left(N_{pre_i} - N_{pre_i}\right)^2}} \tag{19}$$

The number of observations is denoted by *n*, the mean value of the observed data level is denoted by $\overline{N}_{obs}$, and the predicted values are denoted by N_{pre}.

The Validation of the Model

For prediction issues, the main purpose of applying intelligent prediction models is the production of a reliable result that would be impossible to achieve employing traditional methodologies without prior information and a complete comprehension of the subject matter in question. However, many data-driven models suffer from overfitting issues, causing their performance during the calibration stage to diverge from their performance during the verification step, making accurate predictions for unknown datasets unfeasible. This necessitates model validation in order to avoid overfitting problems. Different forms of validation processes have been applied in literature, for example, "cross validation", "holdout validation", "leave one out validation", etc, however the SVR-EANN model in this work used k-fold cross validation and holdout validation applied to the GAEANN. For the first model, the dataset is divided into an equal number of subgroups in this sort of validation process. k-1 subsets are used to calibrate the model, with the remaining subset being used for verification. When the procedure is repeated k times, it is considered complete when all k-subsets have been utilized for modification training and testing. In the verification stage, the final performance is computed by averaging the outcomes of k- subgroups and calculating the average of those values. One of the key advantages of k-fold cross validation is the total independence of the training and testing subsets (Sharma et al., 2018). It is also possible to improve the efficiency of data utilization by using cross validation techniques. When using 4-fold cross-validation, the data set (normalized) is separated into two halves (training = 75% and testing = 25%) for the purpose of improving the performance of models. The number of k values to be employed is determined by the data size, which is typically between 2 and 10 values.

CHAPTER IV
Results

The results of Implementation for the proposed models of SVR-EANN and GAEANN modeling approaches are presented in this chapter. MATLAB2019b software used to implement the modeling approaches and present the simulation analysis of results.

Results of Model1 SVR-EANN

a) Identification of Important Factors

In this study, the researcher performed a left-out sensitivity analysis, which included the use of the SVR model, to indicate the exclusive impact of various “input” parameters in the prediction model of BFP. Four phases were included in the SVR-based left out sensitivity analysis technique. To forecast the BFP, the SVR model was trained and evaluated in the first stage, utilizing all of the possible input parameters as input variables. Second, the model’s prediction error (RMSE) was calculated. The third phase involved deleting a parameter (for example, Abdomen C) from the previously trained and tested model and training and testing a new model without that parameter (Abdomen C), with the resulting error determined. Finally, after eliminating Abdomen C from the testing stage, the equivalent increase in the RMSE value was calculated and used to rank the parameter's relative importance. Table 4.1 shows the outcomes of repeating the method for all of the parameters (Hussain et al., 2021). The increase amount of RMSE value suggests a larger relative significance, whereas a reduction in RMSE values indicates a lower relative importance. This novel strategy should improve feature selection methodology to obtain most related features to BFP prediction.

As can be observed in the SVR sensitivity study findings (Table 4.1), when Abdominal C, weight, height, WHR, and BMI are removed from the input parameters, the RMSE values rise, however when age removed, the RMSE values decrease. When any of the components were removed from the model, the RMSE increased, demonstrating that all of the parameters are required for estimating the BFP. According to the data,” abdominal C”, “gender”, “height”, “WHR”, “BMI”, “weight”,” age” were ranked to be the top relevant variables in BFP prediction as the 1st, 2nd, 3rd, 4th, 5th, 6th, and 7th most important factors, respectively. The age parameter was the least influential, with just a 0.0035 increase in the RMSE value (Hussain et al., 2021).

Table 4.1

Sensitivity Analysis Output

Removed parameter	RMSE
Body Max Index	0.0564
Abdominal Circumference	0.1509
Body weight	0.0341
Gender	0.1223
Body height	0.0708
Waist-hip-Ratio	0.0589
Age	0.0157

Note: RMSE0 = 0.0122 (Normalized).

b) Regression Results and Comparisons

The suggested model’s performance in the first stage was compared against FFNN as a traditional NN, EANN as a new generation of neural networks, and a two-phase hybrid SVR-EANN for BFP prediction . While, the second stage employed more benchmark algorithms in the present investigation. The FFNN model was progressed employing MATLAB19b toolbox, while the EANN models were trained using a MATLAB code. The Levenberg Marquardt backpropagation algorithm was used to train a three-layer FFNN. Likewise, a simpler variant of EANN training utilizing the backpropagation technique was utilized to mimic the BFP. The accuracy of neural network models is dependent on the use of a good model structure. As a result, a hypersensitivity analysis was performed to determine the optimal model structure by varying the number of “hidden neurons” between (5-30), “emotional hormones” between (2-20), “training epochs” between (10-1000), percentage of training data (50-80), and” activation function “tested tansig function, logsig function,and purelin function. The trial-and-error method was used to find the model’s optimum structures. The models' performances for both “training” and “testing” phases were evaluated using four performance criterions (R^2, RMSE, RRMSE, and R), as shown in Table 4.2.

Table 4.2

Modelling Results for Model1

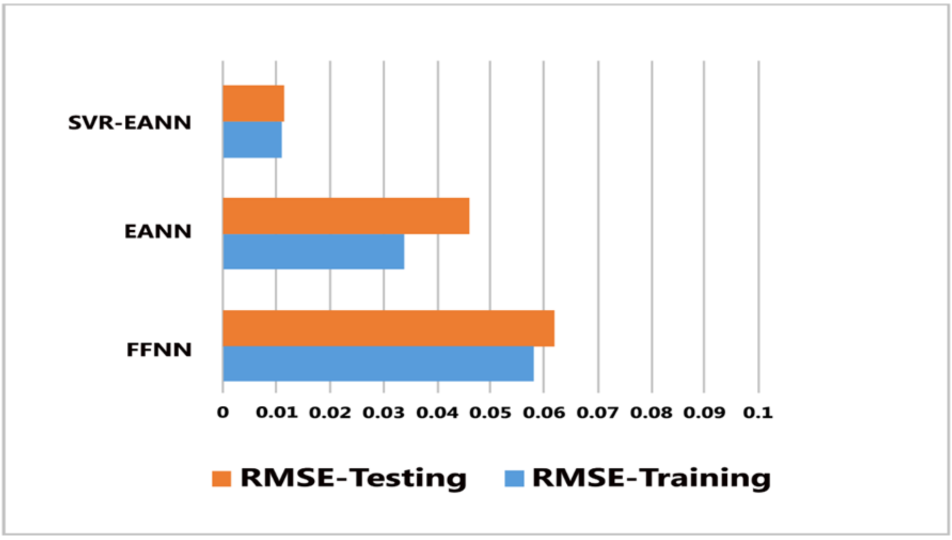
	Training				Testing			
Models	R^2	RMSE	RRMSE	R	R^2	RMSE	RRMSE	R
FFNN	0.850	0.0581	24.3773	0.9836	0.8573	0.0622	19.6919	0.9816
EANN	0.9306	0.0349	8.7888	0.9804	0.8928	0.0464	13.5565	0.9695
SVR-EANN	0.9935	0.0115	3.3470	0.9969	0.9911	0.0125	3.1513	0.9956

Note: data were trained with 4-fold cross validation.

ML-based modeling approaches related to (Model1) generated numerous models, each with a unique structure, and training procedure, but only the best models are shown in the Tables. All models demonstrated trustworthy performance in terms of BFP prediction, with NSE values greater than >0.8 during the training and verification stages. This is because AI-based models are more capable of representing complex processes such as BFP and also have a higher capacity for generalization than linear models (Nourani et al., 2019b). The SVR-EANN model beat all other models in this study during the verification stage, with a higher R^2 value and the lowest error (RMSE). As indicated in the sensitivity analysis result, the SVR-EANN did not include the age, which is the least essential component in the BFP prediction, but the EANN and FFNN models employed all 7 input parameters to model the BFP. At both the training and verification phases, the SVR-EANN outperformed the EANN and FFNN, with R^2, RMSE, RRMSE, and R values of 0.9911, 0.0125, 3.1513, and 0.9956, respectively, at the verification stage. The analysis results of the models depicted the high performance of the hybrid model SVREANN in compare to other single models, it was discovered that the SVREANN outperforms the EANN and FFNN models in the verification stage by up to 10(%) percent and 13(%), respectively. It's possible to combine the unique benefits of each SVR model's selection of important input parameters with the EANN model's higher performance, as illustrated in Figure 4.1.

Figure 4.1

Error Plots for the Models (FFNN, EANN, SVR-EANN)



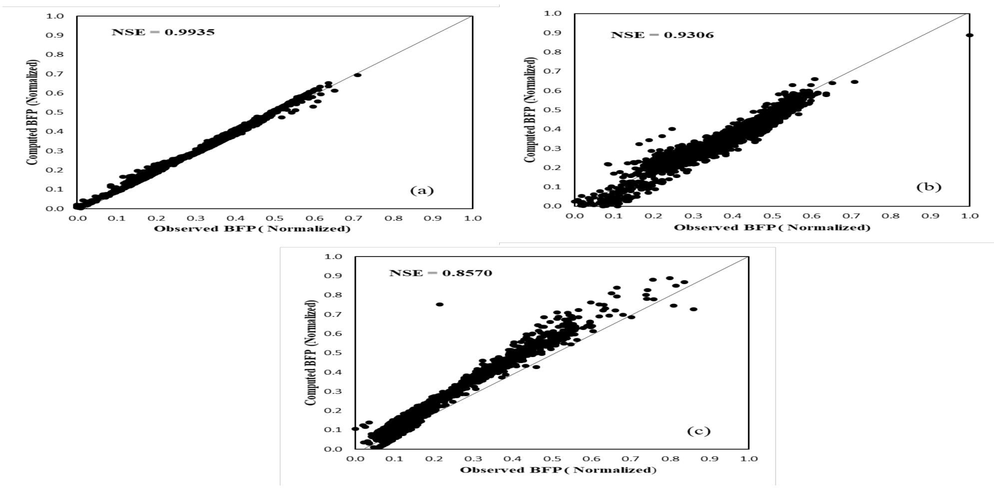
Note: Error plots for Model1, SVR-EANN hybrid model achieved best performance.

In order to capture a greater performance, the scatter plots of fitness presented. The SVR-EANN fits the data better than the other models in the both training and verification steps, as shown in Figures 4.2 and Figures 4.3. The data is more compressed along the diagonal line than in other models, indicating that it has a higher goodness of fit than other data-driven models.

During the second stage of the performance evaluation of SVR-EANN, the findings of the SVR-EANN contrasted with other benchmark models: SVR, DT, RF, LR, XGBoost, and GradBoost. The mean squared error (MSE) criteria were used to build the DT for regression investigations. With 100 and 150 estimators, respectively, GradBoost and RF were trained. The XGBoost and GradBoost algorithms have learning rates of 0.3 and 0.08. The values of "estimators" and "rates of learning" for RF, GradBoost, and XGBoost algorithms were verified. The training models data validated with "4-fold cross validation". Best values registered after several attempts of trial and error.

Figure 4.2

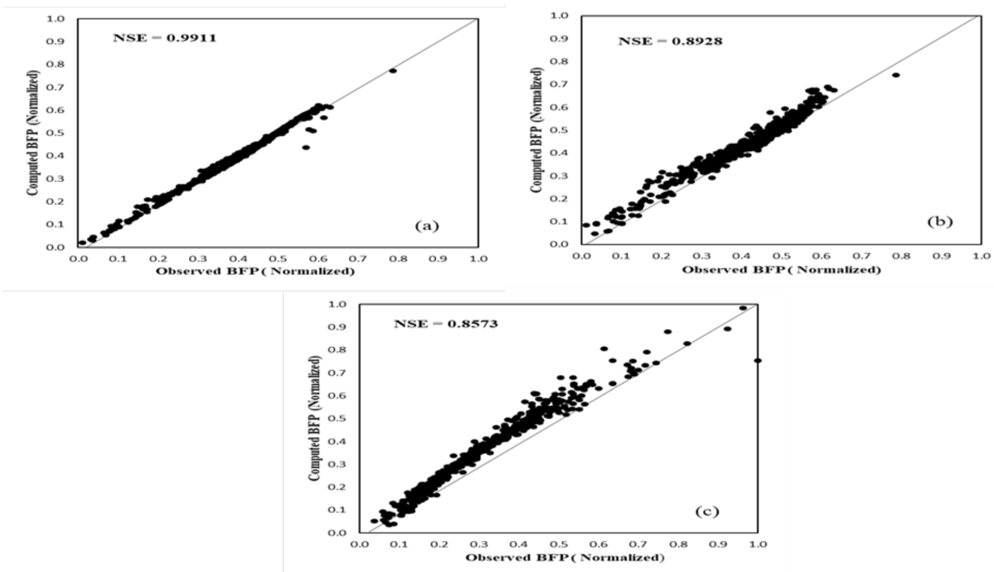
Scatter Plots for Predicted BFP in Training for Model1



Note: Scatter plots figured level of fitness between actual and predicted output. for (a) SVREANN (b) ENN (c) FFNN.

Figure 4.3

Scatter plots for predicted BFP in verification for Model1



Note: The level of fitness between estimated and predicted values of BFP for (a) SVR-EANN (b) EANN (c) FFNN.

SVR's final parameters were adjusted at 5, 0.0313, and 0.25 to match the suggested hybrid model's C, γ, and values. On the EANN search, we utilized the same settings as the FFNN search and trained it for 50 epochs with `twenty hidden and five hormone neurons.

SVR-EANN produced superior findings with "R^2", "RMSE", and "rRMSE" values of 0.991, 0.0125, and 3.15%, respectively, whereas EANN without left out sensitivity analysis produced "R^2", "RMSE", and "rRMSE" values of 0.89280, 0.04640, and "13.55%". The contrast between the SVREANN and EANN outcomes demonstrates the critical nature of choosing data and the effectiveness of the SVREANN. On the other side, the FFNN achieved lower outcomes because to its superior design and settings, achieving 0.8570, 0.06220, and 19.69% R^2, RMSE, and rRMSE, respectively. "LR" got R^2, root mean square error, and root mean square error values of 0.9180, 0.03960, and "11.43%", respectively, whereas "SVR" acquired values of 0.9680, 0.0240, and "7.69%". The "DT" (R^2 = 0.9690, RMSE = 0.02160, rRMSE 6.98%) and tree-based ensemble models, RF (R^2 = 0.9740, RMSE = 0.01980, rRMSE 6.32%), XGBoost (R^2 = 0.98, RMSE = 0.0178, rRMSE 5.91%), and GradBoost (R^2 = 0.980, RMSE = 0.01820, rRMSE 5.99%) obtained much better outcomes than other models, however they were unable to exceed the suggested SVREANN model in terms of overall performance. Table 4.3 contains the performance results of testing module for BFP prediction with benchmark algorithms' for Model1(Hussain et al., 2021). The output results listed in Table 4.3 emphasized the BFP prediction ability of all tested ML models.

Table 4.3

BFP Prediction Results with Benchmark Models for SVR-EANN

Models	R^2	RMSE	rRMSE%
FFNN	0.85730	0.06220	19.69190
EANN	0.89280	0.04640	13.55650
SVR-EANN	0.99110	0.01250	3.15130
SVR	0.96820	0.02450	7.69560
DT	0.96990	0.02160	6.98770
RF	0.97470	0.01980	6.32250
XGBOOST	0.98070	0.01780	5.91250
GRADBOOST	0.98020	0.01820	5.99490
LR	0.91850	0.03960	11.43560

Note: All models trained with 4-fold cross validation.

Using the R^2 score and rRMSE data, we can identify the best and worst prediction rates, respectively, in terms of their R^2 score and rRMSE findings. The linear-hybrid model's performance was evaluated using the rRMSE, which revealed that the hybrid model has outstanding performance and greatest accuracy, with an RRMSE value of 3.15% (excellent), compared to 13% and 19% for EANN and FFNN, respectively, in verification (Hussain et al., 2021).

Results of Model2 GAEANN

The proposed model performance evaluation made in two stages as the main methodology followed in this thesis. First, the performance of the training and testing algorithms of ANN, EANN, and GAENN models for BFP prediction, estimated. Second, the results were compared with benchmark algorithms. It is important to mention that data were normalized and divided into two sets (70%) of the row data occupied for training, and 30% for testing for all modeling algorithms involved in this section.

With respect to building the ANN model for the first stage, we set the maximum number of iterations to 1000 as well as the mean square error and the learning rate to each be 0.0001 and 0.01, respectively. The log sigmoid and purlin activation functions were shown to be the most effective activation functions for the hidden and output layers, respectively, in this study. In order to build the optimum model, having the right amount of hidden nodes is crucial, since having too many neurons can lead to overfitting, while not having enough neurons can result in inadequate information being captured (Nourani et al., 2018). According to (Abdulkadir et al., 2020), there are between (2n+1/2 + m) and (2n+1) nodes in the hidden layer for calculating the optimal number of hidden layers, which is the number of input neurons. As a result, the range of hidden neurons in a typical three-layer ANN model for BFP prediction was discovered to be (5-21). According to GA, the optimal EANN and GA-EANN model architectures were found by trial and error, depending on hyper tuning parameters. Most statistical assessment requirements may be met by an appropriate model, according to the literature (Pham et al., 2019). At the time of both calibration and verification, the model's performance was assessed using the most commonly based performance metrics, such as R^2 and RMSE. Table 4.4 displays the predicted evaluative assessment outcomes depending on model combinations. Statistically, almost all combinations matched the precision of all models.

Table 4.4

Results for the Modeling and Performance Analysis

Datasets	Algorithms	Training-stage			Testing-stage		
		R	R^2	RMSE	R	R^2	RMSE
M1	ANN	0.9942	0.9884	12.7702	0.9947	0.9895	11.1585
	EANN	0.9997	0.9995	2.7272	0.9997	0.9995	2.5168
	GAEANN	0.9998	0.9995	2.5324	0.9997	0.9995	2.5119
M2	ANN	0.9719	0.9445	27.9046	0.9698	0.9404	26.5578
	EANN	0.9567	0.9154	34.4610	0.9534	0.9090	32.8326
	GAEANN	0.9755	0.9515	26.0809	0.9739	0.9486	24.6804
M3	ANN	0.9239	0.8536	45.3236	0.8918	0.7953	49.2297
	EANN	0.9331	0.8707	42.5991	0.9303	0.8655	39.8990
	GAEANN	0.9698	0.9404	26.5578	0.9374	0.8787	37.8976

Note: datasets include input features of M1= (BMI, WHR), M2= (age, gender, BMI, WHR), and M3= (age, gender, height, weight, abdominal C).

According to the obtained results, these methods can handle models with several parameters, minimize the error function, and solve data fitting issues. R2, RMSE, and R were shown to be the most accurate simulations by GA-EANN-M1 in the overall modeling results. GA-EANN-M1 was shown to be the most accurate model, aside the difficulty in ranking the models by accuracy. However, the accuracy of all three models (ANN, EANN, and GA-EANN) was greater than 90% as shown in Figure 4.4.

Figure 4.4

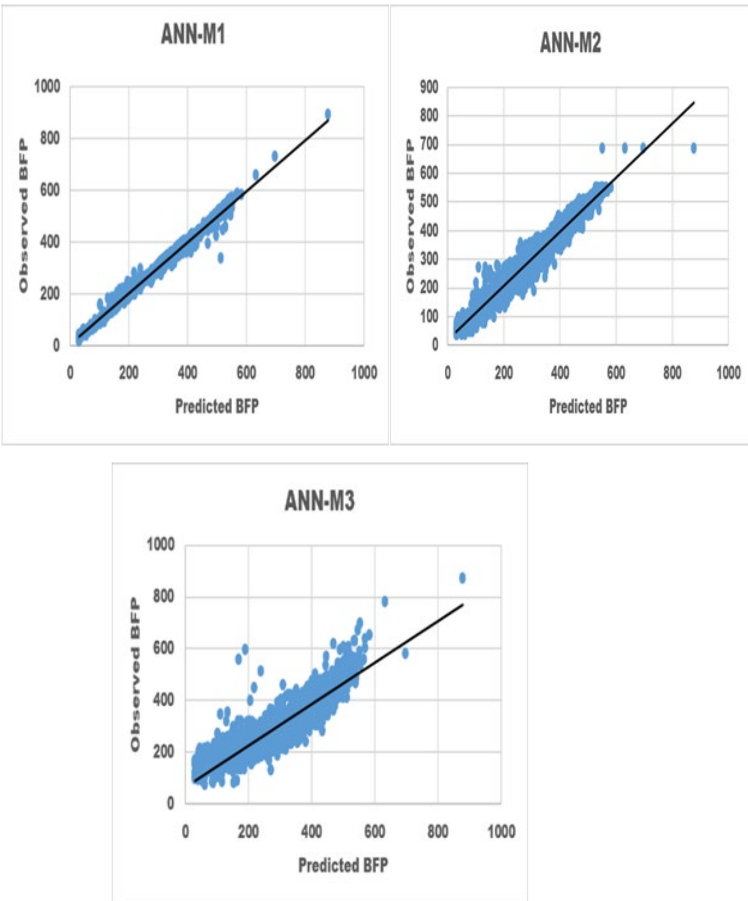
Error Plots for the Models (ANN, EANN, GAEANN)

Note: The groups of (M1(BMI, WHR), M2 (age, gender, BMI, WHR), M3 (age, gender, height, weight, abdominal C)) partitioned to 70% for training and 30% for testing.

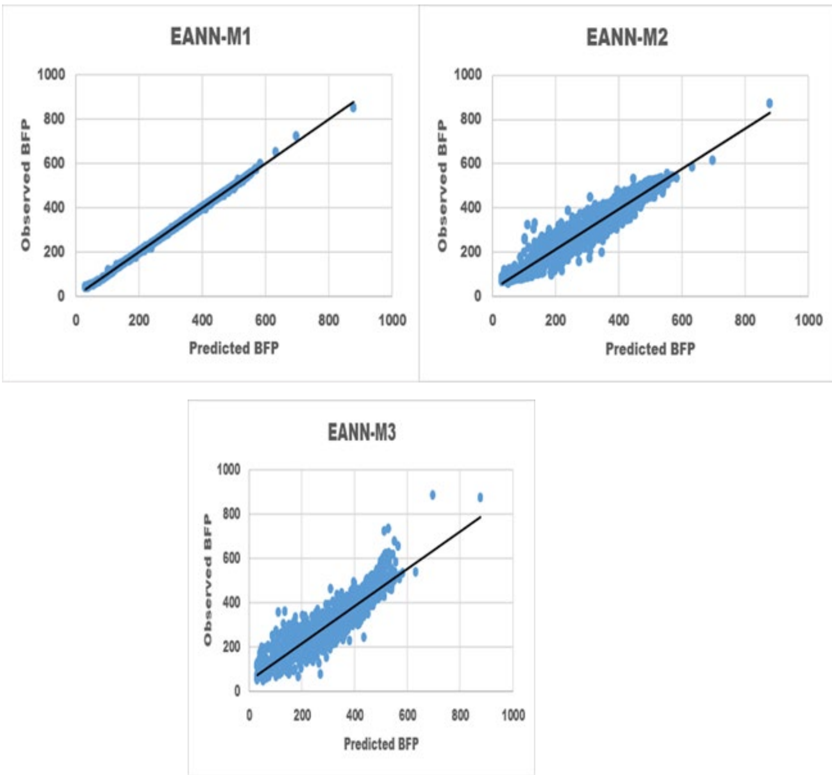
The comparison visualization of the three-model combination, on the other hand, is displayed in scatter plots (Figure 4.5). Using a scatter plot, we can see how well the measured and predicted output correspond. The scatter figure clearly shows that the GA-EANN-M1 model has a higher accuracy than the other best single models.

Figure 4.5

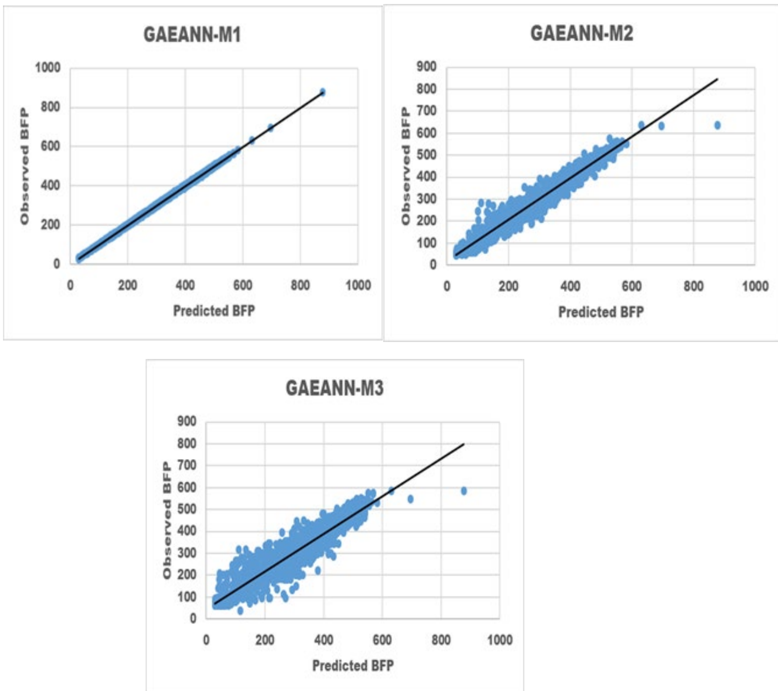
Scatter Plots for GAEANN Model2



(a)



(b)



(c)

Note: The scatter plot expresses the fitness score between actual and predicted output a) ANN (M1-M3), (b) EANN (M1-M3), and (c) GAEANN (M1-M3)

The benchmark algorithms(Chiong et al., 2021) trained in the second level of performance evaluation. The evaluation criteria of (R, R2, MAE, MSE, RMSE) used to determine the performance of the models as presented in (table 4.5). It is important to mention that the structures and parameters of benchmark algorithms used in model1 repeated in model2. The proposed EANN &GAEANN models achieved superior results over all other models for all groups. However, the best results for BFP prediction achieved through the features of M1 which included BMI and WHR. Followed by group M2, and group M3 in sequence.

Table 4.5

Results for the Proposed Models and Benchmark Algorithms

Datasets	Algorithms	Testing		
		R	**R^2**	**RMSE**
M1	ANN	0.9947270	0.98948360	11.158490
	EANN	0.9997326	0.99946500	2.5167943
	GAEANN	0.9997339	0.9994677	2.5119258
	DT	0.995486	0.97785	17.484755
	RF	0.999687	0.9877	13.022
	SVR	0.44566	0.3833	92.2577
	LR	0.9687	0.9431	28.023
	GRADBOOT	0.9996541	0.98489	14.7368
	XGBOOST	0.999452	0.99234	10.2786
M2	ANN	0.9697566	0.9404287	26.557819
	EANN	0.9533905	0.9089528	32.832602
	GAEANN	0.9739360	0.9485539	24.680365
	DT	0.854144	0.80831	51.4368
	RF	0.914456	0.89219	38.574

Table 4.5 -Continue

	SVR	0.899874	0.85252	45.1173
	LR	0.734451	0.72668	61.6227
	GRADBOOT	0.977832	0.93523	29.8266
	XGBOOST	0.92146	0.88245	40.27835
M3	ANN	0.89179	0.7953035	49.229739
	EANN	0.9303	0.8655448	39.898977
	GAEANN	0.9373	0.87869	37.8975750983
	DT	0.788663	0.733672	40.9391
	RF	0.799412	0.75055	39.6207
	SVR	0.80415	0.75785	39.0365
	LR	0.699941	0.655223	46.5799
	GRADBOOT	0.831147	0.787761	36.9879
	XGBOOST	0.79654	0.758391	38.99303

Note: The groups of (M1(BMI, WHR), M2 (age, gender, BMI, WHR), M3 (age, gender, height, weight, abdominal C)) partitioned to 70% for training and 30% for testing for training the benchmark algorithms.

The results of the algorithms showed the variant of the ML models performance for BFP model. Although EANN & GAENN have approximately same performance in term of R^2, But the consistent model of GAENN makes it superior over all other models. For example, the modeling results performance shown for group M2, registered values of (RMSE=24.680365 for GAENN, RMSE =32.832602 for EANN, and RMSE =26.557819 for ANN) which means that the ANN model performed better that the EANN with small ratio, while the GAENN continued in outperforming all models. The proposed GAENN outperformed all benchmark models listed in the Table 4.5.

The average execution time estimated in “seconds” of the models registered (DT:0.01, GBoost:0.05386, LR:0.006002, RF:2.577184, SVR:0.1663525, XGBoost:1.707491, ANN:0.001, EANN:0.046033333, GEANN:0.057166667).

CHAPTER V

Discussion

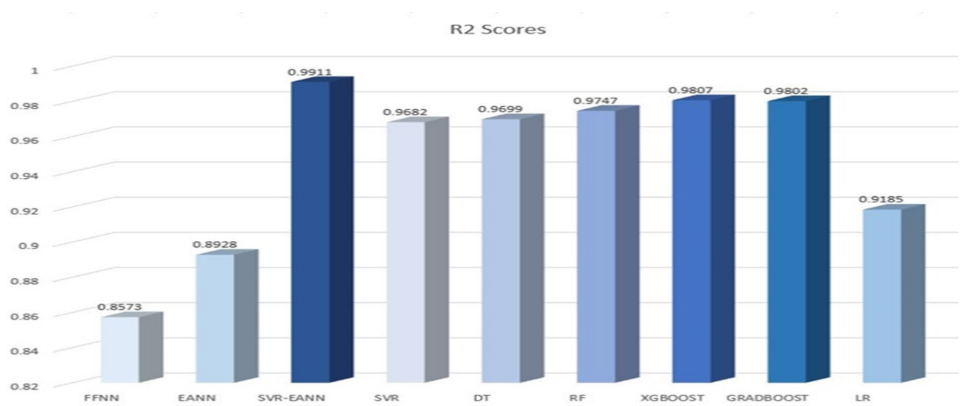
This chapter presents the discussion of the study findings in comparison to the studies in the literature. we'll go over in depth how machine learning modelling approaches are used to predict BFP and discuss the importance of the research findings.

Discussion of Model1: SVR-EANN

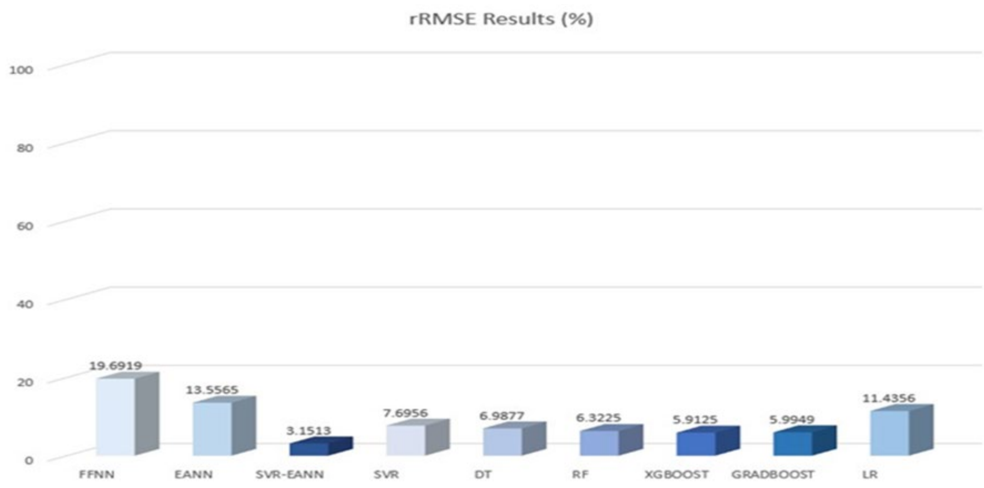
The suggested hybrid SVR-EANN model, as demonstrated by the obtained results, indicate a larger possibility of success than traditional and single-model methods. Figure 5.1 depicts a visual representation of the derived R^2 scores and rRMSE data in order to more clearly show the differences between the models compared. When compared to other models, it was viewed that utilizing the chosen data (i.e., cross validated) options for the training data produced more precious findings than using all data and raised prediction rates from 2.8 to 16%. This displayed that SVR may be employed for data selection and progression. Aside from SVR's data selection efficiency, the EANN, the genuine generation ANN, has been proven to maximize learning capacity by imposing “hormone neurons”.

Figure 5.1

Visualization of the Results Obtained for Model1



(a)



(b)

Note: The verification model of the SVR-EANN in (a) R2 scores and (b)rRMSE, run with best features identified by the sensitivity analysis model.

In fact, it is possible to predict BFP using the primary dataset and ML algorithms, although SVR-EANN beat the other models in this study by achieving superior results for all assessment measures. Real-world data analytics are compared to those found in previous studies.

Several studies using machine learning to predict BFP have been reported in the literature. Artificial neural networks (ANN) and/or SVR were used in conjunction with primary information and some “anthropometric” data such as “gender”, “BMI”, “height” , “weight” and “waist circumference” to forecast the BFP estimation model (Kupusinac et al., 2014; Ferenci and Kovács, 2018; Chiong et al., 2021a). The overall performance of the system varied between 44% to 80.43% in these investigations. Nevertheless, the mention studies contained a random selection of factors, they never managed to find the most significant parameters, which allowed them to arrive at the most accurate predictor variables

Major hybrid models in the field of BFP prediction, mostly proposed feature selection methods. They employed feature selection methodology may contrasts intelligence-based and/or statistical representations to recognize the most important features for better prediction during the application of the ML forecasting algorithm. The hybrid model proposed by (Shao, 2014), the remarkable input features evaluated by the use of MR and MARS model to be included in the feature selection in the “explanatory phase”, resulted in two features sets of five &six “anthropometric parameters”, respectively, from the overall thirteen anthropometric parameters of the trained dataset of Body fat. Their statistical model (MR and MARS) nominated parameters such as “age”, “height”, “weight”, “wrist circumference”, “forearm circumference”, “waist circumference”, “thigh circumference”, and “neck circumference”, as the most crucial parameters for BFP prediction. To put this into perspective, the MRSVR had the best RMSE of 4.6427 when BFP predictions were made usinUçar et al. (2021a) used “Spearman’s correlation coefficients” and “principal component analysis” to choose the best predictive features model from 13 distinct anthropometric subclasses. The ML based-hybrid methodologies employed MLFFNN, DT, and SVM algorithms. The hybrid model DTSVM achieved the best performance with RMSE of 0.482, involving a single significant feature named “waist circumference”.

The hybrid studies (Uçar et al., 2021; Shao, 2014) took into account the dataset obtained by Johnson (Johnson, 1996), which consisted of 252 male subjects. As a result, the issue of gender was not concerned throughout the discussion. Furthermore, despite the fact that the datasets in both experiments were significantly smaller, no research was able to attain greater results than those obtained in the current investigation in terms of prediction rates.

A number of factors have been shown to be beneficial in the estimation of BFP published studies in the literature. The studies of (Stevens et al., 2017; Deurenberg et al., 2001; Linder et al., 2003) have found that gender is the most significant factor. While, other studies (Ferenci and Kovács, 2018; Stevens et al., 2017), proposed generalized equations for men and women.

The dataset utilized in this study contains information about the gender of the participants. Further to gender, benchmark investigations on ML and BFP prediction using single or hybrid intelligent algorithms revealed that other features such as BMI (Kupusinac et al., 2014), waist circumference (Ferenci & Kovács, 2018), and body mass index (BMI) (Uçar et al., 2021a) were found to be of relative importance. This study, is the first study that has identified the most crucial parameter’s foundation in an intelligent model for predicting BFP. First and second most important characteristics in predicting BFP were the abdominal C (waist circumference) and gender, respectively, according to the SVR sensitivity analysis.

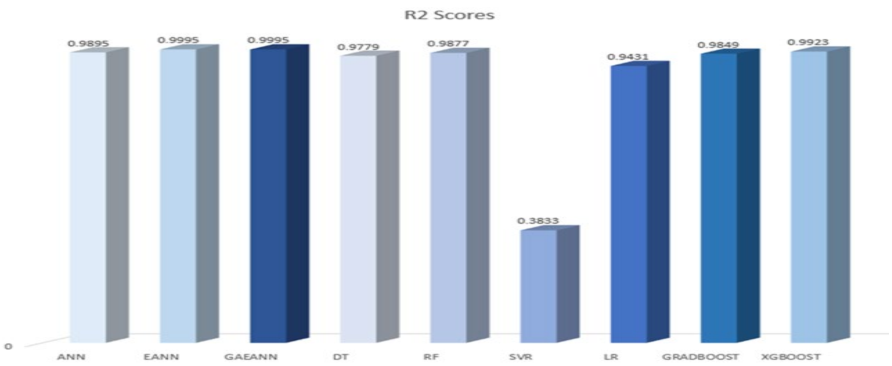
Although the total models of machine learning in this study were able to predict BFP employing the 2000-person primary dataset, the suggested hybrid model SVR-EANN with an intelligence-based feature selection algorithm, outperformed Other models. The feature selection based SVR enhanced the Hybrid model SVR-EANN performance. This yields that SVR can be employed as sensitive feature selection model in prediction models.

Discussion of Model2: GAEANN

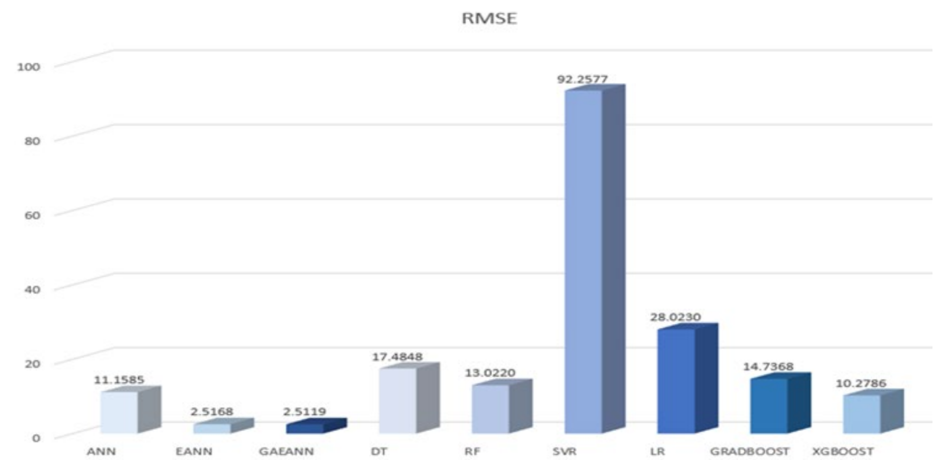
Visualization comparison of the GAEANN and benchmark simulation result depended (R^2, RMSE) for all models shown in figures (Figure 5.2 for M1, Figure 5.3 for M2, and Figure 5.4 for M3). The results supported the best performance results achieved by EANN and GAEANN over other ML models used. Although, all groups have acceptance importance in BFP Prediction and most ML-based algorithms can predict BFP with a large dataset, the GAENN outperformed others in all evaluation criteria used. Despite of their importance in Network structure optimization and performance enhancements of EANN, extensive cost of operations and execution time regularly related to the evolutionary GAs models. GAENN produced best performance results, even though, the training data were not validated (i.e., Cross fold validations not applied). The GAENN model showed high consistency and stableness in performance and even optimized the EANN performance in all sub-groups of BFP prediction data.

Figure 5.2

Visualization of the Results for M1



(a)

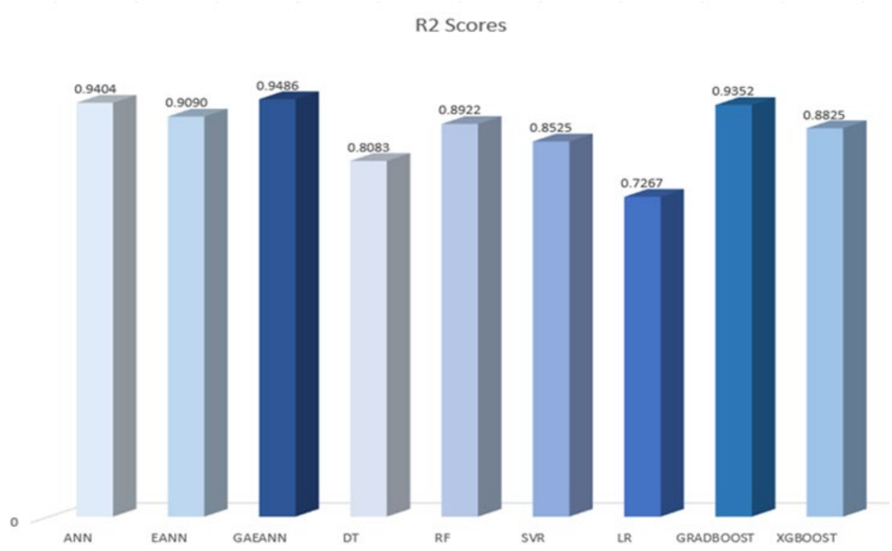


(b)

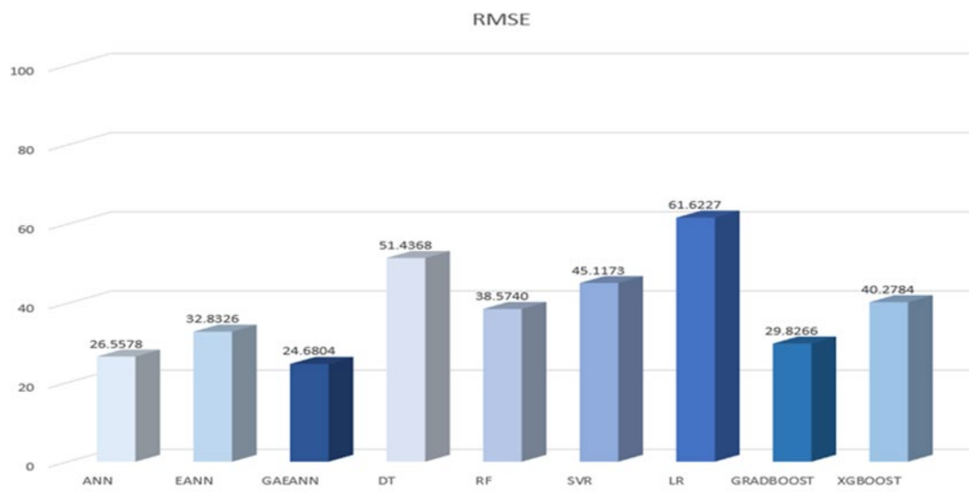
Note: The M1 group in Model2 (GAEANN) included two features (BMI, WHR). M1, (a) R2 scores, and (b) RMSE results

Figure 5.3

Visualization of the Results for M2



(a)

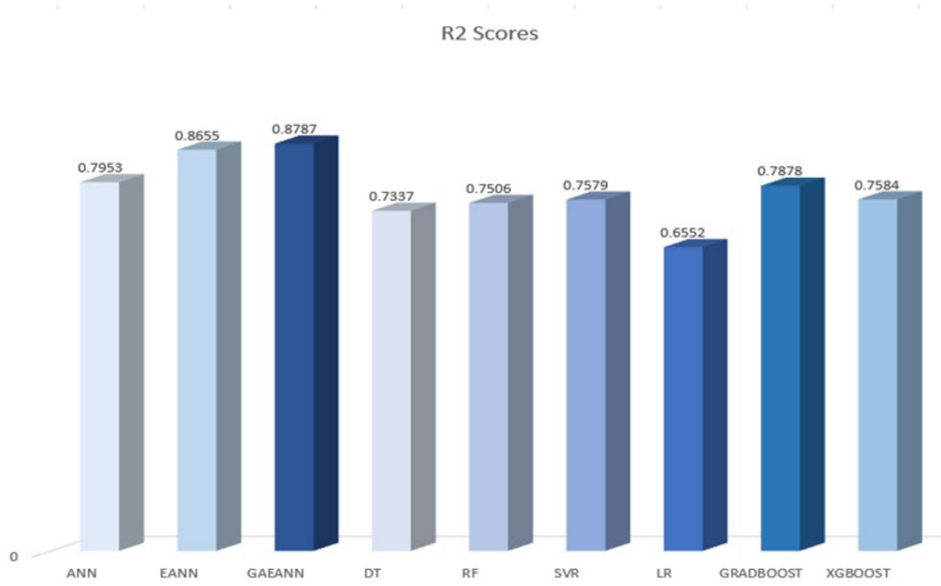


(b)

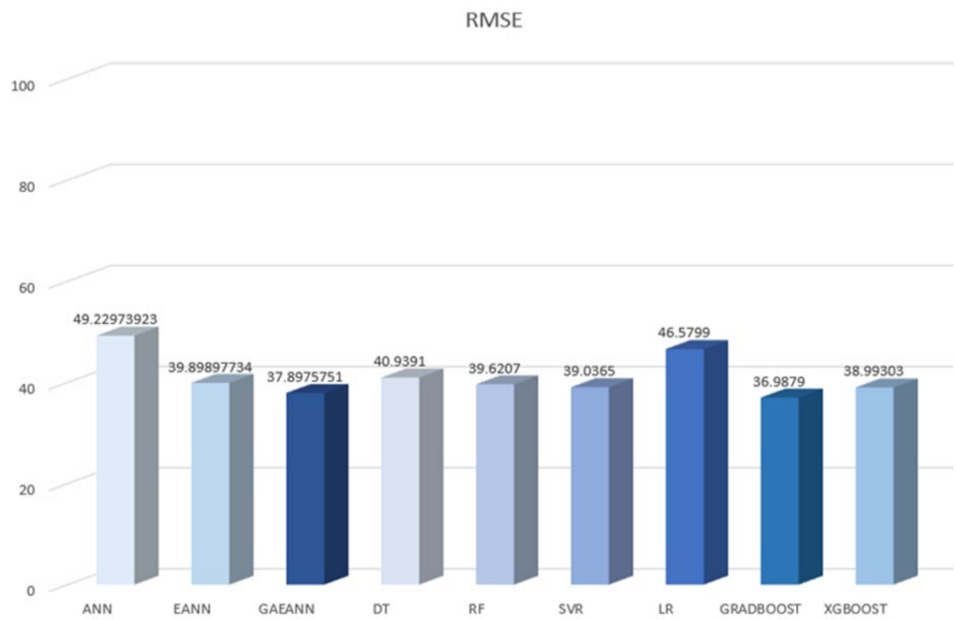
Note: The M2 group in Model2 (GAEANN) included features (age, gender, BMI, WHR). (a) R2 scores, and (b) RMSE results.

Figure 5.4

Visualization of the Results for M3



(a)



(b)

Note: The M3 group in Model2 (GAEANN) included features (age, gender, height, weight, abdominal C), (a) R2 scores, and (b) RMSE results.

The model of GA in context of BFP prediction was not discussed too much in the literature, a single study related to Gao et al. (2020) improved an “adaptive” genetic system for BFP prediction using a sample of 220 participants to improve forecast precision, generalization capability, and convergence speediness. The proposed GA model improves the parameter selection operator from (age, gender, height, weight, eight impedance R1-R8 of BIA device) while also accounting for the retention problem of a characteristic parameter (individual) with high “adaptability” in an initial evolution and degradation of the algorithm in “late evolution”. For the second method, they used a modified adaptive genetic algorithm that took use of both optimal reservations and the roulette strategy to properly estimate parameter weights. As best results in BFP prediction with GA, the adaptive evolutionary algorithm received a score of (MSE =6.27). Although, their study was the first in using GA for BFP prediction they did not discuss the important parameters that affect the prediction model, rather they concerned about the assessment of the algorithm performance. And the dataset sample limited to small size.

The hybrid model of the GAEANN proposed in this study is new and never presented in any research. Results conducted the best performance with (BMI, WHR) as input variables and the primary dataset with 2000 participants, which is a large dataset and the simulation results implied the importance of some anthropometric regressions in the prediction model of BFP. The overall evaluation metrics estimated in the present investigation. When compared to other models, it was viewed that utilizing the chosen data (i.e., cross validated) options for the training data produced more precious findings than using all data and raised prediction, future studies should consider that.

CHAPTER VI
Conclusion and Recommendations

This chapter presents the important conclusions inspired from the study and the recommendations for future work.

Conclusion

Having a consistent & cost-effective BFP estimation model on a large dataset is essential for providing experts with the information they need to take precautions against obesity. The percentage of body fat is a critical indicator of overall health. The study took into account the urgent requirement for a BFP estimation tool to help mitigate the disease caused by an excess or low percentage of fat in the body. It is either difficult or expensive to calculate one's body fat percentage accurately, so more efficient methods are required. Under this consideration, Models depend ML technologies are practically applicable in BFP estimation and provide a significant expansion of the obesity control and management industry.

The most important feature in ML model is the included dataset and the algorithm involved. For this study, a special real dataset for BFP prediction was released and the potential of EANN for modeling BFP through hybrid machine learning models was explored. Emotional neural networks and support vector regression were used to develop the first model for accurately predicting body fat percentage. Sensitivity analysis of the SVR was incorporated to fit best features for emotional artificial neural network prediction model. When associated to seven other ML models, the proposed SVR-EANN model was found to outperform all others in every evaluation metric.

The research showed that the prediction accuracy of the EANN in a hybrid model could increase, when used with feature selection based-SVR for datasets with limited BFP properties and measurements. The SVR-EANN model was used to identify the factors that influence the prediction of body fat percentage. There is a strong correlation between body fat percentage and abdominal C, according to the study's findings. The age attribute, on the other hand, has the least impact and can be omitted from prediction studies. However, the set of significant parameters criticized in this study performed well with most ML models and the primary dataset used in the study.

The second model proposed GA-EANN model with 3 different scenarios for parameter selection. The first scenario depends only secondary parameters like (BMI, WHR). The second scenario used a part of primary parameters (age, gender) that not included in the "secondary parameters namely the BMI and WHR" within the secondary parameters themselves. The third model consider the primary parameters (age, gender, height, weight, abdominal C). It has been shown that while the use of secondary parameters can provide accurate estimation with reduced computational cost, the use of the primary parameters is more reliable. The use of GA optimization algorithms in the tuning of the EANN parameters has improved the accuracy of the models.

As an added bonus, utilizing benchmark ML models as a point of reference highlighted potential differences in model regression abilities for BFP prediction studies, which could pave the way for additional BFP based ML research. it is obtained from experimental results that; the size of dataset did not affect the general performance quality of the hybrid models even with the absence of the validation model for trained data.

Finally, the current study has great impact in BFP estimation and obesity level identification models for body weight control programs. In addition, the ML based models can integrate within statistical-based tools to modify the statistical analyzation of obese data in different societies to estimate the level of obesity and risks related to them.

Recommendations for Future Works

The expected ideas for future work to inspired from the study includes the following:

- A hybrid model for predicting body fat percentage (BFP) based on optical data that should be motivated by the examination of gender effects on BFP.

- Experiments will be conducted to determine the accuracy of the proposed hybrid model in predicting body fat percentage in obese children.

References

Abba, S. I., Linh, N. T. T., Abdullahi, J., Ali, S. I. A., Pham, Q. B., Abdulkadir, R. A., Costache, R., Nam, V. T., & Anh, D. T. (2020). Hybrid machine learning ensemble techniques for modelling dissolved oxygen concentration. *IEEE Access, 8*, 57218-157237, DOI:0.1109/ACCESS.2020.3017743.

Abdulkadir, R. A., Ali, S. I. A., Abba, S. I., & Esmaili, P. (2020). Forecasting of daily rainfall at Ercan airport Northern Cyprus: A comparison of linear and non-linear models. *Desalination and Water Treatment, 177*, 297–305, DOI:10.5004/dwt.2020.25321.

Akman, M., Uçar, M. K., Uçar, Z., Uçar, K., Baraklı, B., & Bozkurt, M. R. (2021). Determination of body fat percentage by gender based with Photoplethysmography Signal using machine learning algorithm. *Innovation and Research in Biomedical Engineering, 1*, 1-18, DOI: 10.1016/j.irbm.2020.12.003.

Amirov, A., Gerget, O., Devjatyh, D., & Gazaliev, A. (2014). Medical data processing system based on Neural Network and Genetic Algorithm. *Procedia - Social and Behavioral Sciences*, *131*, 149–155, DOI: 10.1016/j.sbspro.2014.04.095.

Balasundaram, S., & Kapil, K. (2010). On lagrangian support vector regression. *Expert Systems with Applications*, *37*(12), 8784–8792, DOI: 10.1016/j.eswa.2010.06.028.

Cavus, N., Mohammed, Y. B., & Yakubu, M. N. (2021). An Artificial Intelligence-based model for prediction of parameters affecting sustainable growth of mobile banking Apps. *Sustainability*, *13*(11), DOI:10.3390/su13116206.

Chiong, R., Fan, Z., Hu, Z., & Chiong, F. (2021). Using an improved relative error support vector machine for body fat prediction. *Computer Methods and Programs in Biomedicine, 198*, 105749, DOI: 10.1016/j.cmpb.2020.105749.

DeGregory, K. W., Kuiper, P., DeSilvio, T., Pleuss, J. D., Miller, R., Roginski, J. W., Fisher, C. B., Harness, D., Viswanath, S., Heymsfield, S. B., Dungan, I., & Thomas, D. M. (2018). A review of machine learning in obesity. *Obesity Reviews, 19*(5), 668–685, DOI: 10.1111/obr.12667.

De Koning, L., Merchant, A. T., Pogue, J., & Anand, S. S. (2007). Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: meta-regression analysis of prospective studies. *European Heart Journal*, *28*(7), 850-856, DOI:10.1093/eurheartj/ehm026.

Deurenberg, P., Andreoli, A., Borg, P., Kukkonen-Harjula, K., De Lorenzo, A., Van Marken Lichtenbelt, W.D., Testolin, G., Vigano, R., &Vollaard, N. (2001). The validity of predicted body fat percentage from body mass index and from impedance in samples of five European populations. *European Journal of Clinical Nutrition*, *55*(11), 973–979, DOI: 10.1038/sj.ejcn.1601254.

Eisinga, R., Grotenhuis, M. Te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?. *International Journal of Public Health, 58*(4), 637–642. DOI:10.1007/s00038-012-0416-3.

Elkiran, G., Nourani, V., & Abba, S. I. (2019). Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *Journal of Hydrology*, *577*, DOI: 10.1016/j.jhydrol.2019.123962.

Fan, Z., Chiong, R., & Chiong, F. (2021). A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat prediction. *Applied Intelligence*, DOI: 10.1007/s10489-021-02421-3.

Fan, J. G., Kim, S. U., &Wong, V. W. S. (2017). New trends on obesity and NAFLD in Asia. *Journal of Hepatology*, 67(4), 862–873, DOI: 10.1016/j.jhep.2017.06.003.

Farquad, M. A. H., Ravi, V., & Raju, S. B. (2010). Support vector regression-based hybrid rule extraction methods for forecasting. *Expert Systems with Applications*, 37(8), 5577–5589, DOI: 10.1016/j.eswa.2010.02.055.

Ferenci, T., & Kovács, L. (2018). Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks. *Applied Soft Computing Journal*, 67, 834–839, DOI: 10.1016/j.asoc.2017.05.063.

Freedman, D. S., Dietz, W. H., Srinivasan, S. R., & Berenson, G. S. (2009). Risk factors and adult body mass index among overweight children: The bogalusa heart study. *Pediatrics*, *123*(3), 750–757, DOI: 10.1542/peds.2008-1284.

Gao, X., Xie, W., Wang, Z., Zhang, T., Chen, B., &Wang, P. (2020). Predicting human body composition using a modified adaptive genetic algorithm with a novel selection operator. *PLOS ONE*, *15*(7), 1–23, DOI: 10.1371/journal.pone.0235735.

Ghali, U. M., Usman, A. G., Chellube, Z. M., Degm, M. A. A., Hoti, K., Umar, H., & Abba, S. I. (2020). Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: An ensemble machine learning approach. *SN Applied Sciences*, *2*(11), DOI:10.1007/s42452-020-03690-2.

Gholami, A., Bonakdari, H., Samui, P., Mohammadian, M., & Gharabaghi, B. (2019). Predicting stable alluvial channel profiles using emotional artificial neural networks. *Applied Soft Computing Journal, 78*, 420–437, DOI: 10.1016/j.asoc.2019.03.003

Harty, P. S., Sieglinger, B., Heymsfield, S. B., Shepherd, J. A., Bruner, D., Stratton, M. T., & Tinsley, G. M. (2020). Novel body fat estimation using machine learning and 3-dimensional optical imaging. *European Journal of Clinical Nutrition*, *74*(5), 842–845, DOI: 10.1038/s41430-020-0603-x.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. *National Taiwan University*, Taipei Taiwan. Available at: http://www.csie.ntu.edu.tw/~cjlin/papers.html. Accessed on 14 October 2021.

Hussain, S. A., Cavus, N., & Sekeroglu, B. (2021). Hybrid machine learning model for body fat percentage prediction based on support vector regression and emotional artificial neural networks. *Applied Sciences, 11*(21), 9797. DOI:10.3390/app11219797.

Johnson, R. W. (1996). Fitting percentage of body fat to simple body measurements. *Journal of Statistics Education, 4*(1), DOI: 10.1080/10691898.1996.11910505.

Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2011). Application of genetic algorithm optimized neural network connection weights for medical diagnosis of Pima Indians diabetes. *International Journal on Soft Computing*, *2*(2), 15-23. DOI: 10.5121/ijsc.2011.2202.

Khashman, A. (2008). A modified backpropagation learning algorithm with added emotional coefficients. *IEEE Transactions on Neural Networks*, *19*(11), 1896–1909. DOI:10.1109/TNN.2008.2002913.

Khashman, A. (2010). Modeling cognitive and emotional processes: A novel neural network architecture. *Neural Networks*, *23*(10), 1155–1163. DOI: 10.1016/j.neunet.2010.07.004.

Kupusinac, A., Stokić, E., & Doroslovački, R. (2014). Predicting body fat percentage based on gender, age and BMI by using artificial neural networks. *Computer Methods and Programs in Biomedicine*, *113*(2), 610–619, DOI: 10.1016/j.cmpb.2013.10.013.

Leung, F. H. F., Lam, H. K., Ling, S. H., & Tam, P. K. S. (2003). Tuning of the structure and parameters of a neural network using an improved genetic algorithm. *IEEE Transactions on Neural Networks*, *14*(1), 79-88, DOI: 10.1109/TNN.2002.804317.

Linder, R., Mohamed, E. I., De Lorenzo, A., & Pöppl, S. J. (2003). The capabilities of artificial neural networks in body composition research. *Acta Diabetologica, 40*(1), 9–14, DOI: 10.1007/s00592-003-0018-x.

Liu, Y., Zhang, Q., Song, L., & Chen, Y. (2019). Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction. *Computers and Electronics in Agriculture*, *165*, 104964, DOI: 10.1016/j.compag.2019.104964.

Lu, Y., Hahn, J. K., & Zhang, X. (2020). 3D shape-based body composition inference model using a Bayesian network. *IEEE Journal of Biomedical and Health Informatics, 24*(1), 205–213, DOI: 10.1109/JBHI.2019.2903190.

Lu, Y., McQuade, S., & Hahn, J. K. (2018). 3D shape-based body composition prediction model using machine learning. *In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3999–4002). Honolulu, HI, USA, DOI: 10.1109/EMBC.2018.8513261.

Lotfi, E., & Akbarzadeh-T., M. R. (2014). Practical emotional neural networks. *Neural Networks, 59*, 61–72, DOI: 10.1016/j.neunet.2014.06.012.

Lotfi, E., Khosravi, A., Akbarzadeh, M. R., & Nahavandi, S. (2014). Wind power forecasting using emotional neural networks. *In Proceedings of IEEE*

International Conference on Systems, Man and Cybernetics (pp. 311–316). San Diego, CA, USA, DOI:10.1109/SMC.2014.6973926.

Mantzaris, D., Anastassopoulos, G., & Adamopoulos, A. (2011). Genetic algorithm pruning of probabilistic neural networks in medical disease estimation. *Neural Networks, 24*(8), 831–835, DOI: 10.1016/j.neunet.2011.06.003.

Muhammad, A., Gaya, M. S., Aliyu, R., Umar, I. D., Yusuf, L. A., Ali, M. U., & Khairi, M. T. M. (2018). Forecasting of global solar radiation using ANFIS and ARMAX techniques. *In Proceeding of IOP Conference Series: Materials Science and Engineering* (Vol. 303, No. 1, p. 012016). IOP Publishing, DOI:10.1088/1757-899X/303/1/012016.

Naik, D. L. (2021). A novel sensitivity-based method for feature selection. *Journal of Big Data*, *8*(1), 1-16. DOI: 10.1186/s40537-021-00515-w.

Nourani, V. (2017). An Emotional ANN (EANN) approach to modeling rainfall-runoff process. *Journal of Hydrology, 544*, 267–277, DOI: 10.1016/j.jhydrol.2016.11.033.

Nourani, V., Elkiran, G., & Abba, S. I. (2018). Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach. *Water Science and Technology*, 78(10), 2064-2076, DOI:10.2166/wst.2018.477.

Nourani, V., Elkiran, G., & Abdullahi, J. (2019). Multi-station artificial intelligence-based ensemble modelling of reference evapotranspiration using pan evaporation measurements. *Journal of Hydrology*, *577*, 123958, DOI: 10.1016/j.jhydrol.2019.123958.

Nourani, V., Gökçekuş, H., & Umar, I.K. (2020a). Artificial intelligence-based ensemble model for prediction of vehicular traffic noise. *Environmental Research, 180*, 108852, DOI: 10.1016/j.envres.2019.108852.

Nourani, V., Gökçekuş, H., Umar, I. K., & Najafi, H. (2020b). An emotional artificial neural network for prediction of vehicular traffic noise. *Science of the Total Environment*, *707*, 136134, DOI: 10.1016/j.scitotenv.2019.136134.

Ozsahin, I., Sekeroglu, B., & Mok, G. S. (2019). The use of back propagation neural networks and 18F-Florbetapir PET for early detection of

Alzheimer's disease using Alzheimer's Disease Neuroimaging Initiative database. *PLOS ONE*, *14*(12), DOI: 10.1371/journal.pone.0226577.

Pham, Q. B., Abba, S. I., Usman, A. G., Linh, N. T. T., Gupta, V., Malik, A., Costache, R., Vo, N. D., & Tri, D. Q. (2019). Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. *Water Resources Management, 33*(15), 5067–5087, DOI:10.1007/s11269-019-02408-3.

Rabehi, A., Guermoui, M., & Lalmi, D. (2020). Hybrid models for global solar radiation prediction: A case study. *International Journal of Ambient Energy*, *41*(1), 31-40, DOI:10.1080/01430750.2018.1443498.

Roshni, T. (2020). Neural network modelling for groundwater-level forecasting in coastal aquifer. *Neural Computing and Applications*, 1-18, DOI:10.1007/s00521-020-04722-z.

Safaei, M., Sundararajan, E.A., Driss, M., Boulila, W., & Shapi, A. (2021). A systematic literature review on obesity: Understanding the causes and consequences of obesity and reviewing various machine learning approaches used to predict obesity. *Computers in Biology and Medicine*, *136*, DOI: 10.1016/j.compbiomed.2021.104754.

Sarría, A., García-Llop, L.A., Moreno, L.A., Fleta, J., Morellón, M.P., & Bueno, M. (1998). Skinfold thickness measurements are better predictors of body fat percentage than body mass index in male Spanish children and adolescents. *European Journal of Clinical Nutrition, 52*(8), 573–576, DOI: 10.1038/sj.ejcn.1600606.

Selin, A. G. U., & Abba, I. S. I. (2020). A novel multi - model data - driven Ensemble technique for the prediction of retention factor in HPLC method development. *Chromatographia*, 0123456789, DOI:10.1007/s10337-020-03912-0.

Sekeroglu, B., & Tuncal, K. (2021). Prediction of cancer incidence rates for the European continent using machine learning models. *Health Informatics Journal, 27*(1), DOI:10.1177/1460458220983878.

Shao, Y. E. (2014). Body fat percentage prediction using intelligent hybrid approaches. *The Scientific World Journal*, *1*, 1–8, DOI: 10.1155/2014/383910.

Shin, S. C., Lee, J., Choe, S., Yang, H. I., Min, J., Ahn, K. Y., Jeon, J. Y., & Kang, H. G. (2019). Dry electrode-based body fat estimation system with anthropometric data for use in a wearable device. *Sensors (Switzerland)*, *19*(9), 1–17, DOI: 10.3390/s19092177.

Stevens, J., Truesdale, K. P., Cai, J., Ou, F. S., Reynolds, K. R., & Heymsfield, S. B. (2017). Nationally representative equations that include resistance and reactance for the prediction of percent body fat in Americans. *International Journal of Obesity*, *41*(11), 1669–1675, DOI: 10.1038/ijo.2017.167.

Swainson, M. G., Batterham, A. M., Tsakirides, C., Rutherford, Z. H., & Hind, K. (2017). Prediction of whole-body fat percentage and visceral adipose tissue mass from five anthropometric variables. *PLOS ONE*, *12*(5), 1–12, DOI: 10.1371/journal.pone.0177175.

Uçar, M. K., Uçar, Z., Köksal, F., & Daldal, N. (2021a). Estimation of body fat percentage using hybrid machine learning algorithms Measurement. *Journal of the International Measurement Confederation,167*, DOI: 10.1016/j.measurement.2020.108173.

Uçar, M. K., Uçar, Z., Uçar, K., Akman, M., & Bozkurt, M. R. (2021b). Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence. *Biomedical Signal Processing and Control, 68*, DOI: 10.1016/j.bspc.2021.102650.

Usman, A. G., Işik, S., Abba, S. I., & Meriçli, F. (2020). Artificial intelligence–based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method. *Turkish Journal of Chemistry, 44*(5), 1339–1351, DOI:10.3906/kim-2003-6.

Vapnik, V. N. (1998). Statistical Learning Theory. Wiley: New York, NY, USA.

Wang, C., Hou, X. H., Zhang, M. L., Bao, Y. Q., Zou, Y. H., ZHong, W. H., Xiang, K. S., & Jia, W. P. (2010). Comparison of body mass index with body fat percentage in the evaluation of obesity in Chinese. *Biomedical and Environmental Sciences*, *23*(3),173–179, DOI: 10.1016/S0895-3988(10)60049-9.

Wang, Q., Xue, W., Zhang, X., Jin, F., & Hahn, J. (2021). Pixel-wise body composition prediction with a multi-task conditional generative adversarial network. *Journal of Biomedical Informatics*, *120*, 103866, DOI: 10.1016/j.jbi.2021.103866.

Xu, Y., & Wang, L. (2012). A weighted twin support vector regression. *Knowledge-Based Systems, 33*, 92–101, DOI: 0.1016/j.knosys.2012.03.013.

APPENDICES

APPENDIX A

Ethical Committee Approval Letter

16.11.2021

Dear Solaf Ali Hussain

Your application titled **"Artificial Intelligent based Hybrid Models for Body Fat Prediction"** with the application number NEU/AS/2021/134 has been evaluated by the Scientific Research Ethics Committee and granted approval. You can start your research on the condition that you will abide by the information provided in your application form.

Assoc. Prof. Dr. Direnç Kanol

Rapporteur of theScientificResearchEthicsCommittee

Direnc Kanol

Note: If you need to provide an official letter to an institution with the signature of the Head of NEU Scientific Research Ethics Committee, please apply to the secretariat of the ethics committee by showing this document.

APPENDIX B
Permissions Regarding the Use of Dataset

Dec.24 .2020
Solaf Ali Hussain
Iraq, KRG, Sulaimania
Email:solafrecovery@gmail.com

Permission to use data for PhD thesis

Dear Doctor Mohammed and Baxshin hospital manager,

My name is (Solaf Ali Hussain), I am PhD student at Computer Information System, Near East University. I am in the process of conducting a research about (Body fat prediction using Artificial intelligent models). I need patients' data about body composition analysis data estimated by BIA device of type (**ACCUNIQ BC380**) located in Baxshin hospital under the supervision of Doctor Mohammed Jabari specialist in food and nutrition.

The data will be used in computerized algorithms to get analyzed outputs related to my thesis topic entitled "**Artificial Intelligent based Hybrid Models for Body Fat Prediction**". And the permission includes all future publish researches related to the thesis study. The data will not be used in other researches not related to the thesis.

The needed data description includes body composition analysis for adults of age>18 that computed by the **ACCUNIQ BC380** device for more than 2000 person.

Permission includes non-exclusive world rights to use the data and will not limit any future publications-including editions related to the thesis topic.

I would like greatly appreciate your consent to my request.

Three copies of this request has been provided for your records. If you agree with the terms described above, please sign the release form below.

Sincerely,

Solaf Ali Hussain

Permission granted for the use of the data as described above:

Agreed to :Solaf Ali Hussain to Use the (ACCUNIQ BC380) device data for the thesis researches as described above

Dr. Name : Mohammed I.M. Gubari, **Job Title:** Clinical Dietian (PhD, MSC),

signature 24. 12 . 2020

Agreed to: Solaf Ali Hussain to Use the (ACCUNIQ BC380) device data for the thesis researches as described above

Dr.Name : REBWAR H. KH **Job Title:** G. Manager Baxshin HOSP.

Signature:

24.12-20

APPENDIX C

Turnitin Similarity Report

PhD-THESIS

ORIGINALITY REPORT

11%	6%	10%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	www.ncbi.nlm.nih.gov Internet Source	1%
3	M.A.H. Farquad, V. Ravi, S. Bapi Raju. "Support vector regression based hybrid rule extraction methods for forecasting", Expert Systems with Applications, 2010 Publication	1%
4	scholarspace.library.gwu.edu Internet Source	1%
5	S. I. Haruna, Salim Idris Malami, Musa Adamu, A. G. Usman, AIB. Farouk, Shaban Ismael Albrka Ali, S. I. Abba. "Compressive Strength of Self-Compacting Concrete Modified with Rice Husk Ash and Calcium Carbide Waste Modeling: A Feasibility of Emerging Emotional Intelligent Model (EANN) Versus Traditional FFNN", Arabian Journal for Science and Engineering, 2021 Publication	1%

6	Tamás Ferenci, Levente Kovács. "Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks", Applied Soft Computing, 2018 Publication	1 %
7	Muhammed Kürşad Uçar, Zeliha Uçar, Fatih Köksal, Nihat Daldal. "Estimation of Body Fat Percentage using Hybrid Machine Learning Algorithms", Measurement, 2020 Publication	<1 %
8	www.pubfacts.com Internet Source	<1 %
9	S.I. Abba, R.A. Abdulkadir, Saad Sh. Sammen, Quoc Bao Pham, A.A. Lawan, Parvaneh Esmaili, Anurag Malik, Nadhir Al-Ansari. "Integrating feature extraction approaches with hybrid emotional neural networks for water quality index modeling", Applied Soft Computing, 2021 Publication	<1 %
10	www.springerprofessional.de Internet Source	<1 %
11	Solaf A. Hussain, Nadire Cavus, Boran Sekeroglu. "Hybrid Machine Learning Model for Body Fat Percentage Prediction Based on Support Vector Regression and Emotional	<1 %

Artificial Neural Networks", Applied Sciences, 2021
Publication

12	Ehsan Lotfi, Abbas Khosravi, M-R. Akbarzadeh-T, Saeid Nahavandi. "Wind power forecasting using emotional neural networks", 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2014 Publication	<1%
13	Dayakar L. Naik, Ravi kiran. "A novel sensitivity-based method for feature selection", Journal of Big Data, 2021 Publication	<1%
14	link.springer.com Internet Source	<1%
15	Xu, Y.. "A weighted twin support vector regression", Knowledge-Based Systems, 201209 Publication	<1%
16	Zongwen Fan, Raymond Chiong, Fabian Chiong. "A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat prediction", Applied Intelligence, 2021 Publication	<1%
17	Yao Lu, Scott McQuade, James K. Hahn. "3D Shape-based Body Composition Prediction Model Using Machine Learning", 2018 40th	<1%

Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018
Publication

18	ebin.pub Internet Source	<1 %
19	Muhammed Kürşad Uçar, Zeliha Uçar, Kübra Uçar, Mehmet Akman, Mehmet Recep Bozkurt. "Determination of body fat percentage by electrocardiography signal with gender based artificial intelligence", Biomedical Signal Processing and Control, 2021 Publication	<1 %
20	academic.oup.com Internet Source	<1 %
21	Zhihao Jin, Yiqun Ma, Lingzhi Chu, Yang Liu, Robert Dubrow, Kai Chen. "Predicting spatiotemporally-resolved mean air temperature over Sweden from satellite data using an ensemble model", Environmental Research, 2022 Publication	<1 %
22	www.ijmrhs.com Internet Source	<1 %
23	Yao Lu, James K Hahn, Xiaoke Zhang. "3D Shape-Based Body Composition Inference	<1 %

Model Using a Bayesian Network", IEEE Journal of Biomedical and Health Informatics, 2020
Publication

24	doaj.org Internet Source	<1%
25	mdpi-res.com Internet Source	<1%
26	scholarworks.waldenu.edu Internet Source	<1%
27	Azadeh Gholami, Hossein Bonakdari, Pijush Samui, Majid Mohammadian, Bahram Gharabaghi. "Predicting stable alluvial channel profiles using emotional artificial neural networks", Applied Soft Computing, 2019 Publication	<1%
28	scholar.sun.ac.za Internet Source	<1%
29	www.acnut.com Internet Source	<1%
30	Submitted to King's College Student Paper	<1%
31	Submitted to University of Central England in Birmingham Student Paper	<1%

32	In-Ho Yang, Myoung-Souk Yeo, Kwang-Woo Kim. "Application of artificial neural network to predict the optimal start time for heating system in building", Energy Conversion and Management, 2003 Publication	<1 %
33	Vahid Nourani, Amir Molajou, Selin Uzelaltinbulat, Fahreddin Sadikoglu. "Emotional artificial neural networks (EANNs) for multi-step ahead prediction of monthly precipitation; case study: northern Cyprus", Theoretical and Applied Climatology, 2019 Publication	<1 %
34	www.science.gov Internet Source	<1 %
35	Amir Molajou, Vahid Nourani, Abbas Afshar, Mina Khosravi, Adam Brysiewicz. "Optimal Design and Feature Selection by Genetic Algorithm for Emotional Artificial Neural Network (EANN) in Rainfall-Runoff Modeling", Water Resources Management, 2021 Publication	<1 %
36	One-Ki Daniel Lee, Ramakrishna Ayyagari, Farzaneh Nasirian, Mohsen Ahmadian. "Role of interaction quality and trust in use of AI-based voice-assistant systems", Journal of Systems and Information Technology, 2021 Publication	<1 %

37	Submitted to University of Newcastle Student Paper	<1 %
38	Vahid Nourani, Hüseyin Gökçekuş, Ibrahim Khalil Umar, Hessam Najafi. "An emotional artificial neural network for prediction of vehicular traffic noise", Science of The Total Environment, 2020 Publication	<1 %
39	Ya-min Li, Li-juan Cui, Kai Li. "Creating Diversity in Ensembles using Clustering Method from Libraries of Models", 2006 International Conference on Machine Learning and Cybernetics, 2006 Publication	<1 %
40	embs.papercept.net Internet Source	<1 %
41	mafiadoc.com Internet Source	<1 %
42	www.researchgate.net Internet Source	<1 %
43	www.temjournal.com Internet Source	<1 %
44	Patrick S. Harty, Breck Sieglinger, Steven B. Heymsfield, John A. Shepherd, David Bruner, Matthew T. Stratton, Grant M. Tinsley. "Novel body fat estimation using machine learning	<1 %

and 3-dimensional optical imaging", European Journal of Clinical Nutrition, 2020
Publication

45 Schimanski, Marcher, Monizza, Matt. "The Last Planner® System and Building Information Modeling in Construction Execution: From an Integrative Review to a Conceptual Model for Integration", Applied Sciences, 2020 <1%
Publication

46 Raymond Chiong, Zongwen Fan, Zhongyi Hu, Fabian Chiong. "Using an improved relative error support vector machine for body fat prediction", Computer Methods and Programs in Biomedicine, 2021 <1%
Publication

Exclude quotes	On	Exclude matches	< 5 words
Exclude bibliography	On		

APPENDIX D
Curriculum Vitae

PERSONAL INFORMATION

Name and Surname: Solaf A. Hussain
Date of Birth : September 3,1972
Title : Lecturer
Education Status : Master
Institution : University of Sulaimani
Language Skills : Arabic and English

EDUCATION

Degree	Department/Program	University	Year
Bachelor's Degree	Applied Mathematics	University of Technology	1990-1994
High Diploma	Computer Assistant learning	University of Technology	1995-1997
Master's Degree	Computer Science	University of Sulaimani	2007-2009
PhD student	Computer Information Systems	Near East University	2017-2022

FOREIGN LANGUAGES

- Fluent spoken and written English

PUBLICATIONS IN INTERNATIONAL REFEREED JOURNALS (IN COVERAGE OF SSCI/SCI-EXPANDED, AHCI and ESCI):

- Hussain, S. A., Cavus, N., & Sekeroglu, B. (2021). Hybrid machine learning model for body fat percentage prediction based on support vector regression and emotional artificial neural networks. *Applied Sciences, 11*(21), 9797. DOI:10.3390/app11219797.

PUBLICATIONS IN INTERNATIONAL REFEREED JOURNALS (IN COVERAGE OF SCOPUS):

- Hussain, S. A., & Baban,M. H. M. (2015). Fuzzy rule base system for student grade classification in online test. *International Journal of Scientific & Engineering Research*, *6*(8), 2229-5518.

PUBLICATIONS IN INTERNATIONAL REFEREED JOURNALS

- Kaki, M. N. M., & Hussain, S. A. (2014). Conceptions of transitive maps in topological spaces. *International Journal of Electronics Communication and Computer Engineering, 5*(1), 2278–4209.

THESIS

Master

- "Modelling System Using Adaptive Fuzzy and Neural Network"

COURSES GIVEN AT THE UNDERGRADUATE LEVEL

Academic Year	Course Name	Weekly Hour		Number of Students
		Theoretical	Practice	
2009-2010 2010-2011 2011-2012	AI	2	2	80
	Programming	2	2	120
2012-2013 2013-2014 2014-2015 2015-2016 2016-2017	Discrete Mathematics	2	-	120
	Operating System	2	2	120