NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF BIOMEDICAL ENGINEERING

BIBLIOMETRIC STUDY ON MACHINE LEARNING APPLICATION TO METABOLIC SYNDROME

M.Sc. THESIS

TATENDA MARATA

Nicosia

January, 2023

TATENDA MARATA

BIBLIOMETRIC STUDY ON MACHINE LEARNING APPLICATION TO METABOLIC SYNDROME

MASTER THESIS

2022

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF BIOMEDICAL ENGINEERING

BIBLIOMETRIC STUDY ON MACHINE LEARNING APPLICATIONS TO METABOLIC SYNDROME

M.Sc. THESIS

TATENDA MARATA

Supervisor

Prof Dr. TERIN ADALI

Nicosia

January, 2023

Approval

We certify that we have read the thesis submitted by TATENDA MARATA titled **"BIBLIOMETRIC STUDY ON MACHINE LEARNING APPLICATIONS TO METABOLIC SYNDROME"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Examining Committee	Name-Surname	Signature
Head of the Committee:	Assoc. Prof. Dr. Boran Şekeroğlu	……………………
Committee Member*:	Assoc. Prof. Dr. Melis S. Özdenefe	……………………
Supervisor:	Prof. Dr. Terin Adalı	……………………

Approved by the Head of the Department

15/03/2023

Assoc. Prof. Dr. Süleyman Aşır
Title, Name-Surname
Head of Department

Approved by the Institute of Graduate Studies

…../…../20…

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.

Tatenda Marata

…../01/2023

Day/Month/Year

Acknowledgments

Firstly, I would like to thank God, for enabling me to do this thesis project. There are days where it was difficult to continue, but He continuously gave me the strength to push through.

I would like to thank and appreciate my Thesis Supervisor Prof Dr. Terin Adali, for the job well done. I learnt a lot from her and my level of understanding is no longer mediocre.

Last but definitely not least, I would like to thank my family and friends for their continuous support on this academic journey. Special thanks to my mother and sister for motivating me along the way as well as their unending prayers for my life.

Tatenda Marata

Abstract

Bibliometric Study on Machine Learning Applications to Metabolic Syndrome

Marata, Tatenda
MSc, Department of Biomedical Engineering
January,2023, 51 pages

A disease that has encamped a quarter of the world, drastically increasing the mortality rate. Metabolic Syndrome has caught the attention of researchers, as it is a multifactorial disorder. The use of machine learning in medicine has made remarkable progress for numerous conditions that are dependable on early detection. This study aims to evaluate different applications of machine learning algorithms in detecting metabolic syndrome, the bibliometric aspect being the analysis and survey in order to provide recommendations for the researcher for further research in MetS and machine learning. The bibliometric analysis has been applied between 2011 to 2022 for machine learning applications for metabolic syndrome disease. In this study, different databases have been used, which include SCOPUS, PUBMED, GOOGLE SCHOLAR, IEEC, SPRINGER LINK, and SCIENCE DIRECT. Bibliometric analysis used articles and journals from different subject areas like medicine, engineering, computer science, and mathematics. The data was gathered by a filtered search made on SCOPUS, using specific keywords and VOSviewer was used as the visualization software tool. This analysis found the United States to be the global dominating research country with the highest number of publications which is 48 followed by China with 35. Documents by year were highest in 2022 with 54 published papers.

Keywords: Bibliometric study, artificial intelligence, machine learning, metabolic syndrome

ÖZET

Matabolik Sendrom konusunda Makine Öğrenme uygulamalarının Bibliyometrik Çalışmaları

Marata, Tatenda
Yüksek Lisans, Biyomedikal Mühendisliği Bölümü
Ocak 2023, 51 sayfa

Dünyanın dörtte birini etkileyen hastalık olan metabolic sendrom, hızlı bir şekilde artmaktadır. Çoklufaktör bir rahatsızlık olarak araştırmacıların dikkatini çekmiştir. Erken tanı çalışmalarında Makine öğrenme kullanımı dikkat çekici gelişmeler sağlamıştır. Bu çalışmanın amacı, metabolic sendrom hastalığının erken tanısında, farklı makine öğrenme algoritmalarının bibliometric çalışma uygulayarak değerlendirmesini yapmaktır. Bu çalışma sonucunda araştırmacılara sağlanan verilerin ileri araştırmalara ışık tutması hedeflenmektedir. 2011-2022 yılları arasında yapılan metabolic hastalığına makine öğrenme SCOPUS, PUBMED, GOOGLE SCHOLAR, IEEC, SPRINGER LINK ve SCIENCE DIRECT kullanılmıştır. Tıp, mühendislik, bilgisayar mühendisliği, ve matematik disiplinlerini içeren dergilerden elde edilen makalelere uygulanan bibliometric analiz konuya özel anahtar kelimeler kullanılarak VOSviewer ile filtrelenmiştir. Amerika, dünya çapında 48 yayın ile birinci sırada olup 35 makale ile Çin ikinci sıradadır. Yıllara göre makale sayısında ise 2022 yılında 54 makale ile hızlı bir ivme gözlemlenmiştir.

Anahtar Kelimeler: Bibliyomterik çalışma, yapay zeka, Makine öğrenme, Metabolik Semdrom

Table Of Contents

CHAPTER IV

CHAPTER V

List of Tables

List of Figures

List of Abbreviations

MetS:	Metabolic syndrome
T2D:	Type 2 diabetes
ML:	Machine learning
CVD:	Cardiovascular diseases
NCDs:	Non-communicable diseases
WHO:	World Health Organisation
T1DM:	Type 1 diabetes
WC:	Waist circumference
BMI:	Body mass index
VAT:	Visceral Adipose Tissue
SBP:	Systolic blood pressure
TG:	Triglycerides
HDL-C:	High density lipoprotein cholesterol
LR:	Logistic regression
RF:	Random Forest
XGBoost:	Extreme gradient boosting
ANN:	Artificial Neural Network
KNN:	K-nearest Neighbours
RF:	Random Forest
DT:	Decision Tree
LR:	Logistic Regression
NB:	Naive Bayes
AUC:	Area under the curve

CHAPTER I

1 INTRODUCTION

Metabolic Syndrome (MetS) is a combined disorder, the simultaneous occurrence of components that include hypertension, hyperglycaemia, obesity, and atherogenic dyslipidaemia, that then lead to the risk of different types of cancer, type 2 diabetes (T2D), as well as cardiovascular diseases (CVD) (Akbarzadeh, Evaluating machine learning-powered classification algorithms which utilize variants in the GCKR gene to predict metabolic syndrome, 2022). MetS is very common worldwide having encamped a quarter of the world, not just in developed countries but also in underdeveloped countries as there is an increase in the morbidity and mortality rate due to the disorder that mainly affects the adult population. The pathogenesis of this syndrome has been suggested to be insulin resistance, abdominal obesity, microcirculation dysfunction, endothelial cells dysfunction, and high oxidative stress. (František Babič, 2014). The pathophysiology of MetS is quite complex, and can either be genetic or lifestyle. There are metabolic changes that then cause this disease to pan out. Between 2010 and 2019 there has been an increase in the population exposed to MetS, and it being a contributor to the increase in death rates remains to be a huge concern for the world at large. MetS being an occurrence of multiple diseases the prediction for the individuals at high risk becomes our goal in healthcare. Commonly over the years, the main cause of MetS was obesity, but with time it has also been revealed that there is non-obese metabolic syndrome, making it even more or a concern. (Choe EK, 2018). For early diagnosis, predictive models need to be designed. (Akbarzadeh, Evaluating machine learning-powered classification algorithms which utilize variants in the GCKR gene to predict metabolic syndrome, 2022)

Machine learning (ML) is the scientific study of algorithms as well as statistical models, its decisions are based on meta data with the computer system not being explicitly programmed. Sometimes when humans view the data they cannot comprehend it, that is when machine learning is utilised for interpretation of data. (Park, 2018)These algorithms allow for predictions to be made. It learns about the previous experience giving great outcome compared to the old statistical model approach. There are plenty of models that have been used for medical data with positive outcomes. Machine learning is used for accurate prediction of disease all around, like breast cancer. Like any other system, the algorithms have their positives and negatives (Yu C, 2020). ML being an application of artificial intelligence, can predict, interpret text from records, see patterns in pictures in order to discover an individual's health condition. The ML models deal with meta data of information, different variables and they are then trained in a dataset where the results have already been seen, before being used on a new set of data for testing. Researches have already shown that ML is used on the prediction of MetS based on results from laboratories as well as genetics. The challenges now being that ML models designs only predict the current MetS risk and the future predictions on people with no conditions is low. As for predicting MetS based on lifestyles there has not been much data to lean on. (Daniel Tavares L, Prediction of metabolic syndrome: A machine learning approach to help primary prevention, 2022)

1.1 Thesis Problem

MetS is known to be a biomarker for the major non-communicable diseases (NCDs) the main three being CVD, T2D, and cancer, though there are other diseases that are associated with Mets, being kidney stones, psoriasis, and osteoporosis. The three major NCDs are the cause of motility in a total of over 31 million people yearly. With the studies done in this field and several publications were done on the prediction of MetS using machine learning, a bibliometric analysis will evaluate the growth of machine learning models that will better assist the prediction of MetS for current and future conditions.

1.2 Purpose of the Study

The purpose of this study is to provide other researchers with the advancement of how this problem has evolved over the recent years and inform categorically on the basis of country, author, and years that are top of the field. Also, give a quantitative analysis of the publication made in relation to machine learning applications to metabolic syndrome using a bibliometric approach method.

1.3 Research Questions

Answers to the following questions.

- Are there some advancements in the machine learning models for better prediction of the metabolic syndrome, and which area has the growth?
- Which countries, authors and years have participated most in this research.?

1.4 Limitations of the Thesis

There are some gaps in the literature available and scarce data on predictive models on certain individuals with certain conditions. The study provides a bibliometric analysis for a certain time frame as well as language therefore having to filter out some of the data, which may interfere with the final results of the study.

1.5 Overview of the Thesis

This thesis will consist of chapter I Introduction, chapter II is a literature review of the studies related to machine learning applications in metabolic syndrome, with in-depth information on machine learning applications in metabolic syndrome. Chapter III has the methodology used in this study. In Chapter IV the result will be displayed. A discussion and a conclusion will be in Chapter V.

CHAPTER II

2 LITERATURE REVIEW

2.1 INTRODUCTION

Metabolic syndrome is characterized by several abnormalities, leads to a high risk of cardiovascular attacks. It was discovered in 1988 and defined by World Health Organisation (WHO) in 1998 and the diagnosis feature is being insulin resistance. (Katsimardou, et al., 2020) It is also defined with joint interim statement (JIS) criteria, this is due to the presence of at least 3 of the metabolic conditions: Hypertension, hyperglycaemia, obesity and hyperlipidaemia. There are contributing factors like lifestyle and aging population to the development of MetS (Niazi E, 2019). The concern mainly lies on the ability of this syndrome to lead to deadlier diseases like cancer, heart disease and diabetes. The centre of all symptoms falls to insulin resistance.

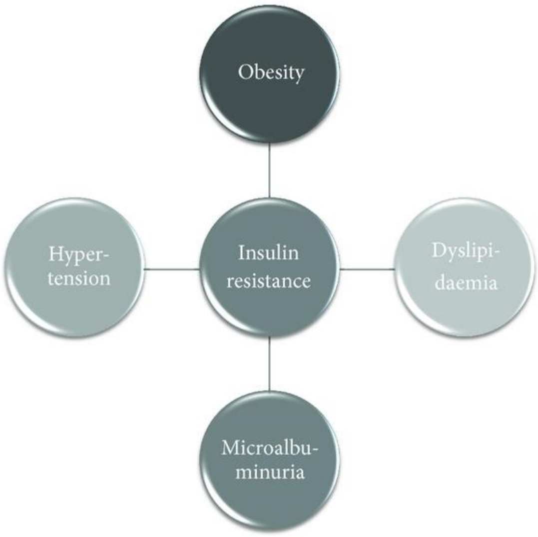

Figure 1 Shows the 4 connections to insulin resistance

Data mining has been of great importance in the early prediction of individuals highly likely to be at greater risk of MetS. Prediction and controlling MetS requires disease risk models, hence the need for machine learning algorithms. For the effectiveness of the built predictive models, certain factors are to be considered for accurate identification. These factors include, family history, age, gender, blood pressure lifestyle and activities like smoking, and exercise which is data that can be modified. Modified data is that which is not permanent and alternates. Machine learning algorithms have potent data analysis skills. (Sghaireen MG, 2022)

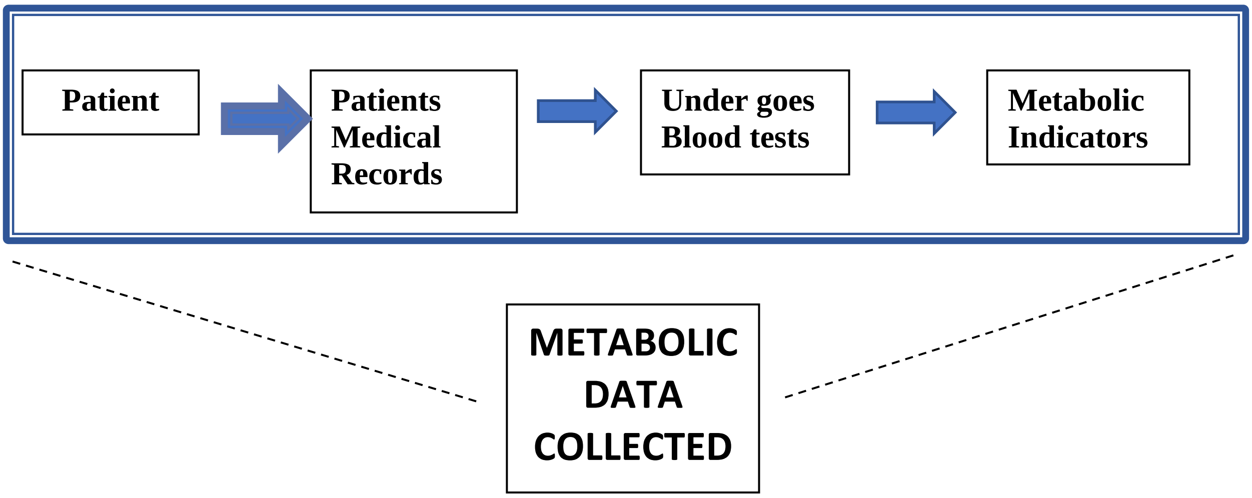

Figure 2 Metabolic Syndrome data collection processes.

2.2 MAIN METABOLIC CONDITIONS

2.2.1 Hypertension

Hypertension is linked to inflammation responses, these can be systemic or vascular and oxidative stress, therefore leading to vascular dysfunction (Zhou, 2014)**.** Though hypertension elusively has been said to be genetic condition. It is also known as high blood pressure, where the blood pressure is higher than what is normal. This condition affects the arteries, blood flow is rushed against the arteries walls, leaving the heart to over exert its energy. Being measured in millimetres of mercury (mm Hg). To be said to have high blood pressure, reading commence at 130/80 mm Hg <, first number between 120- and 129-mm Hg, bottom not being above 80 mm Hg. High blood pressure having 2 stages, primary hypertension and secondary hypertension. 180/120 mm Hg is deemed to be a crisis. Untreated hypertension can lead to stroke or heart attack. 85% of patients with MetS are hypertensive. Hypertension is undiagnosed until its late. Hypertension being a component f or Mets, has assisted with better insight, for early detection. Insulin resistance and obesity are regarded has the major contributors to this pathophysiology of MetS and the high BP, therefore leading to all the arterial damages and renal damages. Insulin resistance induce high BP by the triggering of sympathetic nervous system. Induced hypertension results from sodium retention, volume expansion and endothelial dysfunction. Moderating the secretion of leptin, adiponectin, plasminogen activator inhibitor 1, tumour necrosis factor alfa (TNF-α), interleukin-6 and resist in as they are connected to the inflammation process, endothelial dysfunction and hypertension. Healthy life style, restriction of alcohol and smoking and weight reducing, is the restorative approach for individuals with hypertension and MetS.

2.2.2 Hyperglycaemia

Hyperglycaemia also known as high blood sugar level is a condition where levels of sugar in a patient's blood are too high. It affects individuals with diabetes mainly. A high blood glucose level is being 48mmol/mol. Hyperglycaemia that appears in type 1 diabetes (T1DM) normally evolves without a history of insulin resistance the insulin typically increases over a certain duration of hyperglycaemia, but for type 2 diabetes (T2DM) there is a complete variation for insulin resistance is exists for a couple of years before hyperglycaemia evolves. Hyperglycaemia by itself can't be evaluated to be a cause of cardiovascular disease due to insulin resistance in all patients. Triglyceride levels are increased so is the blood pressure by hyperglycaemia, it also adds on to decreasing the HDL cholesterol levels and taking up the chances of thrombosis formation. This being determined that insulin resistance and hyperglycaemia play quite a huge role in the risk adverse changes in cardiovascular disease. Hyperglycaemia works hand in hand with high cardiovascular connected mortality. The Hyperglycaemia strokes and peripheral diseases are predictable. (Laakso, 2014)

2.2.3 Abdominal Obesity

Obesity is a worldwide major health concern, where by there is and excess gain of weight that then leads to several life-threatening conditions. Obesity has been discovered to be in 61% of cancer patients, 97% of cardiovascular diseases patients and 22% of type 2 diabetes patients. Accumulation of body fat in connected to insulin resistance. Obesity leads to a low life expectancy, and abdominal obesity is one of the components of metabolic syndrome and the percentage in the occurrence in individuals with MetS is to increase in the future. Fat build-up is steadily increasing in subcutaneous tissue more in comparison to visceral fat area (Tanner RM, 2012). The readings of Waist Circumference (WC) and Body Mass Index (BMI) are the major surrogate makers when it comes to the adiposity prognosticating hepatic steatosis. Subcutaneous abdominal adipose has a relation with metabolic risk factor. Visceral fat builds up differently depending on an individual's background. Visceral adipose tissue (VAT) has a notable presence for the evolution of MetS, it is examined with various methods, that being magnetic resonance imaging or computer tomography, where the dissimilarity of VAT is related to MetS. (Engin, 2017)

2.2.4 Dyslipidaemia

High levels of triglyceride, dense low-density lipoprotein and low HDL-cholesterol are factors contributing to metabolic syndrome, and that is how dyslipoproteinaemia is described. These conditions speed the risk of a patient developing atherosclerosis. The irregularity of lipoprotein can be caused by excess production of triglyceride-rich lipoproteins and rise in catabolism of HDL particles. The lifestyle of an individual as well as statin treat, changes lipoprotein movement in MetS. The merge of several lipid- producing agents helps statin on atherogenic dyslipidaemia. Dyslipidaemia is shown to be in connection with diabetes and insulin resistance. With gathered data, most MetS patients have either what is deemed normal or impaired glucose tolerance (IGT). The epidemiological data shows its necessity and a huge independent risk feature for cardiovascular disease within metabolic syndrome. Atherogenicity consists of some mechanism, these being, high oxidative susceptibility, and the impairment of the endothelial function that comes from nitric acid activity. (Kobo O, 2019) There are a few more abnormalities that affect patients, increasing their risk of CAD for instance chylomicron metabolism, which mainly affects individual with insulin resistance. An irregular regulation of lipoprotein in the system can be cause by the collection of overproductions of triglycerides-rich lipoprotein, low catabolism of apoB containing particles but in contrast high catabolism of HDL materials. Some improvements can be done in favour of changing the lipoprotein movement in Mets, one’s lifestyle and statin is considered to have such an impact. Statin has a few agents that allow it to be beneficial when it comes to atherogenic dyslipidaemia, these lipid regulating agents, ezetimibe, niacin and fibrates help with optimisation.

2.3 DATA SOURCE

Physical examinations are taken on patients, surveys done on the life style of an individual. If one is a smoker, or a drinker. Gender, age and activities done. The history of a patient is very critical for proper medical results and comparisons. For instance, if hypertension or diabetes is in the family history. Measurements of height, weight, waist circumference (WC), blood pressure are takes with instruments that are best calibrated before use (Elshawi R, 2019). The readings taken are accurate, with an allowance of 0.1 error. In Table 1 it shows the metabolic indicators, used for the diagnosis of MetS. With the measurements that if exceeded will give the probability that one can be diagnosed with MetS.

Table 1 Metabolic Indicators

Age	>48.07
Systolic blood pressure	SBP ≥130 mmHg
Body Mass Index	BMI ≥ 24.35
High Density Lipoprotein-cholesterol	HDL-C < 1.04mmol/L.
Triglyceride level	TG ≥ 1.7 mmol/L.
Waist Circumference	WC ≥ 90 cm for men and ≥ 85 cm for women.

2.4 MACHINE LEARNING

In previous studies for the identifications of MetS statistical methods like linear and logistic regression were utilised. Unfortunately, there are some hindering factors like multi correlation issues. Therefore, machine learning gives a different view to the Mets prediction, this new model approach being studied in the recent years has shown remarkable outcomes. The access to data sets has given rise to the medical information available and is an aid in the predictive procedure using ML algorithms. Common variables in the dataset that have been seen in the classifiers are triglycerides, HDL cholesterol and waist circumference, due to the reoccurrence they have been deemed as the leading indicators for predicting MetS in individuals. (Chen JH, 2017)

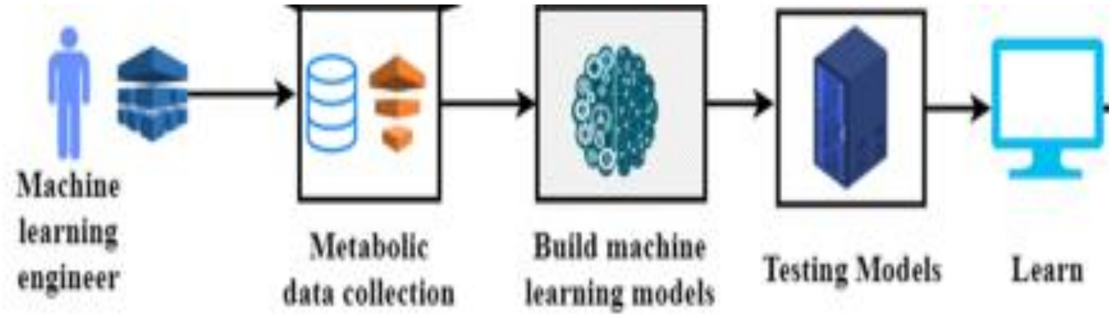

Figure 3 Metabolic Syndrome Diagnosis using machine learning (Sghaireen MG, 2022)

6 different ML classifiers can be utilised together with the collected metabolic data, these include artificial neural network (ANN), K-nearest neighbours (KNN), Bayes (NB), Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM) and Logistic Regression (LR). When now wanting to analyse the success rate of a models, we can look at sensitivity, accuracy and specificity. Values of the confusion matrix are used, showing the results of the prediction. Comparing to the original. Finding true positive (TP), false positive (FP), false negative (FN) and true negative (TN). In order to calculate the sensitivity, accuracy and specificity elements from the matrix are used. Sensitivity (S) is the true rate of the correct diagnosis in an individual with metabolic syndrome, specificity is a true rate of right diagnosis on an individual of normal subjects. Accuracy is the rate of the correct diagnosis of metabolic syndrome and healthy subjects. Area Under the Curve (AUC) also measures a model's performance, the higher the AUC the better the model.

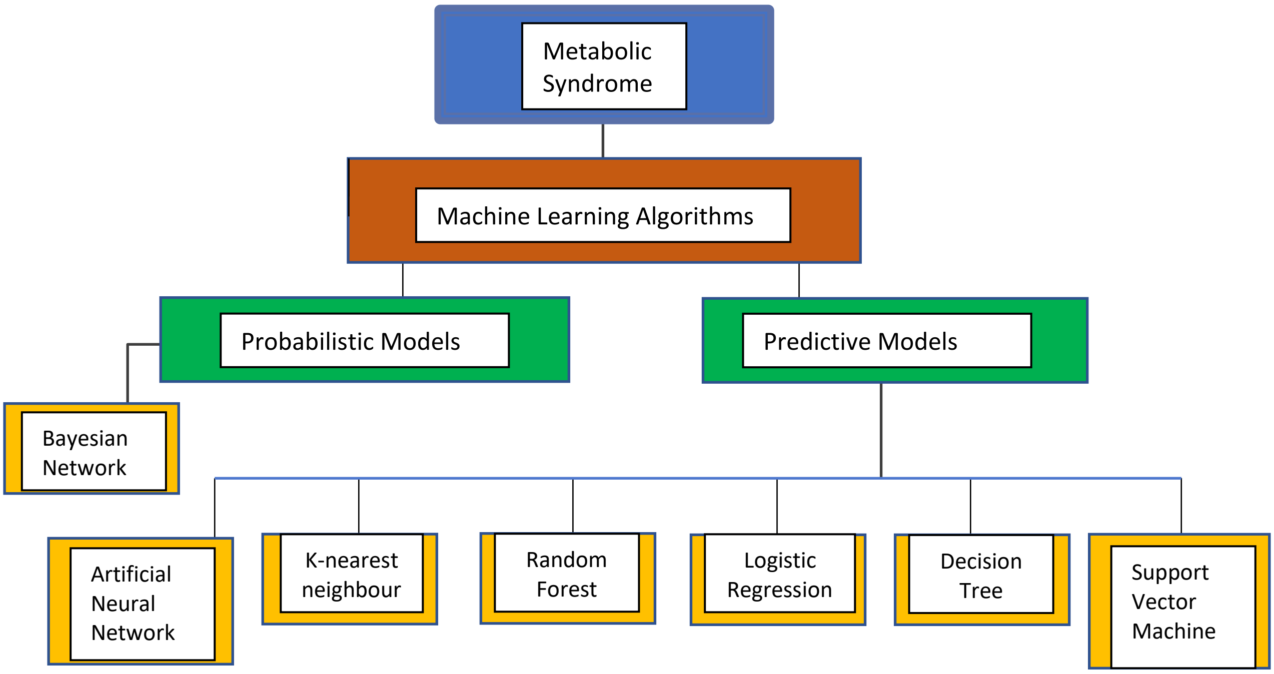

Figure 4 Classifications of metabolic syndrome diagnosis by Machine Learning

2.4.1 Supervised Learning

Supervised learning is controlled and it is an easy technique to learn in ML. This procedure relates to the description of people's respective learning goals before learning. For the first training, it is reliant on the information to understand the learning. For results the learning material is controlled. Supervised learning may start the machines generalised ability to learn in contrast to other learning techniques. The aim being to bring the solutions to problems of classification or regression, in an orderly way when having completed the system learning. BN and KNN as well as other conventional learning techniques, are examples of supervised learning methods and are currently being used. The learning process has a goal making it more methodical.

2.4.2 Unsupervised Learning

Uncontrolled learning is the opposite of the supervised learning methods. Unsupervised means no material is marked in the whole learning procedure in a particular direction, but certainly the machine will finish the knowledge information analysis. The procedure gives aid to the machine to allow essential ideas, then allowing the machine to complete a number of content learning involving concepts and materials quite important as tree roots. This growth of learning has increased the spectrum of machines. Deep-belief networks are an example of uncontrolled learning methods. (Chen et al., 2019)

2.5 MACHINE LEARNING CLASSIFIERS

2.5.1 Random Forest

Random Forest (RF) is a combined learning algorithm established on statistical learning theory. It being a combinatorial classifier consisting of numerous decision trees. RF merges Bootstrap resampling method and decision trees to assemble a collection of tree classifiers accommodates multiple basic classifiers, and the category with more decision votes H(x) is used as the category to which the final sample belongs, using a simple majority voting method. The structure of the models aids in the early detection as well as the prediction of an individual likely to have MetS. (Szabo de Edelenyi, 2008) RF is a parallel ensemble method, utilised mainly for regression as well as classification, build on the makeup of decision tree, removes the probability of an internal unbiased DT model while increasing the prediction power. This is due to the forest building procedure. (Yu C, 2020). It being the most accurate techniques present. It is used for unsupervised learning, it consists of repeating partitioning algorithms tree until the tree gets to the maximum size (Akbarzadeh, Evaluating machine learning-powered classification algorithms which utilize variants in the GCKR gene to predict metabolic syndrome, 2022). For accuracy each tree starts from a random set of the training set data, the classification. The mean vote of the tress is what will be considered as the RF prediction. (Shimoda A, 2018)

2.5.2 Artificial Neural Network

Artificial neural network uses the computing systems in order to copy the human brain, the patterns are made similar to those of the neural networks in the brain. Currently there are many different types of ANN architecture however the one for better use is the multi-layer perception type because the method can easily understand smooth measurable functions. This methodology has surpassed all the traditional ones due to its effective performance. Their set up is made to work just like the biological neural networks, that being a natural wiring system that contains of feature in the human like axons, synapse and dendrites, which are present there for the need of communication through the electric pulses. Now there are some signals that are to be passed through but these signals will depend on how strong is the pulse's transfers to the neuron for there to be an output signal that transfers through the synapse through to the axon of the proximal neuron. (Hirose H, 2011)

Interconnected artificial neurons are a simplification of the brain that are part of a multi-layer artificial neuron that comprises of nodes that constitutes a non-linear mapping, showing the input and the output. The nodes are connected and they receive input from one node to another. These nodes being like switches, their weight matches the multiplication of input by the node. The weight can be both excitatory, promoting the electrical signal or can be inhibitory, preventing it. The outcome of the output signal is as a result of the sum of the inputs into the nodes and their modification by the non-linear activation function. The output will definitely be our probability of the input as a predictor of MetS. What is usual is starting ANN training with random weights then next stage is the backpropagation this is done in order to update the weights towards effectively mapping all inputs to the outputs. Backpropagation is an algorithm that is mostly used for the training phase of ANN this is done for the accurate search for the optimal weight values. (Eyvazlou, 2020)

2.5.3 K- Nearest Neighbor

The K-nearest neighbors (KNNs) classifier, works by a comparison process that is constructed on the distance between the two datasets. KNN classifier is a very simple application yet robust, because it stores data samples all through the training phase and then processes it in assessment phase. Highly known as being one of the slowest classifiers. A number of stages are undergone, initially the K value is found based on the samples most similar to the new data and test the deviation between them both. The prediction values are made based on the samples nearest to the selected ones. Most class values are used in predicting values, looking at the value of K, the values are set and samples that are considerably physically closest to each other are picked. In KNN the classes don't need to be linearly separated. If there is more that 10% of the values missing that variant is ignored. This supervised ML algorithm is used for health systems to determine the prediction level. For accuracy there are several variables that are to be considered, for MetS these include age, BMI, and blood glucose levels. Rules are made by training samples and no other additional data. The training phase uses vectors with a class label placed on both sides. The classification phase is based on an unlabeled vector created by seeing the most presented label in the KNN training samples. This part shows us which categories the data belongs to. The value of K is dependent on the data, but the higher the value of K the better the noise reduction for the classification procedure. our method offers more flexibility and transparency in the system of detection. (Ormskirk)

2.5.4 Logistic regression

Logistic Regression (LR) this being a standard classification method is used in machine learning for classification because the probability of some obtained event is represented as a linear function of a combination of predictor variables. Variables for instance can be weight and age. The technique is used when the response variable is categorical in nature, for instance, when it has the value yes/no or true/false, obesity or no obesity. LR is contrary to linear regression in terms of the correlation between independent and dependent variables not being a necessity. However, the purpose of logistic regression is similar to that of linear regression model, which is the prediction. The predictions of LR not being continuous, inserting a s-shaped line on the graph for the yes or no outcome. This line is fitted using the sigmoid function (S(x) = ((1/(1+e-.x))). This allows a probability that z = 1 which would be correct depending on the value of x. the predicted value would 0 or 1. If predicted that y is 0.5 is implicates that the value of x was 0.5 hence a 50% probability z = 1. A low value of x implicates low probability of z = 1, but a high value of x, implicates a high probability of z = 1. Any value of x above 0.5 gives a 50%< probability that z =1. (Sghaireen MG, 2022)

2.5.5 Support Vector Machines

When it comes to solving problems, AI has been looked at as a solution, and it is the most recent way of problem modelling. For the prediction of diseases and also diagnosis, AI continues to be used. Machine learning techniques like support vector machine (SVM) are used for the prediction of diseases. SVM has been used for the diagnosis of diseases like break cancer and the central nervous system permeability. (Farzaneh Karimi-Alavijeh(1), 2016). Early detection being the key. In cases were MetS is outnumber diabetes, it is seen to be of concern. And for some instance the symptoms of the disease are not even seen. SVM is a supervised ML technique that is also widely used not just in the medical field. Predicting the progression of metabolic syndrome and analysing the important parameters is one of the main functions of ML. For MetS patient’s datasets are used and evaluation is done using Area Under the Curve (AUC). In all being said monitoring obesity, and hypertension and some of the affecting factors will lower the chances of one having MetS. (Shahzadi Bakhtawar, 2022).

SVM is not so clear like a few of the other methods, the information taken from SVM is not directly understood by professionals. However, the accuracy of the technique has caused it to be used in a wide range of applications. Each data is a vector and the dimensions are equal to the number of factors to be put forth. SVM makes a hyperplane that divides the samples present into 2 categories. The benefit of using SVM is its simplicity when it comes to using it compared to other networks, that can freeze in local maximum, but the choosing of the kernel type is what should be mainly looked at for SVM. This is because of how much the kernel type can affect the prediction accuracy. For SVM, radial basis function (RBF), polynomial, linear and sigmoid are the kernels mainly used. (Futoma J, 2015) Now a downfall that is being faced now is the dataset, when it comes to combining the data and the machine learning, many of the samples in this field are negative, yet for accuracy of result positive values quite important. When it comes to the health sector, a high sensitivity is more beneficial than high specificity.

2.5.6 Naïve Bayes

The naïve Bayesian classifier (NB) is a powerful probabilistic model that has been applied in various medical studies. The superiority of the NBC is that it takes all information into account to reach a decision, which is natural way for physicians to make diagnostic and prognostic decisions.

2.5.7 Decision Tree

For years now AI has taken over and the healthcare system has several ways of predicting as well as diagnosing a disease. Decision Tree is used as a machine learning (ML) technique, and has been used to predict diseases like pancreatic cancer and in this case metabolic syndrome. Decision tree is said to have great efficiency. According to the paper published by (Farzaneh Karimi-Alavijeh(1), 2016) there are 3 of the 5 things listed below that need to be present for one to be diagnosed with metabolic syndrome. These include:

1. Hypertriglyceridemia
2. Low- and high-density lipoprotein cholesterol (HDL-C)
3. Fasting glucose
4. Blood Pressure

5. Waist circumference (WC).

Now when it comes to collection of data, there are certain criteria's that are used and looked at in order to make comparisons or rather to deeper understand the numbers. Featured data may include, age, gender, WC, Body mass Index (BMI), physical activity like smoking, fasting blood glucose just to name some. Decision tree being a supervised ML technique, the result is shown in form of a tree or a set of rules. What stands, being the most valuable more than every other part is the root of this method. The placements are in order of the significance, so meaning other features are placed lower while the tree leaves show the classification output. Its use has been seen in other fields, not just the healthcare system, and due to its effectiveness, its commonly used. One of its advantages is the way that it is easy to understand, and a bonus being its accuracy. The process starts from the roots, next is the conditions in the nodes that are followed, making it an easy process to be followed. This method is used to predict metabolic syndrome, and the things that are most effective in determining the chances of MetS are what is needed. There is a tool known as the entropy equation that is used to determine the tree nodes.

2.5.8 Gradient Boost

Gradient Bosting is an ensemble technique used for classification as well as regression. It utilises pseudo residuals (PR). GB varies from AdaBoost whereas Adaboost starts with a stump but GB begins with one leaf that then shows the median value for the predictions. GB makes a tree of a specific size which is also the behaviour of AdaBoost but the major variation being that each tree may be more fundamental than the stump. The base learners utilised are decision trees. The true values are compared to the predicted values; in that way the probability ratio is set. The method is very effective because it shows the prediction errors in GB that have to be decreased, to the lowest point through modifying the PR values on trees.

2.5.9 Adaptive Boosting

Adaptive Boosting also called Adaboost is an ensemble technique that is used for classification. In comparison to random forest and DT this method is ordered. A couple of decision trees are combined together in a decentralised way. Each being called a stump, that constitutes of one node in addition to a pair of leaves. Assembling tree trunks that have been cut down to their roots and left connected together is what is considered a forest of stamps. However, stumps are not strong learners for the classification procedures, due to the output of one tree having an impact on the output of the next tree. AdaBoost algorithm put a high weight on the orders that create the stumps. For ranking process, each data test is allocated a specific weight (w) it being proportionate to the complete tests as a whole. A Gini index method can also be utilised to calculate the quality of a stump, normally a decreased Gini score shows more importance.

2.5.10 XGBoost

XGBoost, which is extreme gradient boosting is an ensemble method that utilises boosting. The ensemble methods constitute of gradient and adaptive boosting, and XGBoost being the most recent technology has replaced other techniques. Extreme Gradient Boosting is calculated to manage extensive and complex data. The variation of adaptive and gradient boosting is that XGBoost utilizes revolutionary regression tree in its algorithm, being converted into reborn trees. Nevertheless, the difference of the gradient boosting is that XGBoost starts with one leaf instead of many leaves. Regularization as well as the gradient boosting are very important procedures for XGBoost. It has soon to be one of the most successful boosters amongst others for quite some time. The accuracy of the predictions done are very high and it makes it easy to use it practically. Boosting has the potential of controlling mega datasets with variance target classes. (Sghaireen MG, 2022)

Table 2 Summaries research related to metabolic syndrome.

This tables showcases the names of the papers that were found to be as informative in the topic of metabolic syndrome with the use of machine learning applications. The methods used are presented with the best outcomes from the variety present. MetS indicators are noted down as well as the references of these papers.

No	Name of the Paper	Method(s) Used	Outcomes	Metabolic Indicators	Ref

1	Prediction of Metabolic Syndrome in a Mexican Population Applying Machine Learning Algorithms	Random Forest, C 5,4	RF s outperforms other methods	• Triglycerides • HDL cholesterol • Waist circumference	(Guadalupe Obdulia Gutiérrez-Esparza 1, 2020)
2	Classification and Prediction on the Effects of Nutritional Intake on Overweight/Obesity, Dyslipidemia, Hypertension and Type 2 Diabetes Mellitus Using Deep Learning Model: 4–7th Korea National Health and Nutrition Examination Survey	Artificial Neural Network, Support Vector Machine, Decision Tree, Random Forest and Naive Bayes	NB was best performing, Area Under the Curve of 69%	No	(Hyerim Kim, 2021)
3.	Developing a Novel Continuous Metabolic Syndrome Score: A Data	Artificial Neural Network and Support	Neither ANN or SVM performed well	• Waist circumference • High-density lipoprotein	(Mohsen Saffarian, 2021)

	Mining Based Model	Vector Machine			
4	Machine learning-aided risk prediction for metabolic syndrome based on 3 years study	Random Forest and XGBoost	XGBoost has the best with and AUC 93%	• Fasting triglyceride level • Body mass index • Abdominal obesity	(Yang H, 2022)
5.	Prediction of metabolic syndrome: A machine learning approach to help primary prevention	Logistic Regression, XGBoost, K-nearest neighbor, Decision Tree, Linear Analysis and LGB	LGB had best results with an area under the curve of 86%	• Waist circumference • Triglycerides • HDL-cholesterol	(Daniel Tavares L, 2022)
6.	Predicting metabolic syndrome using decision tree and support vector machine methods	Support Vector Machine and Decision Tree	SVM shows the best results with an accuracy of 75%	No	(Karimi-Alavijeh F, 2016)
7	Prediction of metabolic and pre-metabolic syndromes using machine learning models	Decision Tree, Naïve Bayes, K-nearest neighbor,	XGBoost had the best results with and area under	• Waist-to-hip ratio • BMI	(Kim, 2022)

	with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea	XGBoost, Random Forest, Logistic Regression, Support Vector Machine, Artificial Neural Network	the curve of 85%		
8.	Opening the black box: interpretable machine learning for predictor finding of metabolic syndrome	Logistic Regression, Random Forest, XGBoost	XGBoost had the highest accuracy of 99.7%	• Fasting triglycerides • Central adiposity • Systolic blood pressure	(Yan Zhang, 2022)
9.	Prediction models to identify individuals at risk of metabolic syndrome who are unlikely to participate in a health intervention program	GBT, Random Forest, Logistic Regression.	GBT outperformed the others, with the best AUC of 89.4%	No	(Shimoda A, 2018)
10.	Predicting Metabolic	Random Forest	RF had and	• Triglyceride level	

	Syndrome Using the Random Forest Method		accuracy of 98.11%		(Worachartcheewan A, 2015)
11.	Machine Learning Approach for Metabolic Syndrome Diagnosis Using Explainable Data-Augmentation-Based Classification	Logistic regression, KNNs, SVM, DTs, RFs, AdaBoost, GB, SGB, XGBoost	KNNs outperformed others, with an accuracy of 94.4% and AUC of 84.4%	• Patient age • Metabolic panel blood test • High blood pressure	(Mohammed G. Sghaireen, 2022)

Table 3 Shows metabolic data descriptions

Features	Description
Insulin test	When the patients has any insulin test recordes availabble
Gender	If patient is (male/female).
High triglceride	A patient with high triglyceride.
Reduced HDL	A patient with a low cholesterol level, can indicates chances of a heart disease
High BP	A patient with high blood pressure.

2.6 CLASSIFIER PERFORMANCE

Where the AUC was high its recorded as an excellent performance in distinguishing between positive and negative classes. Where AUC is 0.5 its seen as a neutral area. 07. to 0.8 is good and 0.8 to 0.9 is great while more than 0.9 is outstanding. AUC being used to evaluate the effectiveness of a classifier. Table 3 shows that RF and GB have expressed an outstanding performance as classifiers with an AUC of 0.969 highest of all others.

Table 4 Classifier AUC results.

Classifiers	AUC
Logistic regression	0.837
K-nearest neighbour	0.868
Support Vector Machine	0.839
Decision Trees	0.802
Random Forest	0.947
AdaBoost	0.761
Gradient Boost	0.969
XGBoost	0.896

CHAPTER III

3 RESEARCH METHODOLOGY

3.1 BIBLIOMETRIC

The bibliometric analysis has become very common and mostly popular in research, in the past years, and this popularity can be because of the upgrades to the software used which includes, VOSviewer being visualisations software, and as well as the scientific databases which consist of databases like Scopus and Web of Science. The benefits of a bibliometric study is that it handles large volumes of scientific data and thus producing high research impact. There is a lot more to it, like discovering the trends in articles, the patterns and the extent the research has reached. The data presented includes the number of citations, and publications, the times the keywords appeared and the topics, just naming a few. Bibliometric analysis is needed for deciphering and mapping the increasing scientific knowledge of various well-known fields with vast amount of data that were not organised to make them sensible. Useful for identifying gaps and investigations and where more efforts and solutions are needed. Other techniques include cluster analysis and time series analysis. (MarcLimd, 2021)

VOSviewer, VOS standing for visualisation of similarities, which is a software tool used to construct and visualise bibliometric networks. The networks include journals that are constructed based on for example, citations. This software greatly assists scholars construct maps that are displayed in various ways, each stating a different aspect of the map. VOSviewer is used to analyse CSV files that were downloaded from SCOPUS and visualised. We chose to use VOSviewer as it is the most used software currently.

3.2 INTRODUCTION

In this chapter we will discuss all the methods used in the research, these being the design of this study, as well as the procedure of analysing data with specific tools in this research.

3.3 STUDY DESIGN

Bibliometric and visualisation analysis was used in this research.

Published research articles on applications of machine learning to metabolic syndrome, were searched and analysed on SCOPUS database. SCOPUS being an easy-to-use interface and providing wider range coverage. Other examples of databases include Web of Science and PubMed. The analysis was done on an eleven-year time frame, starting from 2011 to 2022 from a Scopus database. Once the data was collected, the analyses was displayed visually using VOSviewer tool through graphs, tables, images and charts. This allows better understanding of the results, giving information on trends, patterns and progress within the time frame.

3.4 BIBLIOMETRIC DATA COLLECTION

The data was gathered by a search made in SCOPUS database. This data was extracted and downloaded from the search made on machine learning applications to metabolic syndrome. An extensive search was done by entering specific key words that are in relation to the topic; machine learning applications to metabolic syndrome. The key words used on the SCOPUS database were displayed using “OR” operation between the like words and “AND” for an additional word. Key words were entered “machine learning” OR “artificial intelligence” AND “metabolic syndrome” taking consideration of the selection criteria as shown on Table 2. Stage one of the search had no filters, but before analysing the data, there had to be some eliminations of irrelevant articles.

3.4.1 Machine learning in Metabolic Syndrome research filtered by subject area

This set was filtered by subject areas that are in relation to the subject of interest, the subject areas namely being: medicine, biochemistry, immunology and microbiology, mathematics, multidisciplinary, nursing, engineering, pharmacology, health professions, biological sciences, chemistry, physics and anatomy, neuroscience, material science, environmental science and psychology.

3.4.2 Inclusion Criteria

1. Filtered by types of documents: articles.
2. Filtered by year of publication: 2011 to 2022.
3. Filtered by language: English.

All the data that was outside these criteria was eliminated and not analyzed.

Table 5 Shows the number of papers extracted from the SCOPUS database using specific keywords.

Keywords	**No of Articles**
machine learning OR artificial intelligence AND metabolic syndrome OR syndrome X	187

3.5 METHODOLOGY AND SOFTWARE TOOLS APPLIED

Data was extracted from the SCOPUS database. Analysis made in the SCOPUS database gave results of the number of publications made within the time frame, the publications in each of the years, the authors with the most journals, the subject area with interest in the topics and the countries with publications. The data extracted was exported to Microsoft Excel (MS Excel) in a csv file format. This software formatted and arranged the data, allowing a visualisation of graphs and tables. From MS Excel, the extracted data was imported in the VOSviewer software where it was analysed with a visual output. Carrying out a Co-authorship and Co-occurrence analysis.

CHAPTER IV

4 FINDINGS AND DISCUSSION

In this chapter we will discuss the results from the analyses made from the SCOPOS database and give detailed explanations. The data extracted was analysed from a period of 2011 to 2022 using SCOPUS. The results are presented in graphs, pie charts and visualisation software images. A detailed view of the studies done on machine learning applications to metabolic syndrome will be shown based on the articles available that were exported from the SCOPUS database in the specified criteria.

4.1 Number of documents by year

A total of 187 articles that met the specific criteria on machine learning applications to metabolic syndrome where analysed. Figure 5 shows the yearly publications made within the 11 years' period, specifically from the specified criteria of 2011 to 2022 on machine learning applications to metabolic syndrome. As seen in the figure below, 2011 and 2012 had the same number of publications below 5 and in 2013 there were no publications made. In 2014 there was an increase in the publications but a drop in 2015. 2016 had over 5 publications being the highest seen at that point. There was a slight increase in 2016, but decreased in 2017. 2017 had a steady increase throughout to 2020 reaching the highest number of publications of over 20 articles at that point. 2020 to 2021 had a drastic increase in the number of publications by year, standing at 55 published papers and 2022 showed the highest number publications of 60 to date.

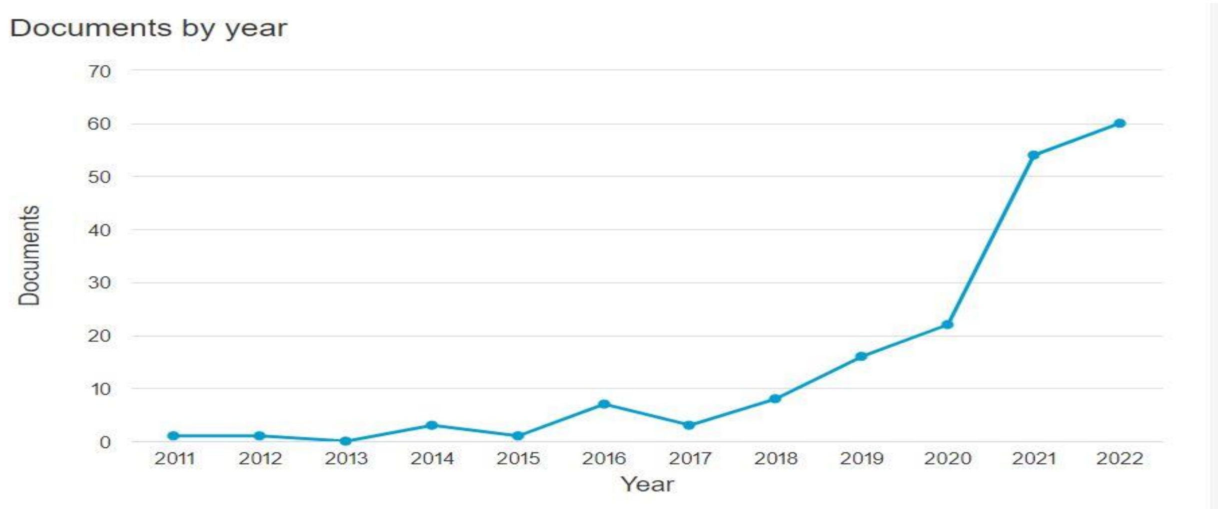

Figure 5 Documents by Year

4.2 Number of Publications by country between 2011 and 2022

Out of the 187 articles that were extracted from the SCOPUS database and analysed. A number of countries took part in the publication of articles. Figure 6 shows the top 10 countries with the highest number of publications made with the specified criteria of 2011 to 2022. The USA had the highest number of publications of 52. It was then followed by China that had 37 publications within the 2011 to 2022 period on machine learning applications to metabolic syndrome. These 2 countries are toping the charts as their publications are more than 50% of the other countries. UK had 15 publications as it sat as the 3rd leading country with publications in the specified subject. Spain had 12 publications and Germany had 11. Canada, India and Taiwan had the same number of publications, being 9. The lowest in the top 10 countries was Australia and Italy, they both had 8 publications on the specified subject. The number of countries publishing articles on machine learning applications to metabolic syndrome shows there is an interest in the growth of this topic worldwide.

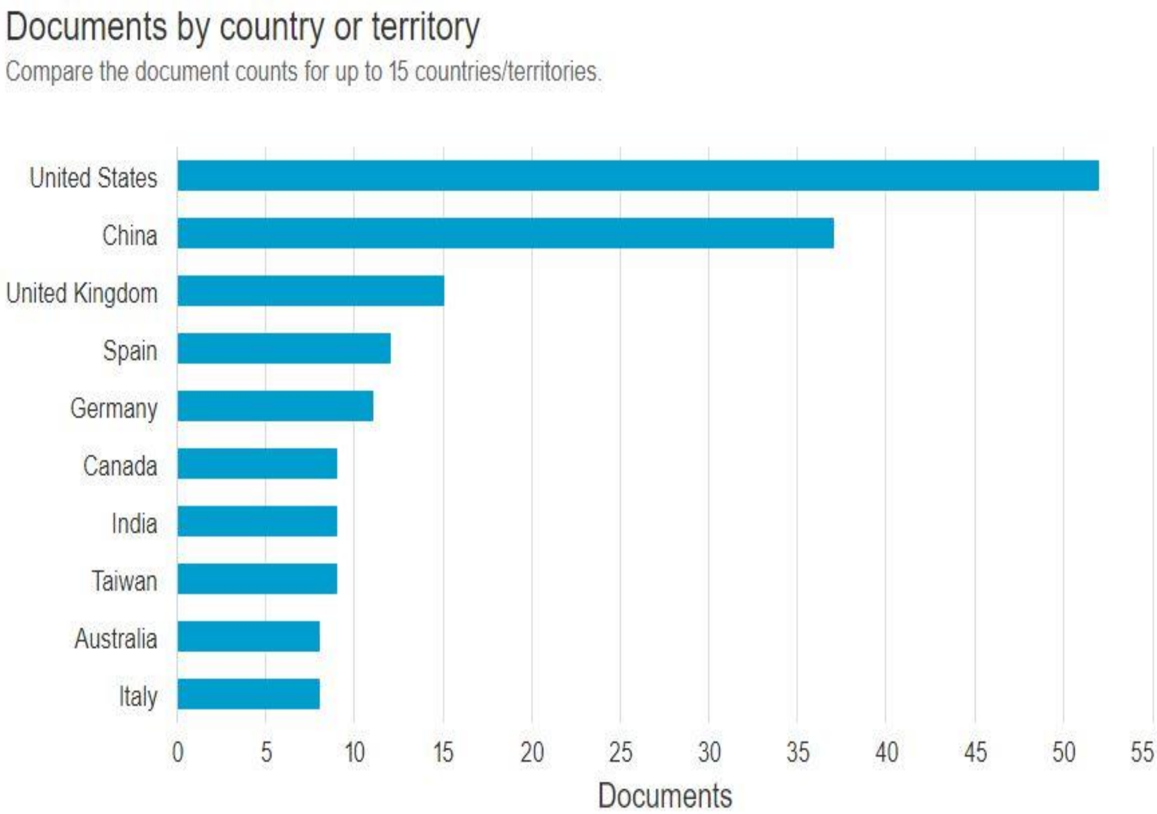

Figure 6 Shows the Documents by country or territory

4.3 Number of Publications based on Authors

There are many researchers with publication related to machine learning applications to metabolic syndrome. Figure 7 shows the top10 authors with the most publications based on the data that was extracted from the SCOPUS database. From the 187 articles exported, the highest publications were made by Nieuwdorp, M. who made a publication of 4 papers, the rest of the authors following in the top 10 each had publications of 3 papers each. There a several authors who have taken up an interest in the research of machine learning applications to metabolic syndrome.

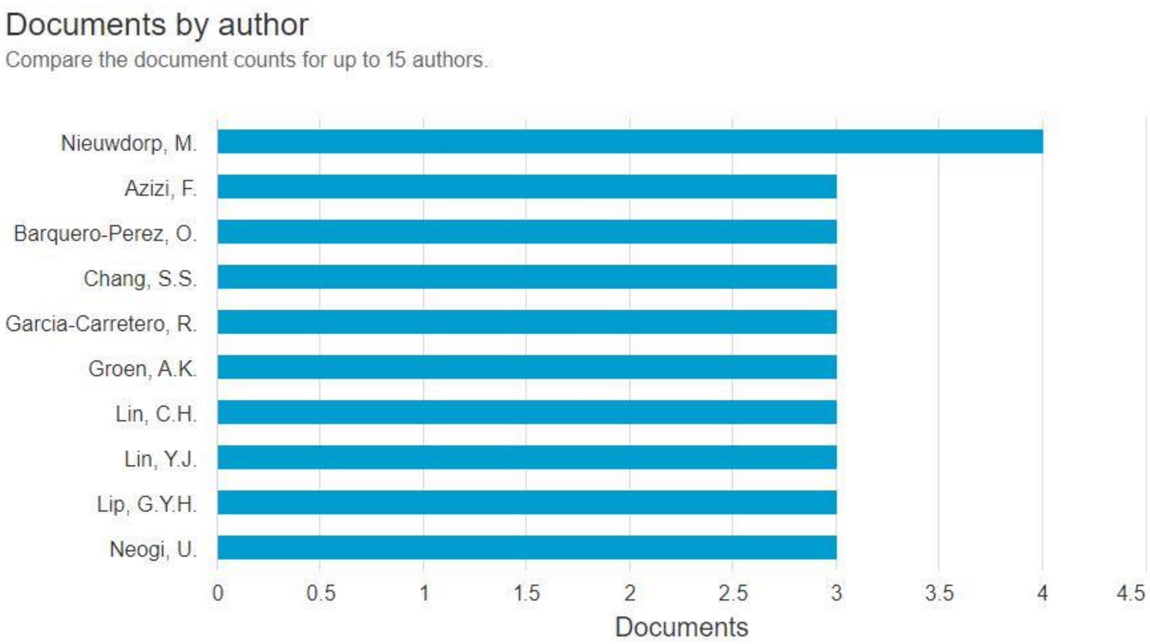

Figure 7 Shows the documents by author

Author Specialisation

1. Nieuwdorp, M.- Diabetes & metabolism and Internal Medicine
2. Azizi, F. - Cardiac Electrophysiology, Medicine
3. Barquero-Perez, O. - Biomedical Signal Processing Machine Learning
4. Chang, S.S - Urology
5. Garcia-Carretero, R. – Internal Medicine
6. Groen, A.K. - Internal Medicine, Biochemistry, Cholesterol and Insulin Resistance.
7. Lin, C.H. - Medicine
8. Lin, Y.J.- Medicine
9. Lip, G.Y.H – Cardiovascular Medicine
10. Neogi U - Molecular Biology

4.4 Subject area of publications related to machine learning applications to metabolic syndrome

There are many research areas that focus on the machine learning applications to metabolic syndrome in the existing literature. This search had exactly 18 subject areas. The lowest being psychology with 1 publication. The field of engineering had 9 papers, 3% of research publications. Mathematics having the 5th highest publications of 4%. Immunology and microbiology having the 4th highest publications of 12 papers, 4% of publications, computer science having the 3rd highest number of publications of 25 papers, 8% of research publications. 2Nd most published papers where from the field of biochemistry, genetics and molecular biology, having 49 papers, 17% of the publications. The leading subject area is medicine with 113 publications, 38% of the research papers.

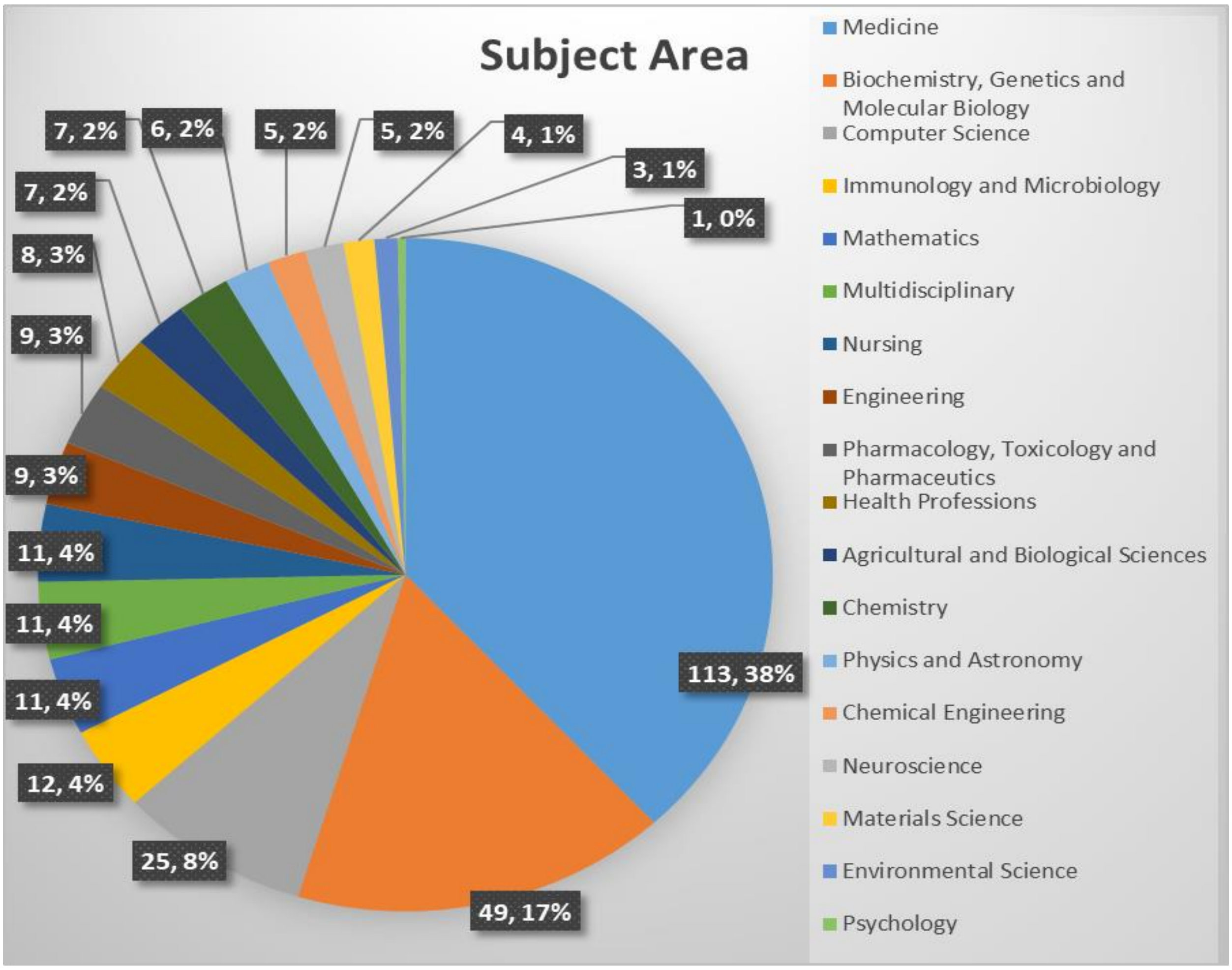

Figure 8 Shows the Subject area of the published articles

4.5 The countries co-authorship analysis on machine learning applications to metabolic syndrome

Figure 9 shows the countries co-authorship analyses results, that were obtained using the VOSviewer software. The analyses were made in relation to the machine learning applications to metabolic syndrome. from the map we can see the strong collaborations between the researchers from the varies countries. It is seen that USA has the most links of 18 with other countries. The next country being China with 6 links. The bigger the size the more the links the country has shown by the connecting lines. The countries represented here are mostly from the developed countries where more of the published paper where recorded.

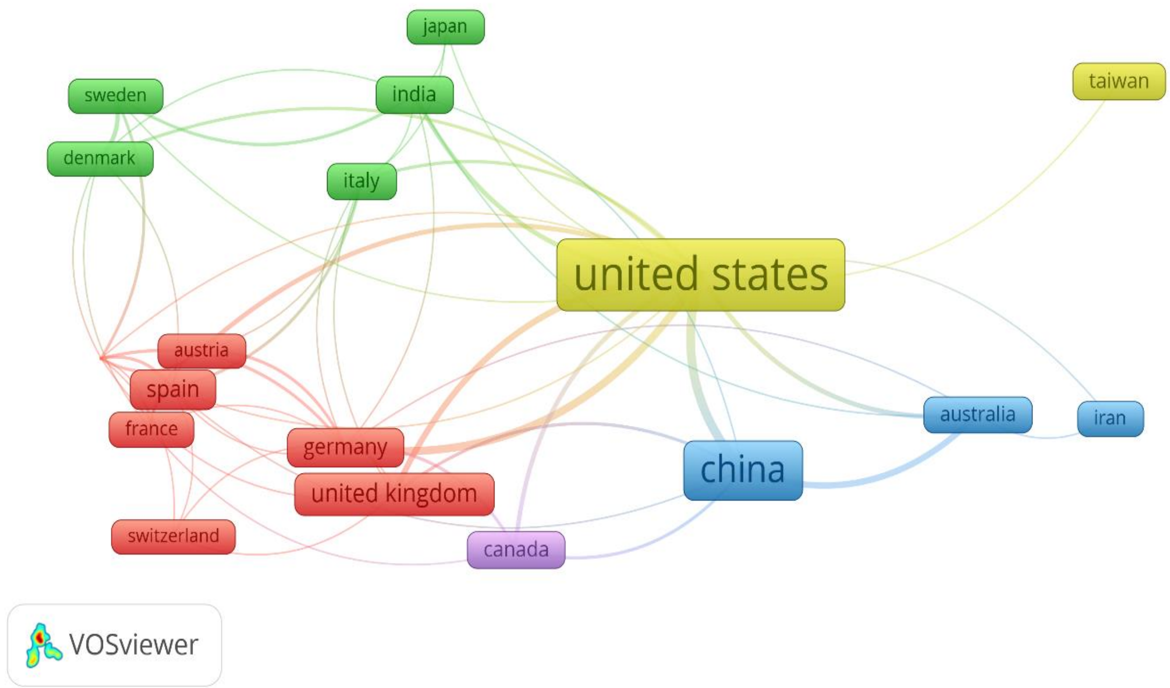

Figure 9 VOSviewer Countries Co-authorship network on the published articles.

4.6 Authors CO-authorship on VOSviewer

Figure 10 it shows the authors co-authorship made from the analysis on the VOSviewer software. There are 3 main authors that showed a link, Liu Y having 2 links with both Zhang Y, and Wang Y. From the analyses shown the publication of Zhang Y. was made in 2020 and the publication of Liu Y was done in 2021. The latest publication with co-authorship was done by Wang Y, in 2022.

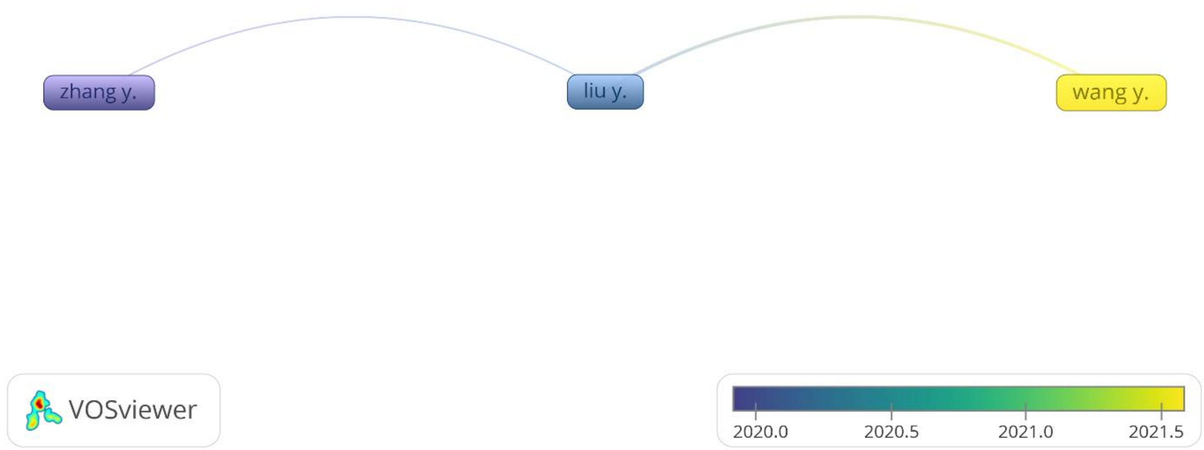

Figure 10 Authors Co-authorship

4.7 Keyword Co-occurrence analysis on VOSviewer

Figure 11 shows the keywords in the co-occurrence analyses that was done on VOSviewer on machine learning application to metabolic syndrome. the results are displayed showing the main search words used in the publication made on this specific subject. Main keywords being “human” and “metabolic syndrome” that have been proven to be present in most published articles. Keywords also include “insulin resistance”, “cardiovascular disease” as mentioned in chapter 2.

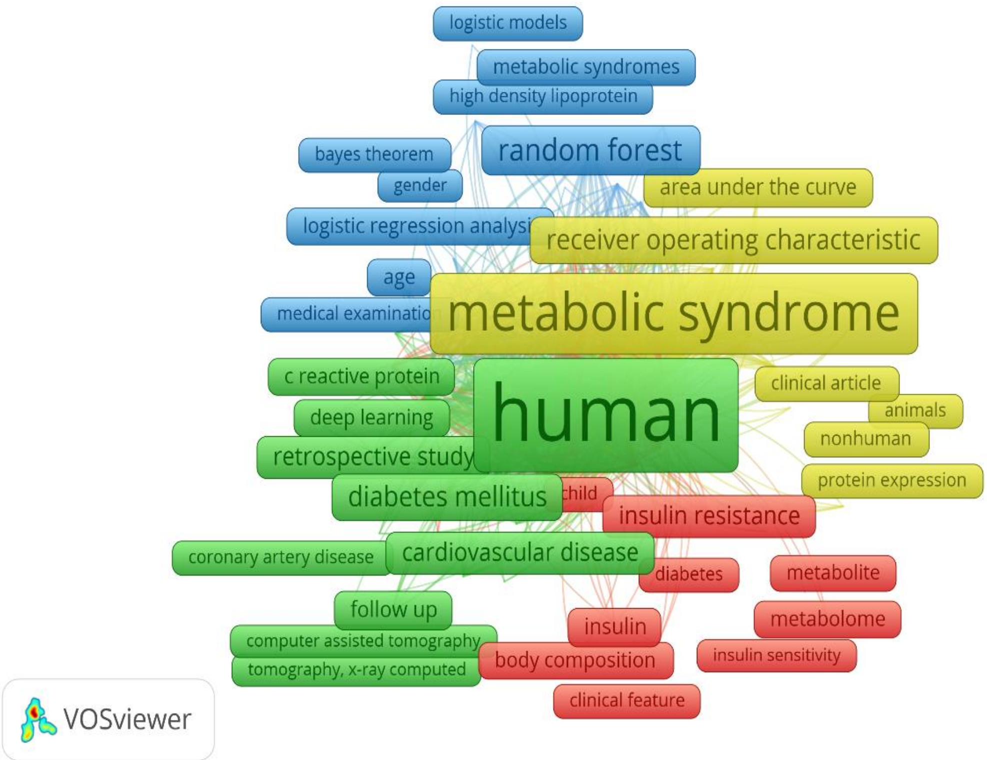

Figure 11 Shows the keywords co-occurrence

CHAPTER V

5 CONCLUSION AND RECOMMENDATIONS

According to this study, a bibliometric analysis on the machine learning applications to metabolic syndrome was undergone. With the aim being the analysis and survey in order to provide recommendations to researchers for further research in metabolic syndrome using the applications of machine learning. After all the analyses that were done it was seen that in the earlier year's machine learning was not well known, it is only in the recent years that there have been several publications made on the relation of ML to MetS. There has been an increase in the number of countries involved in this new technique, and a number of authors taking interest in the prediction of MetS using ML. Measurements of accuracy, sensitivity, specificity and area under the curve have been used has the differentiating factors in order to show the algorithms with the best outcome. ML is highly recommended for healthcare as it aids at early detection and prediction of life-threatening conditions like MetS. According to our study USA was the leading country followed by China in this research. Not much collaboration has been done between authors. As for the future there seems to be a positive response to machine learning and metabolic syndrome, leading to more publications. XGBoost having the best outcomes seen in its AUC readings. It would be recommended for countries to share their research, which will help developing countries.

6 REFERENCES

1. Akbarzadeh, M., Alipour, N., Moheimani, H. *et al.* Evaluating machine learning-powered classification algorithms which utilize variants in the GCKR gene to predict metabolic syndrome: Tehran Cardio-metabolic Genetics Study. *J Transl Med* **20**, 164 (2022). https://doi.org/10.1186/s12967-022-03349-z
2. Babič, F., Majnarić, L., Lukáčová, A., Paralič, J., Holzinger, A. (2014). On Patient's Characteristics Extraction for Metabolic Syndrome Diagnosis: Predictive Modelling Based on Machine Learning. In: Bursa, M., Khuri, S., Renda, M.E. (eds) Information Technology in Bio- and Medical Informatics. ITBAM 2014. Lecture Notes in Computer Science, vol 8649. Springer, Cham. https://doi.org/10.1007/978-3-319-10265-8_11

3. Park, C., Took, C.C. & Seong, JK. Machine learning in biomedical engineering. Biomed. Eng. Lett. 8, 1–3 (2018). https://doi.org/10.1007/s13534-018-0058-3

4. Yu C, Lin Y, Lin C, Wang S, Lin S, Lin S, Wu J, Chang S. Predicting Metabolic Syndrome With Machine Learning Models Using a Decision Tree Algorithm: Retrospective Cohort Study. JMIR Med Inform 2020;8(3):e17110 URL: https://medinform.jmir.org/2020/3/e17110 DOI: 10.2196/17110

5. Choe EK, Rhee H, Lee S, Shin E, Oh SW, Lee JE, Choi SH. Metabolic Syndrome Prediction Using Machine Learning Models with Genetic and Clinical Information from a Nonobese Healthy Population. Genomics Inform. 2018 Dec;16(4):e31. doi: 10.5808/GI.2018.16.4.e31. Epub 2018 Dec 28. PMID: 30602092; PMCID: PMC6440667.
6. Eyvazlou, M., Hosseinpouri, M., Mokarami, H. *et al.* Prediction of metabolic syndrome based on sleep and work-related risk factors using an artificial neural network. *BMC Endocr Disord* **20**, 169 (2020). https://doi.org/10.1186/s12902-020-00645-x
7. Ormskirk, U. K. Automatic Diagnosis Metabolic Syndrome via a K− Nearest Neighbour Classifier.

8. Zhou, MS., Wang, A. & Yu, H. Link between insulin resistance and hypertension: What is the evidence from evolutionary biology?. Diabetol Metab Syndr 6, 12 (2014). https://doi.org/10.1186/1758-5996-6-12
9. Sghaireen MG, Al-Smadi Y, Al-Qerem A, Srivastava KC, Ganji KK, Alam MK, Nashwan S, Khader Y. Machine Learning Approach for Metabolic Syndrome Diagnosis Using Explainable Data-Augmentation-Based Classification. Diagnostics (Basel). 2022 Dec 10;12(12):3117. doi: 10.3390/diagnostics12123117. PMID: 36553124; PMCID: PMC9777696.
10. Kobo O, Leiba R, Avizohar O, Karban A. Relative fat mass is a better predictor of dyslipidemia and metabolic syndrome than body mass index. Cardiovasc Endocrinol Metab. 2019 Sep 10;8(3):77-81. doi:
11. 10.1097/XCE.0000000000000176. PMID: 31646301; PMCID: PMC6779840.
12. Engin, A. (2017). The Definition and Prevalence of Obesity and Metabolic Syndrome. Advances in Experimental Medicine and Biology, 1–17. doi:10.1007/978-3-319-48382-5_1
13. Gutiérrez-Esparza, G. O., Infante Vázquez, O., Vallejo, M., & Hernández-Torruco, J. (2020). Prediction of Metabolic Syndrome in a Mexican Population Applying Machine Learning Algorithms. Symmetry, 12(4), 581. https://doi.org/10.3390/sym12040581
14. Yang H, Yu B, OUYang P, Li X, Lai X, Zhang G, Zhang H. Machine learning-aided risk prediction for metabolic syndrome based on 3 years study. Sci Rep. 2022 Feb 10;12(1):2248. doi: 10.1038/s41598-022-06235-2. PMID: 35145200; PMCID: PMC8831522.
15. Daniel Tavares L, Manoel A, Henrique Rizzi Donato T, Cesena F, André Minanni C, Miwa Kashiwagi N, Paiva da Silva L, Amaro E Jr, Szlejf C. Prediction of metabolic syndrome: A machine learning approach to help primary prevention. Diabetes Res Clin Pract. 2022 Sep;191:110047. doi: 10.1016/j.diabres.2022.110047. Epub 2022 Aug 24. PMID: 36029889.
16. Kim, H.; Lim, D.H.; Kim, Y. Classification and Prediction on the Effects of Nutritional Intake on Overweight/Obesity, Dyslipidemia, Hypertension and Type 2 Diabetes Mellitus Using Deep Learning Model: 4–7th Korea National Health and Nutrition Examination Survey. Int. J. Environ. Res. Public Health 2021, 18, 5597. https://doi.org/10.3390/ijerph18115597

17. Karimi-Alavijeh F, Jalili S, Sadeghi M. Predicting metabolic syndrome using decision tree and support vector machine methods. ARYA Atheroscler. 2016 May;12(3):146-152. PMID: 27752272; PMCID: PMC5055373.
18. Kim, J., Mun, S., Lee, S. et al. Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea. BMC Public Health 22, 664 (2022). https://doi.org/10.1186/s12889-022-13131-x
19. Zhang Y, Zhang X, Razbek J, Li D, Xia W, Bao L, Mao H, Daken M, Cao M. Opening the black box: interpretable machine learning for predictor finding of metabolic syndrome. BMC Endocr Disord. 2022 Aug 26;22(1):214. doi: 10.1186/s12902-022-01121-4. PMID: 36028865; PMCID: PMC9419421.
20. Shimoda A, Ichikawa D, Oyama H. Prediction models to identify individuals at risk of metabolic syndrome who are unlikely to participate in a health intervention program. Int J Med Inform. 2018 Mar;111:90-99. doi: 10.1016/j.ijmedinf.2017.12.009. Epub 2017 Dec 30. PMID: 29425640.
21. Worachartcheewan A, Shoombuatong W, Pidetcha P, Nopnithipat W, Prachayasittikul V, Nantasenamat C. Predicting Metabolic Syndrome Using the Random Forest Method. ScientificWorldJournal. 2015;2015:581501. doi: 10.1155/2015/581501. Epub 2015 Jul 28. PMID: 26290899; PMCID: PMC4531182.
22. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med. 2017 Jun 29;376(26):2507-2509. doi: 10.1056/NEJMp1702071. PMID: 28657867; PMCID: PMC5953825.
23. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. BMC Med Inform Decis Mak. 2019 Jul 29;19(1):146. doi: 10.1186/s12911-019-0874-0. PMID: 31357998; PMCID: PMC6664803.
24. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. J Biomed Inform. 2015 Aug;56:229-38. doi: 10.1016/j.jbi.2015.05.016. Epub 2015 Jun 1. PMID: 26044081.
25. Hirose H, Takayama T, Hozawa S, Hibi T, Saito I. Prediction of metabolic syndrome using artificial neural network system based on clinical data including insulin resistance index and serum adiponectin. Comput Biol Med. 2011

Nov;41(11):1051-6. doi: 10.1016/j.compbiomed.2011.09.005. Epub 2011 Oct 13. PMID: 22000697.

26. Tanner RM, Brown TM, Muntner P. Epidemiology of obesity, the metabolic syndrome, and chronic kidney disease. Curr Hypertens Rep. 2012 Apr;14(2):152-9. doi: 10.1007/s11906-012-0254-y. PMID: 22318504.
27. Niazi E, Saraei M, Aminian O, Izadi N. Frequency of metabolic syndrome and its associated factors in health care workers. Diabetes Metab Syndr. 2019 Jan-Feb;13(1):338-342. doi: 10.1016/j.dsx.2018.10.013. Epub 2018 Oct 10. PMID: 30641722.
28. Szabo de Edelenyi F, Goumidi L, Bertrais S, Phillips C, Macmanus R, Roche H, Planells R, Lairon D. Prediction of the metabolic syndrome status based on dietary and genetic parameters, using Random Forest. Genes Nutr. 2008 Dec;3(3-4):173-6. doi: 10.1007/s12263-008-0097-y. Epub 2008 Nov 26. PMID: 19034549; PMCID: PMC2593021.
29. Katsimardou A, Imprialos K, Stavropoulos K, Sachinidis A, Doumas M, Athyros V. Hypertension in Metabolic Syndrome: Novel Insights. Curr Hypertens Rev. 2020;16(1):12-18. doi: 10.2174/1573402115666190415161813. PMID: 30987573.
30. Laakso M, Kuusisto J. Insulin resistance and hyperglycaemia in cardiovascular disease development. Nat Rev Endocrinol. 2014 May;10(5):293-302. doi: 10.1038/nrendo.2014.29. Epub 2014 Mar 25. PMID: 24663222
31. Chen JH, Asch SM. Machine Learning and Prediction in Medicine - Beyond the Peak of Inflated Expectations. N Engl J Med. 2017 Jun 29;376(26):2507-2509. doi: 10.1056/NEJMp1702071. PMID: 28657867; PMCID: PMC5953825.
32. Shahzadi Bakhtawar, Komal Bashir,Fatima Anjum, Faria Kanwal, Ayesha Afzaal, Ayesha Haider Ali. Machine Learning for Predicting Metabolic Syndrome and its Prognostic Factors. Research Square. 2022 July. https://doi.org/10.21203/rs.3.rs-1800914/v1
33. M. Saffarian 1 V. Babaiyan 2 K. Namakin 3 F. Taheri 3 T. Kazemi. Developing a Novel Continuous Metabolic Syndrome Score: A Data Mining Based Model. Journal of AI and Data Mining. 2021 April. https://doi.org/10.22044/jadm.2021.10433.2179
34. Naveen Donthu. Satish Kumar, Debmalya Mukherjee, Nitesh Pandey, Weng ,Marc Lim. How to conduct a bibliometric analysis: An overview and guidelines.

Journal of Business Research. journal homepage. 2021 April. doi.org/10.1016/j.jbusres.2021.04.070

35. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285–296. doi:10.1016/j.jbusres.2021.04.070

7 APPENDICES

Appendix A:

Similarity Report.

NOW VIEWING: HOME > THESIS PLAGIARISM REPORT > BIBLIOMETRIC ANALYSIS SIMILARITY REPORT

About this page

This is your assignment inbox. To view a paper, select the paper's title. To view a Similarity Report, select the paper's Similarity Report icon in the similarity column. A ghosted icon indicates that the Similarity Report has not yet been generated.

Bibliometric Analysis Similarity Report

INBOX | NOW VIEWING: NEW PAPERS ▾

Submit File

Online Grading Report | Edit assignment settings | Email non-submitters

AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Tatenda Marata	Abstract	0%	--	--		2020365318	22-Feb-2023
Tatenda Marata	chapter 5	0%	--	--		2020366546	22-Feb-2023
Tatenda Marata	chapter three	0%	--	--		2020366137	22-Feb-2023
Tatenda Marata	chapter four	1%	--	--		2020366377	22-Feb-2023
Tatenda Marata	chapter one	8%	--	--		2020365621	22-Feb-2023
Tatenda Marata	all chapters	18%	--	--		2020369678	22-Feb-2023
Tatenda Marata	chapter two	18%	--	--		2020365960	22-Feb-2023