Tampere University

Atiqa Ashfaq

PREDICTION OF OXYGEN UPTAKE (VO_2) USING NEURAL NETWORKS

Master of Science Thesis
Faculty of Information Technology and Communication Sciences
Examiners: Dr. Philipp Müller, Prof. Neil Cronin
April 2022

ABSTRACT

Atiqa Ashfaq: Prediction of oxygen uptake (VO_2) using neural networks
Master of Science Thesis
Tampere University
Master's Degree Education in Computing Sciences
April 2022

This thesis focuses on using neural network models for the prediction of oxygen uptake (VO_2). The predictions are made using regression techniques. The dataset contains independent predictor variables such as heart rate (HR), energy expenditure (EE), height, body mass, gender and age. VO_2 is the output dependent variable. The goal is to evaluate and compare the performance of neural networks to other machine learning techniques such as support vector machines and multiple linear regression.

Few neural network models have been tested previously in the literature for maximal oxygen uptake (VO_2max) prediction. During the last decade, most approaches have focused on support vector machines and linear regression equations. In this thesis, data collected at the University of Jyväskylä is used to create a dataset for the prediction of VO_2. A detailed statistical analysis has been performed to see the relationship between speed, VO_2 and energy expenditure. Using 8 different combinations of predictor variables, neural network's performance and the effect of predictor variables on the performance is measured. Data pre-processing is performed. R^2 value and root mean square error value is used for measuring the performance of the machine learning models. Same data set is used for all models to ensure accurate results.

The results of this thesis show that speed, VO_2 and energy expenditure have a direct relationship. Males show higher energy produced as compared to females. The neural network model outperformed support vector machine and multiple linear regression by resulting in accurate predictions, high R^2 value and low root mean square value. The highest accuracy is achieved with the model containing all predictor variables. The inclusion of HR as a predictor variable is important due to its effect on the performance of the model.

Further advancements in neural networks can allow more accurate VO_2 predictions, the model can also be used in a wearable device for real-time VO_2 prediction. The same approach can be extended to predict VO_2max values.

Keywords: Neural Network, Machine Learning,VO_2, VO_2max, Prediction Model, Regression

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

PREFACE

Because of my passion in the subject, it has been a pleasure to work on this thesis. I would like to sincerely thank Dr. Philipp Müller and Prof. Neil Cronin in particular for giving me the opportunity to work on this thesis, for their honest advice and great support during this project. In addition, I'm grateful to my family, particularly my spouse, for the never-ending support during my studies.

Tampere, 6th April 2022

Atiqa Ashfaq

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

R^2 value	Coefficient of Determination
AI	Artificial Intelligence
ANN	Artificial Neural Network
CRF	Cardio-respiratory fitness
EE	Energy Expenditure
HRmax	Maximum Heart Rate
JYU	University of Jyväskylä
LSTM	Long Short Term Memory
ML	Machine Learning
MLR	Multiple Linear Regression
R-value	Correlation Coefficient
RMSE	Root Mean Squared Error
SEE	Standard Error of Estimate
SVM	Support Vector Machine
TAU	Tampere University
TUNI	Tampere Universities
VO_2	Oxygen Uptake
VO_2max	Maximal Oxygen Uptake

1. INTRODUCTION

Oxygen uptake (VO_2) is the oxygen consumed during daily life activities and exercises. Maximal oxygen uptake (VO_2max) is the maximum amount of oxygen that a body can extract from the air during intense exercise. VO_2max is of significant value in measuring aerobic fitness level. VO_2max is often used to assess athletes' aerobic endurance and is considered a measure of cardio-respiratory fitness (CRF) in both medical and non-medical applications [1]. Severe cardiac events such as stroke can occur as a result of low CRF [2]. In many pieces of research, authors have emphasized the importance of improving or maintaining an optimal level of CRF because a higher fitness level significantly reduces the risk of heart and lung diseases, therefore improving the quality of life. Thus, CRF should be monitored regularly. The prediction of VO_2 and VO_2max is an important component of health monitoring including CRF and fitness improvement [3].

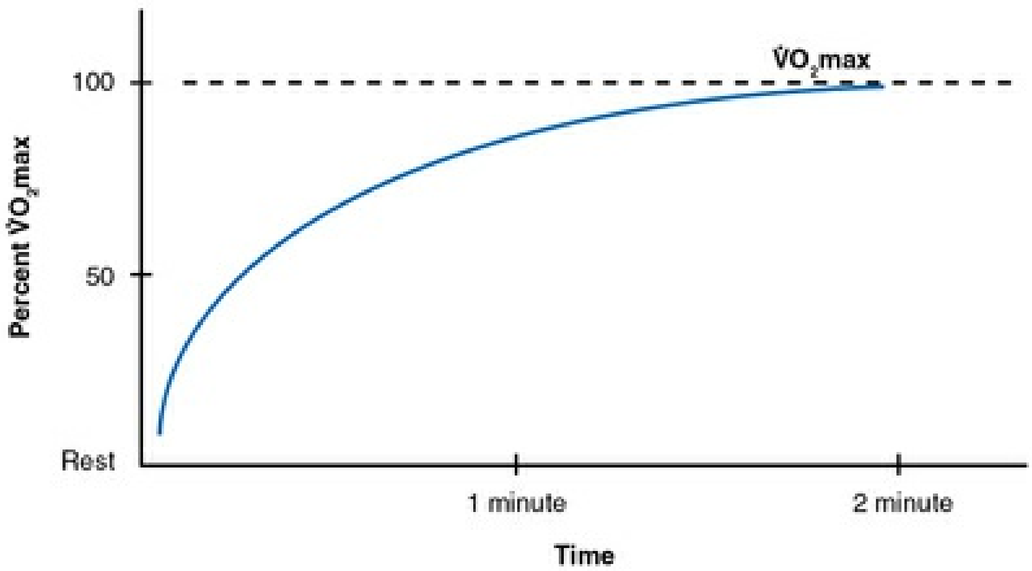


Figure 1.1. *VO_2max graph*

VO_2 is directly measured during the graded exercise test (GXT). The test is performed using a maximal intensity exercise on a treadmill or cycle ergometer in a laboratory under strict monitoring. VO_2max is measured when no increase in oxygen uptake is noticed despite an increase in exercise intensity (steady-state peak oxygen uptake shown in Figure 1.1) [4]. These direct tests are expensive as they require administered protocol and monitoring. They can also pose severe health risks as the participant reaches its maximal heart rate [5]. However, direct testing using maximal tests gives accurate VO_2max prediction. Due to the disadvantages of the maximal test, other methods such as the non-exercise test and sub-maximal test have been developed for determining VO_2max. These methods indirectly predict VO_2max without requiring the need for maximal exertion of the participant.

Sub-maximal tests use exercise predictor variables to predict VO_2. The test is performed on a treadmill or a track. The advantages include easy to administer test, less time consuming and safer as compared to maximal tests. Furthermore, sub-maximal exercise testing allows evaluating participants' adaptations to exercise and instructs them to choose a suitable exercise intensity. These tests have proven to be less precise than maximal tests while being more precise than non-exercise tests [6]. However, the use of sub-maximal tests with modern machine learning techniques such as neural networks can result in accurate VO_2 predictions.

Non-exercise tests are dependent on honest self-reporting of the participant regarding exercise. These tests are independent of using expensive laboratory equipment and can be easily administered to a larger population. Non-exercise tests may be preferred by researchers over maximal and sub-maximal tests due to their simplicity. Hybrid models (combination of non-exercise and sub-maximal tests) can give accurate results for VO_2 prediction without posing severe health risks. However, the selection of a suitable predictor variable plays a role in accurately determining VO_2 [7].

In the literature, many prediction models based on maximal, sub-maximal and hybrid testing have been developed for VO_2max. [6] lists machine learning models for VO2max prediction from 2010 - 2015. Models between 2016 - 2021 are listed in [8]. Machine learning approaches, such as artificial neural networks (ANN), have seldom been utilized as prediction models to predict VO_2max with given input characteristics such as age, gender, and body mass index (BMI) as the majority of published models are built using regression equations or machine learning techniques such as Support Vector Machines (SVMs) and Multiple Linear Regression (MLR).

Deep neural network-based artificial intelligence (AI) applications have dominated a variety of industries in the last decade such as face recognition, language processing, and financial analysis. While the notion of representing a decision function with a network of artificial neurons was first articulated several decades ago, artificial neural networks have only recently acquired popularity. ANNs learn from existing data and are analytic methodologies capable of predicting new observations from other observations. They are based on the cognitive system's learning process and the brain's neurological functions. In the last decade, there has been a spike in interest in ANN. Because ANNs may generate nonlinear input-output mappings, they have shown significant advantages over more standard regression techniques. Traditional strategies, which rely on linear or basic nonlinear models, are incapable of capturing complicated connections within obtained data, i.e., they are vulnerable to the underfitting problem. In medical fields, ANN based prediction models have been extensively employed. However, there is not enough prior research to assess the ability of ANN for prediction purposes in sports science to estimate physical fitness data and VO_2.

1.1 Objectives of the thesis

The objective of this thesis includes developing an ANN model for predicting VO_2 from the available data. The performance of the ANN model is then compared with the LSTM based neural network.

An extensive literature review is included regarding previous studies involving neural networks for VO_2max prediction. Dataset is created for training, validation and test set. The effect of different combinations of predictor variables on model performance and VO_2 prediction is evaluated. Evaluation metrics such as R^2 and RMSE are used to measure the accuracy of the models.

Furthermore, the outcomes of the neural network based model are compared with different machine learning models proposed in the literature specifically, multiple linear regression (MLR) and support vector machine (SVM). This research aims to address three main hypothesis.

Hypothesis 1: An increase in running speed results in an increase in VO_2 which increases energy expenditure.

A detailed statistical analysis is performed to observe the increase in energy expenditure as the running speed and VO_2 increase. Statistical analysis includes the use of graphs, mean and standard deviation.

Hypothesis 2: Neural network outperforms other machine learning techniques, as indicated by higher R^2 and lower RMSE values.

Machine learning techniques used to predict VO_2max in the last decade are identified and an extensive literature review is carried out. The results are compared to investigate how neural networks performed as compared to other machine learning models (SVM, MLR).

Hypothesis 3: Increasing the number of independent predictor variables results in an increase in prediction accuracy.

The effect of predictor variables on prediction accuracy is investigated by using 8 different combinations of predictor variables.

1.2 Organisation of the thesis

The following is how this thesis is organized: Chapter 2 covers the important conceptual background of VO_2, a literature review on the VO_2max, and machine learning. The methods for data gathering and processing are described in Chapter 3. It is followed by Chapter 4, which offers the results, discussion and analysis of the models that have been developed. Following that, Chapter 5 discusses limitations and future suggestions. Finally, the conclusion is presented in Chapter 6.

2. MACHINE LEARNING FOR VO_2 PREDICTION

Supervised learning is the use of known data to do predictions on fresh data. The requirement of the input/output configurations contained in the known data differentiates supervised learning from unsupervised learning. The primary goal of the supervised approach is to learn how to predict an output variable y from n input variables $x_1, x_2, ..., x_n$. For both regression and classification, supervised machine learning algorithms are applied. Classification is the process of determining which category an item belongs to. The goal of regression is to predict a property associated with an object. The main difference between classification and regression problems is that in regression, the output variable y is continuous, whereas, in classification, it is discrete. The regression task in this thesis is to predict the VO_2 value.

2.1 Oxygen Uptake (VO_2) $\&$ Maximal Oxygen Uptake (VO_2max)

The term "oxygen consumption" is the amount of oxygen that can be extracted from the air. Many physically demanding jobs use VO_2 values to determine an individual's ability to execute work tasks successfully. In this thesis, relative VO_2 values are used as a dependent variable [9]. VO_2 is the difference between oxygen inhaled and oxygen exhaled in a specified time. Absolute VO_2, measured in L min^1, refers to the quantity of oxygen consumed by a body, independent of size, gender, or age, whereas relative VO_2 refers to the absolute VO_2 value modified to a reference (mL kg^{-1} min^1). Relative VO_2 is dependent on several factors such as age, gender, environmental conditions, and fitness level [10].

VO_2max is the largest volume of oxygen utilized by a person while exercising at maximal capacity, and it is generally measured as a relative rate in millilitres of oxygen per kilogram of body mass per minute (mL kg^{-1} min^1) [7, 11]. It refers to the maximal rate of oxygen consumption during exercise. VO_2max essentially assigns a monetary value to this rate of energy expenditure. The larger the VO_2max, the greater the oxygen consumption at all levels of exertion. This means that the muscles get more oxygen to transform nutrients into fuel which is used by muscles to perform. The capacity to supply oxygen to active muscles, as well as the effectiveness with which muscles utilize the oxygen, is measured by VO_2max [12].

2.2 Neural Networks

The brain is capable of performing complex perceptual activities such as face recognition and speech as well as control operations (e.g. body movements and body functions). The brain has the advantage of having a highly parallel computing structure for information processing. The human brain has massive neural networks capable of executing cognitive, and control tasks at which humans excel. Brain neurons receive messages (electric impulses sent by chemical processes) from other neurons, process them, and transfer them to their connections [13]. A simple diagram of a human neuron is presented in figure 2.1 A.

Neural networks have been constructed as expansions of mathematical models of the human nervous system. Artificial neurons, often known as neurons or nodes, are the fundamental processing units of neural networks inspired by human neurons. Synaptic effects are depicted in a simple mathematical model of the neuron by connection weights that alter the effect of associated input signals. Neurons' non-linear behavior is represented by a transfer function [14]. The neuron impulse is then calculated as the weighted sum of the input signals after they have been converted by the transfer function. The learning ability of an artificial neuron is acquired by adjusting the weights in accordance with the learning method of choice. A simple diagram of an artificial neuron is presented in figure 2.1 B.

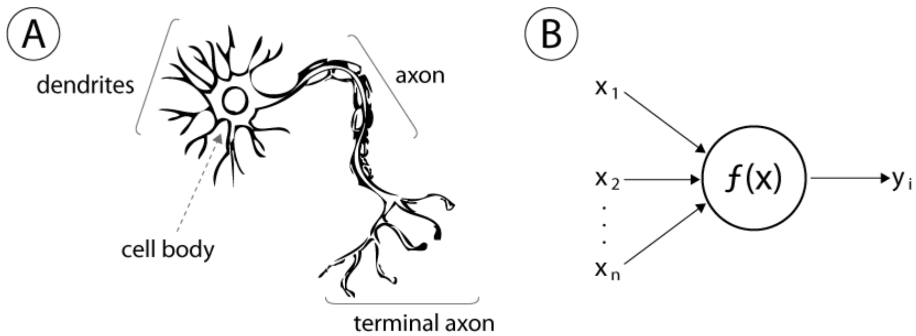


Figure 2.1. *Model of human neuron cell (A) and artificial neuron (B) [13]*

2.2.1 Artificial Neural Networks

Artificial Neural Network (ANN) approaches, especially deep ANNs, which are generally referred to as Deep Learning, are largely responsible for recent key achievements in Artificial Intelligence. ANNs are optimization algorithms modelled after biological brain networks [15]. A neural network contains computer processing elements known as neurons, with the architecture of the network created by the links (directed edges) between neurons. Layers are commonly used to arrange neurons, and a network is normally made up of several layers that are separated into three categories: input, hidden, and output layers. The input layer consists of neurons that take task input data in the form of scalars, vectors, tensors, or a mix of these. The input data for an object identification task can be an image, but the input data for time series prediction could be time-series data of recent historical values. As a result, the number of input neurons or the size of the input layer is comparable to the magnitude of the input signal. In a neural network, there could be no hidden layers, a single hidden layer, or multiple hidden layers. A hidden layer neuron gets the output of one or more other neurons as inputs, which are then adjusted in a nonlinear way to form its output. The output layer neurons are responsible for producing the neural network's correct outcome, which can be any expected quantity such as the likelihood of things appearing in the input image or the projected value of VO_2 [16].

In a neural network, each neuron symbolizes a mathematical formula that is typically described by weights. A neural network works hierarchically, processing data from the input layer to the output layer through hidden layers. Because of its architecture, when the configurations of a neural network are properly adjusted for a certain job, the neural network is able to obtain data representations of greater degrees of abstraction. The act of modifying the characteristics of a neural network to accomplish a given problem using a collection of sample data is known as training or optimization. As a consequence, the training set refers to the data used to improve neural networks. Training a neural network necessitates the use of a cost function, also known as a loss function. This function measures how well a network performs against a set of observations. Training is sometimes presented as a cost-benefit analysis, with lower cost/benefit values equating to higher performance. The cross-entropy function is commonly used as a loss function in classification tasks, although the root mean squared error is commonly used in regression assignments. The most often used methods for training neural networks are stochastic gradient descent approaches because the majority of neural networks represent highly nonlinear functions. When minimising the loss function, these approaches are based on gradient information to adjust the parameters' values on a frequent basis, as the name implies [17]. The most common optimizers are stochastic gradient descent (SGD) with momentum and ADAM.

Dedicated neural network architectures have been designed to handle a variety of input data formats, including vectors, images (2D/3D tensor), and time series. For common input data encoded as a vector, the traditional Multilayer Perceptron (MLP), which is constructed by entirely connected layers, is widely utilized. "Fully connected" means that every neuron in a completely connected layer is coupled to every other neuron in the preceding layer.

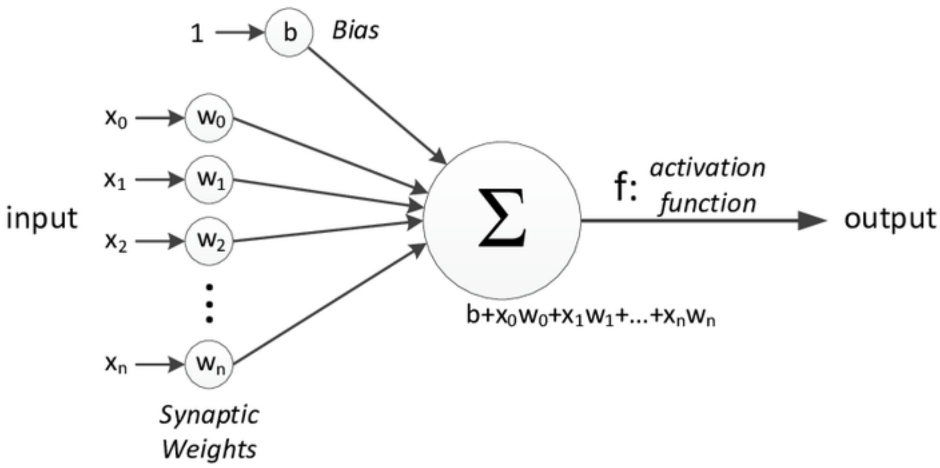


Figure 2.2. *Simple mathematical model of a neuron [18]*

Convolutional Neural Network

Convolutional neural networks (CNN) are used for pictures, videos, or any general input material with considerable spatial information. A CNN is composed of convolution layers, each of which has a number of convolution neurons/filters that slide over multiple dimensions of the input tensor [18]. This layer combines convolution with a learnable kernel, allowing the network to learn patterns e.g. edges, corners, arcs in pictures. Other layers, such as pooling and dense layers, may be present in a convolutional neural network.

Before transmitting input to the next layer through an activation function, a normal neural network conducts a linear combination of the previous layer's output and the current layer's vectors. Convolution (also known as correlation in signal processing) is performed between the output of the previous layer and the kernel (a small matrix) of the current layer in a CNN, which then delivers input to the next layer through an activation function. They differ from ANNs' typical (dense) layers such that they were created to collect and process image data.

Recurrent Neural Network

In time-series data processing, recurrent neural networks (RNNs) are a popular choice for interpreting input data that contains essential temporal information. These neural networks were created with the sequential type of information in input data in consideration. Feedback links are present in recurrent networks. RNNs are neural networks that feed their output back into their inputs recursively and are commonly employed for sequence data such as text, audio, video, and so on [19].

Interestingly, recurrent neural networks have been adjusted to take into account prior input data and also new entered data when performing a task. RNNs do this by storing the context of the data being fed into the network in an "internal state," which is then used to predict the outputs for the following inputs. The most prevalent type of recurrent layers is LSTM (Long Short Term Memory): their cells include tiny, in-scale ANNs that decide how much previous data should flow through the network.

Artificial Neural Network

In the modelling of complicated real-world scenarios, ANNs are critical in comprehending the dynamic response of time series data. Artificial neural networks (ANNs), like biological nervous systems, are a type of artificial intelligence that can learn on its own. They are inspired by real neurons and have a comparable architecture to the nervous system. ANNs are data processing and knowledge representation structures made up of densely linked adaptive simple processing components that can execute massively parallel computing [16]. The capacity to adapt is the most important feature of an ANN.

High parallelism, resilience, the capacity to manage complicated data, failure tolerance, the ability to accept inaccurate input, and generalisation are some of the other advantages of ANNs. There are three parts to an ANN: node properties, network structure, and learning rules.

Design of Artificial Neural Network

An ANN's architecture is made up of numerous layers and may be built in a variety of ways. Layers, which are linear arrays, are used to arrange the nodes of the ANN. The first layer of an ANN is the input layer, and the final layer is the output layer, with one or more hidden layers in between. Signal flow from input to output units is strictly in the feed-forward direction in feed-forward networks. Because there are no feedback linkages, data processing can take place across several layers. [20].

In other applications, changes in the activation levels of the output neurons are substantial enough that the dynamical behaviour is the network's output. Depending on the qualities and requirements of the application, there are alternative neural network topologies. A neural network must be constructed in such a way that applying a collection of inputs results in the intended set of outputs. Figure 2.3 shows a simple diagram of an artificial neural network with input, hidden and output layers.

Choosing the number of ANN layers, the number of nodes in each layer, and the node links are all part of the architectural design process. There is no such thing as a universal model design that solves all problems. Instead, the design of the neural network must be adapted to the unique challenge at hand, and it is common to go through numerous iterations and changes before finding the best architecture for the task.

An ANN's basic simple processing unit is referred to as a node. A node gets many inputs from the ANN's other nodes, which are coupled with various weights, and computes their

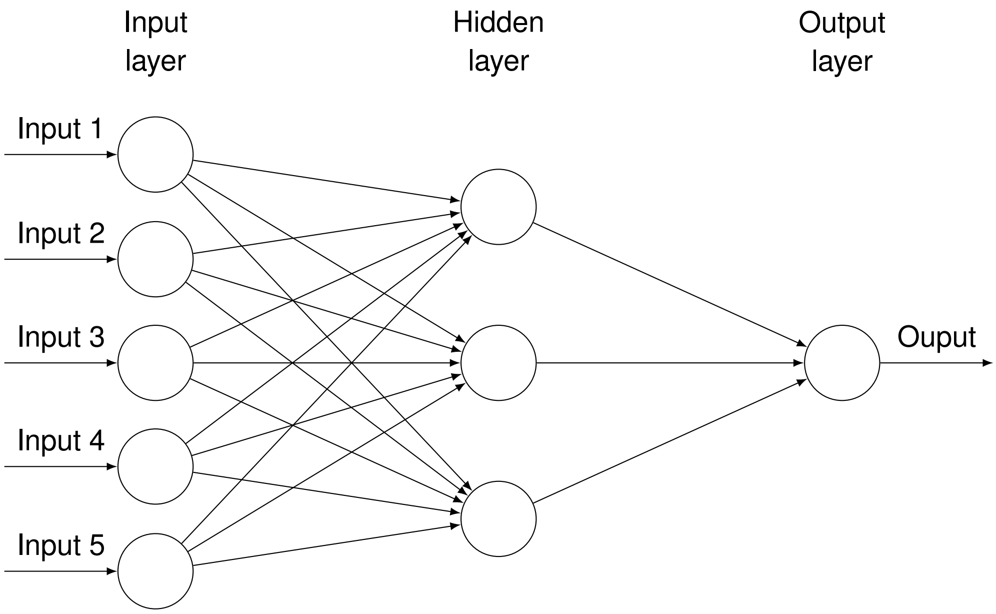


Figure 2.3. *Simple diagram of an artificial neural network*

weighted total [18].

A mathematical formula that aids in determining the output is known as an activation function. Each node in a network is given an activation function. In its most basic form, it decides whether or not a node should be functional; for example, a value of zero means the node will not be active, but a value of one means the node will be enabled. In the construction of ANNs, the sigmoid function is the most widely employed activation function. Figure 2.4 shows a sigmoid activation function. The sigmoid function returns a number between 0 and 1.

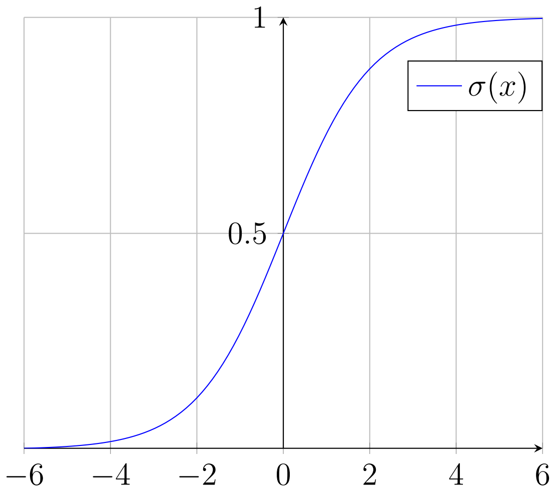


Figure 2.4. *Sigmoid Activation function*

The hyperbolic tangent function is another popular activation function shown in Figure

2.5. The hyperbolic tangent function can also yield negative values, with a range of -1 to 1.

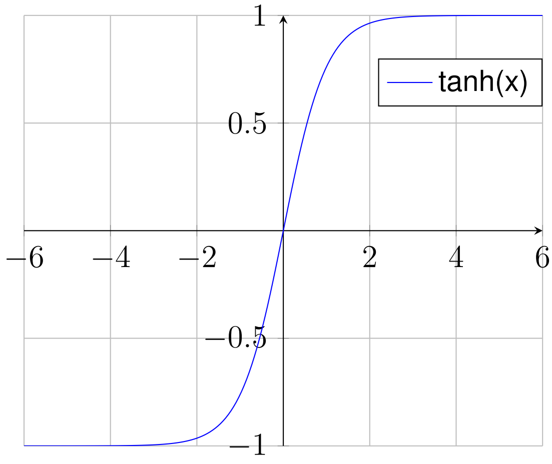


Figure 2.5. *Hyperbolic Tangent Activation function*

The activation function is determined by keeping in consideration the task, training data, and other parameters. Two more activation functions are the exponential linear unit (ELU) and the rectified linear unit function (RELU).

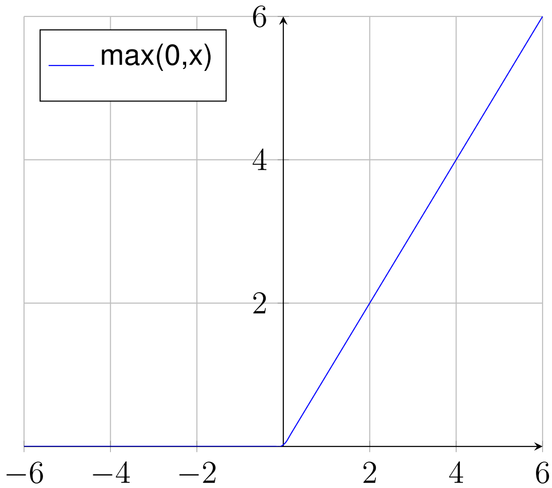


Figure 2.6. *RELU function*

Because of non-linear activation functions such as rectified linear unit (RELU, Figure 2.6), hyperbolic tangent, and exponential linear unit (ELU, Figure 2.7), ANN-based prediction model may incorporate not just linear but also non-linear characteristics of the target system. This is the most significant benefit of the ANN-based prediction which allows the ANN model to be utilized as a strong instrument, radically altering the availability of research and practice in many disciplines.

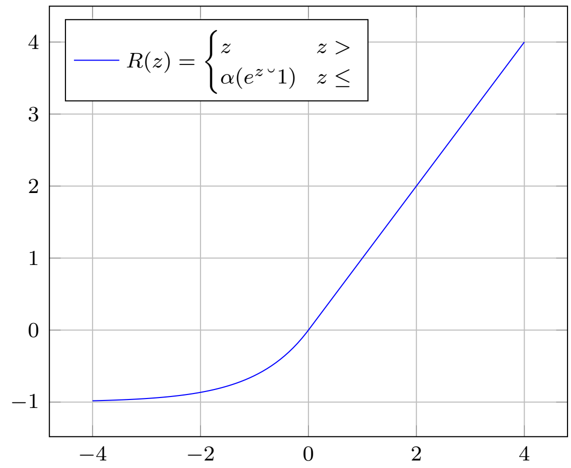


Figure 2.7. *ELU*

Generalisation

Although a neural network may be trained with an existing dataset, its true usefulness is achieved when it can perform efficiently for fresh data sets. The ability of ANNs to generalize is well known. In order for the ANN to generalize, it must be well-suited to the dataset, as underfitting or overfitting may result in inefficient performance. Overfitting of the data sets reduces the model's generalization ability, resulting in unreliable performance when applied to fresh unanticipated data [17].

Overfitting occurs when an ANN learns too much about the training dataset, accumulating complexities and noise to the point that it can no longer properly recreate the original patterns. It would then have a detrimental impact on the ANN's performance with the new data. The model acknowledges and learns noise and unexpected fluctuations in the training set as ideas, even though they have nothing to do with the incoming data sets, which explains the poor performance. Underfitting is a learning process in which an ANN is not able to obtain enough information from the training dataset. As a result, given more data samples, the underfitted ANN will be less efficient [17]. Figure 2.8 shows an example of underfitting, good fit, and overfitting.

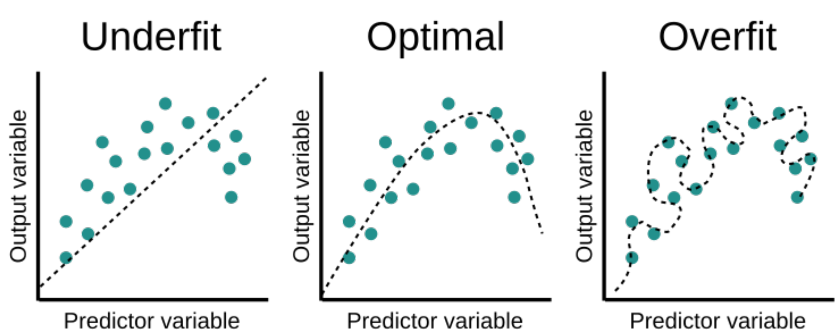


Figure 2.8. *Generalisation [21]*

The weights of an ANN are modified to train it, which is done using an optimization approach to determine the ideal weights. Stochastic gradient descent (SGD) is a very successful optimization approach. The gradient descent technique iteratively takes the negative gradient to continue in the descending direction and eventually reach the desired minimum. The gradient shows how a variation in weight impacts the total reduction. Critical considerations include learning rate and weight initialization. The gradient descent learning rate controls how much the weights change from iteration to iteration [22]. The gradient descent method will take an exceptionally long time to attain convergence if the learning rate is too low. The method, on the other hand, fails to converge if the learning rate is too high.

With a lower learning rate, the gradient descent approach may converge to the minimum. However, there is a small risk that gradient descent will find a local minimum instead of a global minimum. This can be overcome by running SGD several times. When the learning rate is too high, the approach diverges. The learning rate can be modified throughout the training process to improve learning [23]. However, the best learning rate for a model on a particular dataset cannot be determined analytically. Therefore, finding a suitable or best learning rate is done using trial and error. The phrase "learning rate decay" describes how the pace of learning decreases with time. The most straightforward technique to implement learning rate decay is to progressively lower the learning rate value from a large initial value to a small one [24].

"Adam," a stochastic gradient-based optimizer, was created as a substitute approach for SGD. Adam is a stochastic objective function first-order gradient-based optimization method based on adaptive lower-order moment predictions. Adam is well-suited to scenarios involving huge datasets because of its efficiency and robustness, as well as the fact that it does not require a lot of processing RAM. It may also be used to solve problems with gradients that are too noisy or sparse.

To prevent overfitting the training data, a dropout regularization is employed for network optimization. Dropout is a regularization method for fully connected network layers. It is accomplished by learning network weights, which provide a tradeoff between prior layer outcomes and hidden layer outcomes obtained by using dropout with a likelihood value p, commonly known as the dropout rate [25]. This implies that the dropout operates by eliminating hidden layer inputs on a probabilistic basis, making network nodes more robust to inputs in general.

To avoid further overfitting, an early stopping approach is employed to discontinue training before it becomes excessive. It can take a long time to train an ANN with a large number of epochs, and after a certain number of epochs, the training will no longer enhance performance but instead will increase the loss.

The early stopping strategy monitors, after each epoch, the validation loss and terminates training if the loss begins to rise. However, because the ANN's performance may first decline before improving owing to noise in the training data, the initial signal that the loss is rising is not always the best moment to halt training. To remedy this, the early halting strategy now includes a patience option [26]. This is the number of epochs to wait before ending if the model's performance does not improve.

2.3 Literature Review

Based on a single-stage sub-maximal treadmill running test, Akay et al., (2010) [27] built a reliable ANN-based prediction model for VO_2max of healthy adults. A maximal graded exercise test (GXT) was performed on participants (81 men and 45 women), age ranging from 17 to 40 years. The ANN prediction model was built using gender, age, body mass, stable heart rate, and running pace. The measurement findings of standard error of estimate (SEE) and correlation coefficient (R) are more accurate when compared to previous prediction models constructed using the Multiple Linear Regression (MLR) model. SEE and R values were reported to be 1.8 and 0.95, respectively. Good performance is indicated by a high R and a low SEE rating.

Jensen et al., (2021) [28] employed a maximum rowing ergometer test for VO_2max prediction model. In this regard, 34 male club rowers (18–30 years) provided data on maximal power output (MPO) in an individual 72-minute incremental (INCR) test and mean power in a 2k (W2k) rowing ergometer test, and the rowers' maximum VO_2 values were then estimated using a regression equation. The provided data was utilized to train the model using a linear regressor, and the following extremely simple prediction models were created using MPO and W2k:

$$VO_2max = 11.49 * MPO + 810 \tag{2.1}$$

$$VO_2max = 10.96 * W2k + 1168 \tag{2.2}$$

These models have relative errors of 3.1% and 3.6%, R value of 0.95 and 0.94 and SEE values of 0.136 and 0.157 for both equations respectively.

The goal of Park et al., (2021) [29] was to create an ANN model to estimate VO_2max in healthy persons using a multistage 10m shuttle run test (SRT). This study included 118 healthy people (59 men and 59 women) (38.3 $\pm$ 11.8 years, men 37.8 $\pm$ 12.1 years, and women 38.8 $\pm$ 11.6 years). Age, gender, blood pressure (systolic and diastolic), waist and hip circumferences, body composition (height and weight) and, waist-to-hip ratio (WHR) were all predicting variables. Case 3 produced the best prediction outcomes with an R^2 = 0.8206, adjusted R^2 = 0.7010 and RMSE = 3.1301 using waist and hip circumference, age, height, weight, gender, BMI, WHR, blood pressure (systolic and diastolic), number of round trips and final speed in 10 m SRT. Case 1 produced the worst results with R^2 = 0.7765, adjusted R^2 = 0.7206 and RMSE = using age,height, weight, gender, BMI, number of round trips and final speed in 10 m SRT. Case 2: using age, height, weight, gender, BMI, WHR, waist and hip circumference, number of round trips and final speed in 10 m SRT resulted in R^2 = 0.7909, modified R^2 = 0.7072, RMSE = 3.3798 [29]. The prediction accuracy of Case 2 is lower than Case 3's but higher than Case 1. However, according to the prediction results, all cases performed admirably. The model's performance was efficient in this brief article, which created an ANN-based prediction model for healthy people's VO_2max.

Beltrame et al., (2016) used an ANN model to predict VO_2max [30]. 10 healthy young adults (5 males, height 178.4 $\pm$ 11.2 cm, 29.8 $\pm$ 7.6 years old, body mass 75 $\pm$ 11.3 kg; 5 females, height 165.2 $\pm$ 7.5 cm, 22.8 $\pm$ 0.7 years old, body mass 62.1 $\pm$ 5.8 kg) provided pulse rate and treadmill data. The data collected was utilized to train an artificial neural network (ANN) to estimate VO_2max based on the speed and grade of the treadmill, gender, heart rate, exercise time and BMI. Because of its minimal bias and good linear correlation, the ANN had an R-value of 0.97, suggesting that the predictions were correct. Based on data from nine individuals, the authors developed a separate model with seven predictor variables: BMI, gender, exercise duration, recovery time, grade and speed of the treadmill, and HRmax. Cross-validation was performed using a 10-fold leave one out technique. With 11 hidden neurons and one output neuron, the model has an R-value of 0.98. Because of its simplicity, the suggested model would operate for a wide variety of people regardless of parameters such as height and weight, but more study is needed to validate this hypothesis and enhance forecast accuracy.

Borror et al., (2019) [16] used ANN to predict answers using a sub-maximal test. 12 healthy adult adults (age 21.1 $\pm$ 2.5 years, height 179.3 $\pm$ 8.9 cm, body mass index 82.1 $\pm$ 11.7 kg) cycled for 50 minutes at different intensities while having on heart rate monitors. Among the variables used to test, train and validate the ANN were heart rate, power output, the time derivative of heart rate, body mass and cadence. The model's accuracy was evaluated using a 12-fold hold-out cross-validation procedure. SEE and R-values were used to measure the model's accuracy. The model produced an R-value of 0.91 $\pm$ 0.04 and a SEE of 3.34 $\pm$ 1.07. A wide variety of exercise intensities and durations were investigated, resulting in the development of more robust models employing ANN. The approach was shown to be less reliant on rigid procedures. The projected and goal data showed reduced variation, suggesting that ANN might greatly improve energy expenditure estimations. As a result, this straightforward method has the potential to improve the practicability of tracking VO_2. The study's limitations were small sample size and a narrow range of subject ages. More research needs to be conducted to put the proposed ANN to the test on a larger dataset with individuals of various ages and levels of fitness. Other exercises, such as walking, might be accommodated by the approach.

Zignoli et al., (2020) [31] employed a recurrent neural network (RNN) to develop VO_2max prediction models utilizing cardiovascular parameters from seven male volunteers (body mass 76 $\pm$ 6.6 kg) during cycling using easily obtained inputs (intensity Levels, peak power output, weight, respiratory frequency and HR). An RNN model with three hidden layers of 32 neurons each, one hidden layer of 10 neurons, and one output neuron correctly predicted VO_2max. The authors also demonstrated that a bigger dataset may be utilized to construct an accurate model without the need for sophisticated methods to create the training and testing datasets. The models had R-values of up to 0.94.

Haneen Alzamer et al., (2021) [32] focused on some recent advancements in VO_2 prediction using machine learning in studies published between 2005 and 2020. The research includes an in-depth look into the underlying ideas of oxygen uptake measurements and ML application in sports sciences. As previously stated, several effective prediction models for VO_2max have been developed. Complex processes and health hazards connected with direct prediction methods are no longer impediments since models are commonly built utilizing data obtained through non-exercise, exercise, or hybrid methods. The paper revealed that selecting between different machine learning algorithms required finding the right combination of high R and low SEE. Because of the small sample size, more research on sophisticated machine learning techniques for VO_2max prediction is needed.

Brabandere et al., (2018) [33] developed a new model for estimating VO_2max based on accelerometer and heart rate data collected during submaximal running. They studied data from 31 recreational runners (16 women and 15 men) aged 19 to 26 who completed a treadmill maximum incremental test. During the test, the patients' heart rates and acceleration were continually recorded at three different sites (the upper back, the lower back, and the tibia). A wide variety of variables were collected during the warm-up and the first three phases of the test, and the most important ones were picked using a data-driven strategy. The researchers observed that combining accelerometer and HR data generated the best model, with a mean absolute error of 2.33 ml.$kg-1.min-1$ and a mean absolute percentage error of 4.92 %. The model takes into account gender, the inverse of the average heart rate, body mass and the variation of total acceleration during the warm-up section of the treadmill test [34]. The approach is a useful tool for recreational runners of the same age group who want to estimate their VO_2max from moderate intensity treadmill running. It recommends using two body-worn sensors: a monitoring system and an accelerometer placed on the tibia. The results showed that VO_2max is predictable using a mix of descriptive data, heart rate variables, and accelerometer features derived from moderate intensity running data. The model is constrained by two factors. First, given the study's participants were recreational runners aged 19 to 26, the model is unlikely to contribute to elite runners or those outside this age range. Developing models to predict VO_2max from elite athletes as well as people of all ages might be a fascinating future direction. The treadmill running activities are also incorporated into the model. The authors suggested that it would be interesting to look into forecasting VO_2max depending on outside running in the future [34].

Bradshaw et al., (2005) [35] developed a regression equation for estimating VO_2max using N-EX data. To determine VO_2max, all 100 individuals (ages 18-65) underwent a maximal graded exercise test (GXT). An N-EX prediction equation was created using multiple linear regression. In their investigation, they employed cross-validation. Using standardized beta-weights, they discovered that the PFA variable best predicted VO_2max, followed by age, BMI, gender, and PA-R. They created an N-EX regression model that produced satisfactory results and was a simple method for calculating VO_2max in adult women and men. The SEE and R values in their analysis were 3.63 and 0.91, respectively.

VO_2max prediction models were created using support vector regression (SVR) and multilayer feed-forward neural networks (MFFNN) in Akay et al., (2009) [36]. The VO_2max values of 100 subjects were measured using a maximal graded exercise test (50 men and 50 females). Two prediction models were developed using gender, BMI, age, PFA to walk, run or jog particular distances, and PA-R. 10-fold cross-validation was used on the dataset. The predictions of the SVR and MFFNN models were compared. Their best SEE and R were reported to be 3.23 and 0.91, respectively [37].

In Ashfaq et al., 2022 [8], a detailed review of recent advances in past five years (2016 - 2021) for VO_2max prediction is available. The article explains, in detail, the most used machine learning approaches (SVM, MLR) for VO_2max prediction. A comparison of all the machine learning approaches used in the past five years has been done, and it is seen that neural network based models have shown the highest performance in terms of SEE/RMSE and R values. The use of diverse data set and difference in sample size in different models made it difficult to assess the accuracy of the comparison. Hence, it was suggested to have the same dataset and sample size for all models in future. In this thesis, same dataset and sample size has been used for all models.

3. METHODS

3.1 Data Collection and Understanding

The data used in this thesis was collected using the data logger that has been developed and tested during the previous Academy-funded project OpenKin (2015-2019). The data logger integrates accelerometers, gyroscopes, and a GPS receiver, and can accurately measure and store most of the important kinematic parameters. It has been extensively tested at the University of Jyväskylä for walking and running tests. The design of the data logger and some field test results are described in [38]. The data used in this thesis is recorded simultaneously on the data logger and measurement system available at the University of Jyväskylä and Tampere University: Oxygon mobile oxygen uptake measurement system. This system provides gold standard data for training and evaluating the machine learning algorithms. The experimental part was performed at the University of Jyväskylä.

The dataset includes some independent (input) parameters and one dependent (output) parameter (VO_2). Different neural network models are built to predict VO_2 utilizing different input parameters of datasets, as will be explained in the following sections. The dataset utilized in this thesis involves 18 healthy participants ranging in age from 22 to 34 years (9 females and 9 men). The description of the dataset is given in Section 3.2.

During the exercise testing, all individuals were given a questionnaire and an informed consent statement to complete. All individuals were provided information on the exercise test before testing. The subjects walked and ran at a range of speeds same for all subjects (shown in 4.1). The highest VO2 value at each speed was obtained from each subject's VO_2 values. The subjects walked/ran at each speed for 3 minutes. The VO2 values used are the average over the last 1 minute, which ideally represents steady state oxygen consumption.

Oxygon mobile measurement system

The processed files in the Openkin folder contains height, body mass, age, gender. Respiratory gas data from the resting collection sitting at the edge of the field and the field measurements at different walking and running speeds are available.

Programming Environment

Python (ver.3.6) was used to create and build the neural network. The neural network was officially designed and tested using Keras, an open-source framework. Users may define and train neural networks with Keras, which is based on TensorFlow 2.0. Sequential neural networks, RNNs, and CNNs are all supported by Keras. Google designed and published TensorFlow, a python library for rapid numerical computing. It is a foundation library that may be used to directly create deep learning models or to make the process easier by using wrapper libraries developed on top of TensorFlow.

3.2 Data Analysis and Pre-processing

Data analysis was done using the data available in the OpenKin folder from Oxygon mobile measurement system. Questionnaire data (height, body mass, age and gender) was analysed using statistical analysis tools. A separate analysis was done for both males and females. The standard deviation and mean of each variable were taken for better understanding. Table 3.3 shows questionnaire data from 20 participants (males and females), including height, body mass, age and gender. Each participant has a unique identification number. Letter "M" is used for Males and "F" is used for females in the gender column.

Total Participants:	20	Mean	Std. Deviation
Age (yrs)	22 - 34	27.8	3.651
Height (cm)	162.6 - 191.0	176.24	7.67
Body mass (kg)	54.0 - 98.6	74.21	12.945

Table 3.1. *Standard Deviation and mean of data variables for all participants*

	Male	Female
Count	10	10
Age (yrs)	27 - 34	22 - 30
Height (cm)	164.8 - 191.0	162.6 - 177.4
Body mass (kg)	73.9 - 98.6	54 - 78.3

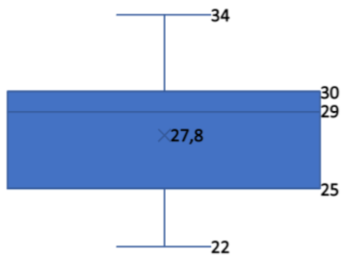
Table 3.2. *Data (Range) from participants (male and female)*

This information about the participants is listed in Table 3.1 and Table 3.2. Figure 3.1 shows box plots for age, body mass and height of the participants for better visualisation.

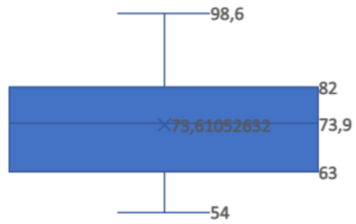
Null values were eliminated and data from 18 participants were used for VO_2 prediction models. Table 3.4 shows the predictor variables (Height, body mass, age, gender, EE, HR, VO_2) for 18 participants used for prediction models.

Identification	Height (cm)	Body Mass (kg)	Age (yrs)	Gender
OK001M	188	78.4	32	M
OK002F	175	59	22	F
OK003F	170	58.1	23	F
OK004F	168.5	54	30	F
OK005M	191	98.6	30	M
OK006F	178	63	26	F
OK007M	176.7	82	27	M
OK008F	177.4	66.3	29	F
OK009F	162.6	57.3	25	F
OK010F	167.7	63.7	25	F
OK011M	178	89.2	27	M
OK012F	174	67.5	22	F
OK013M	179.8	73.9	29	M
OK014M	179	79	29	M
OK015F	168.8	78.3	23	F
OK016F	175.1	72.4	29	F
OK017M	182	77	34	M
OK018M	180.6	95.4	30	M
OK019M	187.5	85.5	34	M
OK020M	164.8	85.6	30	M

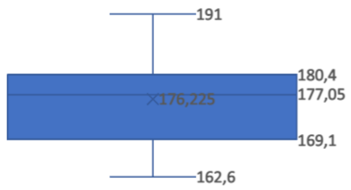
Table 3.3. *Data from 20 participants*



(a) Age



(b) body mass



(c) Height

Figure 3.1. *Box plot for height, body mass and age*

Identification	Height (cm)	Body mass (kg)	Age (yrs)	Gender	EE	HR	VO_2
OK001M	188	78.4	32	M	17.580	163.5	44.848
OK002F	175	59	22	F	14.239	189.4	48.269
OK003F	170	58.1	23	F	12.081	162.7	41.588
OK004F	168.5	54	30	F	12.164	182.3	45.053
OK005M	191	98.6	30	M	19.572	198.537	39.7
OK006F	178	63	26	F	13.342	154.3	42.355
OK007M	176.7	82	27	M	16.079	199.8	39.219
OK008F	177.4	66.3	29	F	14.250	151.2	42.986
OK009F	162.6	57.3	25	F	11.789	173.354	41.148
OK011M	178	89.2	27	M	19.162	186.564	42.965
OK012F	174	67.5	22	F	12.912	168.461	38.260
OK013M	179.8	73.9	29	M	15.811	156.571	42.789
OK014M	179	79	29	M	16.503	190.9	41.779
OK015F	168.8	78.3	23	F	13.015	175.843	33.245
OK016F	175.1	72.4	29	F	17.754	188.662	45.292
OK017M	182	77	34	M	16.912	188.845	43.927
OK019M	187.5	85.5	34	M	18.147	152.759	42.448
OK020M	164.8	85.6	30	M	16.344	198.4	38.187

Table 3.4. *Predictor variables for 18 participants including exercise and questionnaire data*

3.2.1 Evaluation metrics

Model assessment measures are frequently used to assess the model's performance. Machine learning algorithms' efficiency may be tested using evaluation metrics, which are relevant performance indicators. The metrics evaluate the model's performance for unobserved data, or how effectively it predicts values. For regression tasks, the R^2 value, mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE) are the most commonly used terms.

R^2 Value

The R^2 value is an important metric for determining the performance of a machine learning-based regression algorithm. It is sometimes referred to as the coefficient of determination and is represented R^2. It works by determining the amount of variance in the predictions based on the dataset. Simply put, it is the difference between the dataset's samples and the model's predictions. If the R^2 value is one, the model is perfect; otherwise, the model will perform poorly on an unknown dataset. This also implies that the closest the r squared value is to 1, the better trained the model is. Equation 3.1 shows the formula for R^2 value where Y_i is the actual value, $\hat{Y}_i$ is the predicted value and $\bar{Y}_i$ is the mean of all values.

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \bar{Y}_i)^2} \tag{3.1}$$

Mean Absolute Error

Absolute error in machine learning refers to the magnitude of the difference between observation's prediction and its true value. MAE computes the mean of absolute errors given a collection of predictions and observations to calculate the total amount of the errors. Equation 3.2 shows the formula for MAE where N is the number of samples, Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

$$MAE = \frac{\sum_{i=1}^{N} abs(Y_i - \hat{Y}_i)}{N} \tag{3.2}$$

Mean Squared Error

The associations between variables in regression problems are represented by an equation that calculates the distance between predictions and the actual data point. The MSE is a measure of how well the regression curve fits the data as well as how accurate the predictions are. MSE denotes the mean deviation from the actual data. Equation 3.3 shows the formula for MSE where N is the number of samples, Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 \tag{3.3}$$

Root Mean Squared Error

The Root Mean Square Error (RMSE) is a popular statistic for measuring predictive performance. Using the Euclidean distance, it estimates how far the predictions depart from the actual values. it is measured by calculating the residual (the difference between prediction and actual value) for each point, the norm of the residual for each data point, the mean of the residuals, and the square root of that mean to produce the RMSE. Because it uses and takes an actual measurement at each projected data point, the RMSE is frequently utilized in supervised learning applications. Equation 3.4 shows the formula for MAE where N is the number of samples, Y_i is the actual value and $\hat{Y}_i$ is the predicted value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|Y_i - \hat{Y}_i\|^2}{N}} \tag{3.4}$$

Even minor flaws that might lead to an overestimation of the model's inaccuracy are discouraged by RMSE. The RMSE is the preferred assessment metric in this study since it is differentiable and so may be improved. The RMSE is an acceptable statistic when large errors are undesirable due to its sensitivity to outliers. The RMSE is always positive, and the lower the value, the more accurate the model is. The RMSE value reveals when the model is underperforming since the error is squared, and the RMSE value grows when predictions are inaccurate. RMSE does not indicate if the projected values are too high or too low as an assessment metric. To better comprehend the direction of the inaccuracy, predictions must be shown alongside actual results.

4. RESULTS AND DISCUSSION

4.1 Relationship between speed, VO_2 and energy expenditure

The accurate measurement of energy expenditure (EE) is critical in the study of human behaviour. Although the rate of oxygen consumption (VO_2) indicates EE during aerobic metabolism, it is not always practical. An increase in VO_2 is related to higher 24-h EE [39]. During exercise, both oxygen intake and energy expenditure (EE) of working muscles rise. Thus, oxygen use is inextricably tied to energy expenditure [40]. Both oxygen consumption and energy expenditure may be measured directly using expensive laboratory procedures, but both variables can also be correctly calculated using analysis of data such as heart rate level and respiration rate. Cross-sectional research has shown that people with high fitness (as evaluated by VO_2max) are more physically engaged in their everyday lives than people with poor fitness [41]. There is a well-known linear link between oxygen consumption (VO_2) and running speed. The oxygen cost of running increases as the pace of the runner increases. However, an endurance athlete runs at a lower proportion of his or her VO_2max at equivalent submaximal speeds than an untrained individual, yet maintaining a similar VO_2.

As part of the first research question, the relationship between speed, VO_2 and energy expenditure was defined using statistical analysis. It was found that an increase in running speed accounts for an increase in VO_2 which increases energy expenditure.

4.1.1 Statistical Analysis Results

Statistical analysis was done using the data available in the OpenKin folder. VO_2 was analysed for all participants collectively, and gender-wise for identifying the variation in VO_2 among males and females.

The relationship among speed, VO_2 and energy expenditure was identified in two steps. First, the average VO_2 in relation to speed was determined. Data was collected for each participant for six speeds. Mean was taken to calculate the average VO_2 for different speeds. Mean allows to incorporate data from each participant and is the best measure of central tendency. The average VO_2 against each speed for both males and females is shown in Table 4.1. The table shows the increase in velocity resulting in an increase in

VO_2. This is because the rate of oxygen consumption increases as the speed increases.

Velocity	Average VO_2 $(ml\ kg^{-1}min^{-1})$
1.0	10.942
1.3	13.712
1.5	16.140
2.2	30.927
2.5	34.224
2.8	37.498
3.1	40.211
3.3	41.892

Table 4.1. *Average VO_2 over velocity for participants*

Table 4.2 show average VO_2 for males and females respectively. The average VO_2 for both male and female (aged 22 - 34) showed no major difference in values. Several factors can account for variation in VO_2 including metabolism, fitness, body mass etc.

Velocity	Average VO_2 $(ml\ kg^{-1}min^{-1})$ (Female)	Average VO_2 $(ml\ kg^{-1}min^{-1})$ (males)
1.0	11.002	10.864
1.3	13.573	13.851
1.5	15.893	16.360
2.2	30.618	31.274
2.5	33.543	34.905
2.8	37.704	37.292
3.1	40.348	40.073
3.3	42.022	41.762

Table 4.2. *Average VO_2 over velocity for female and male participants*

For better visualisation, the graph was plotted as shown in figure 4.1 for both genders.

No visible difference was found in VO_2 except at velocity V = 2.5, The gender male shows high VO_2 as compared to females. This can be due to the high metabolism and physical fitness of males as compared to females.

The next step involved determining energy expenditure for participants. Energy expenditure (EE) is calculated as

$$EE = (VO_2 * 0.001) * bodymass(kg) \tag{4.1}$$

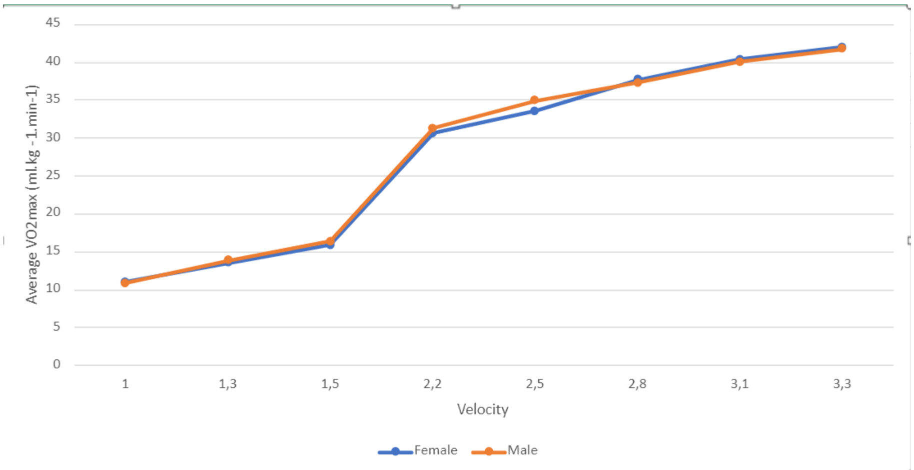


Figure 4.1. *Average VO_2 over velocity for male and female participants*

where EE is in $(kcal\ min^{-1})$ and VO_2 is in $(ml\ kg^{-1}min^{-1})$. VO_2 is multiplied by 0.001 to calculate absolute VO_2 expressed in $L.min^{-1}$.

EE is calculated for each participant for each speed. An average of the EE for each speed for the participants is taken. The results are displayed in Table 4.3. It can be seen that EE increases as speed increase thus showing a linear relationship.

Velocity	Average EE $(kcal\ min^{-1})$
1.0	3.213
1.3	4.567
1.5	5.058
2.2	9.666
2.5	11.389
2.8	10.999
3.1	11.801
3.3	13.883

Table 4.3. *Average EE over velocity for participants*

Table 4.4 show average EE for males and females. Figure 4.2 shows EE against velocity for both genders. It can be seen that energy expenditure increases as velocity increases (VO_2 increases). Males show higher energy produced as compared to females. Haemoglobin levels are another element that contributes to the variance in VO_2 across genders. A study of top cross-country skiers found that female athletes' haemoglobin levels were 10% lower, resulting in lower VO_2 values when compared to male racers [42].

4.2 Performance of Neural Network

The development procedure starts with the creation of a sequential model in Keras, which is then followed by the addition of layers. The input layer is guaranteed to have the right

Velocity	Average EE $(kcal\ min^{-1})$ (Female)	Average EE $(kcal\ min^{-1})$ (Males)
1.0	3.202	3.224
1.3	3.939	5.195
1.5	3.994	6.121
2.2	8.875	10.458
2.5	9.729	13.050
2.8	9.478	12.521
3.1	10.189	13.413
3.3	12.154	15.611

Table 4.4. *Average EE $(kcal\ min^{-1})$ over velocity for male and female participants*

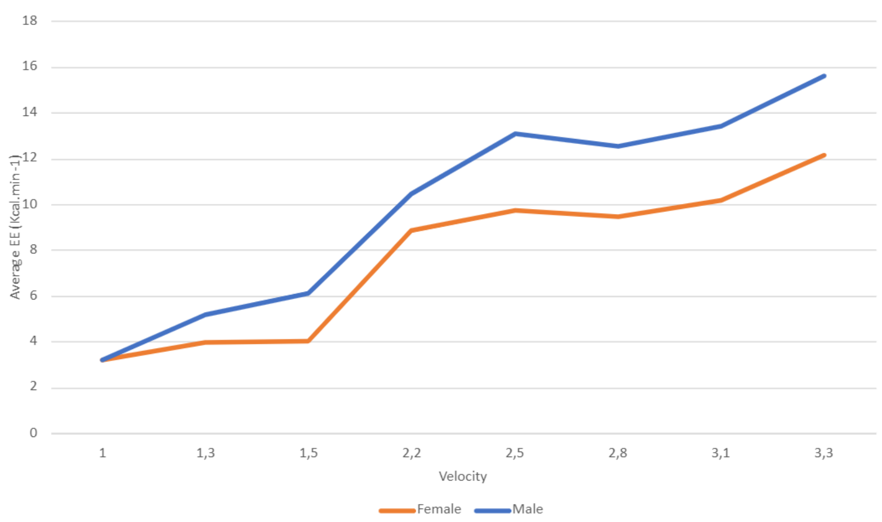


Figure 4.2. *Average EE over velocity for male and female participants*

input size based on batch size, time steps, and feature count. Trial and error experimentation is performed to establish the number and types of hidden layers to create a network large enough to understand the dynamics of the prediction problem. The number of nodes in completely linked layers is determined by the expected output.

The activation functions are influenced by the type of the problem and the kind of layer. For single-output regression, the fully connected layer must have a linear activation function. On higher levels, additional options are offered, and the function is established through trial and error. After determining the network design, the network is built and fitted with Keras using the TensorFlow backend. Before the model can be constructed, several supplementary characteristics must be defined. A loss function is supplied to evaluate a set of weights, and an optimizer is built to discover the network's desired weights. To evaluate the model's accuracy, a metric argument is also defined. The fit function is used to train the model once it has been developed and constructed.

The training phase consists of a set of iterations called epochs that are fed into the model as a parameter. Another method is to keep track of the training duration and stop early if the accuracy of the training outcomes begins to decline. The final training technique parameter to be selected is the batch size. The batch size refers to the number of train samples analyzed before updating the model's weights. Finally, once trained, the network may be utilized to generate predictions on data that has never been seen before.

MLP and LSTM were the first models created. Following preliminary testing, both the LSTM model and the MLP model yielded the most promising results, and the models were further imroved.

4.2.1 Multilayer Perceptron

To anticipate VO_2 levels, an artificial intelligence regressor was developed and used. The most popular type of neural network is the multi-layer perceptron (MLP), which is a feed-forward network constructed of consecutive layers of neurons with no feedback loops between neurons. Each neuron in the input layer gets the information from neurons in the previous layer and generates an activation for the neurons in the layer above it [43]. Finally, the network's outputs are generated by the final layer. The hidden layers occur between the input and output layers. The form of the input data determines the number of neurons in the input layer, whereas the size of the output layer should be proportionate to the number of output classes. Background information and trials guide the selection of the appropriate number of hidden layers. A simple MLP with a single hidden layer is shown in figure 4.3.

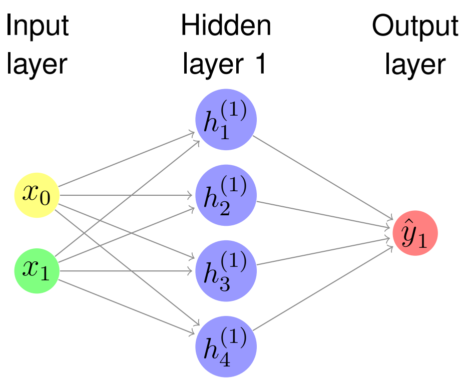


Figure 4.3. *Simple Multilayer Perceptron with 2 Inputs, 1 hidden layer and 1 output*

To predict VO_2 values, a simple and robust multilayer perceptron was developed. The network has six predictor variables, one output variable, and two hidden layers of nodes 32 and 16 with elu activation function and regular kernel initialization. Label Encoder and normalization with mean 0 and variance 1 were done to preprocess the input data. Separating a dataset into two subgroups (training set and test set) is the test train split. The training dataset is used to fit the model. The test dataset is used to test the model. Since, there is no optimal split percentage, It is a good practise to choose a split percentage which gives optimal performance. A test train split of 0.33 (33% of data assigned to test set) was employed. The Adam optimizer with a learning rate of 0.02 was utilized. The loss was calculated using the mean squared logarithmic error. Validation data from the test train split was utilized for cross-validation. Mean squared error and R^2 value were obtained to measure model performance. Figure 4.4 shows the plot of predicted values vs actual values. The model resulted in an RMSE value of 1.1734 R square score of 0.9311.

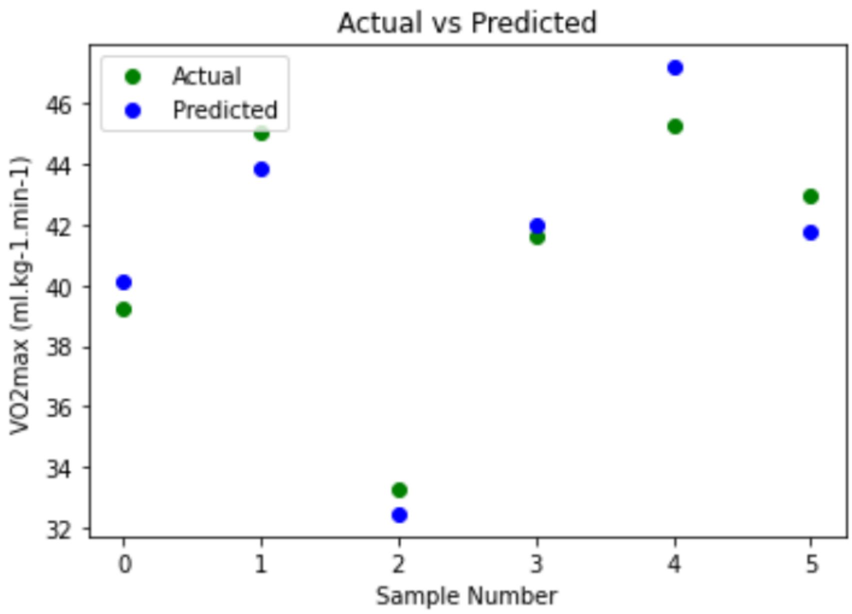


Figure 4.4. *Predicted Vs Actual Values (ANN)*

4.2.2 LSTM based Neural Network

LSTMs are a complex subset of deep learning. Long Short-Term Memory (LSTM) networks are recurrent neural networks capable of learning order dependency in sequence prediction tasks. The Recurrent Neural Network (RNN) is a neural sequence model that excels at key tasks such as language modelling, audio recognition, and machine translation.

An LSTM layer consists of memory blocks that are linked recurrently. These blocks can be compared to differentiable memory chips in a computer system. Each one has one or more recurrently connected memory cells, along with three multiplicative elements – the input, output, and forget gates – that offer continuous equivalents to the cells' write, read, and reset operations. The most critical LSTM hyperparameters to configure are learning rate and network size [31].

The neural network is made up of neurons with long and short term memories, making it ideal for time series analysis and sequence identification. A total of two long-short term memory layers of 32 and 16 neurons, respectively, as well as one output layer of one neuron. Adam was used to train the neural network, which optimizes a mean squared logarithmic loss with a learning rate of 0.01. The entire dataset was cross-validated in 5000 epochs. The batch size (the number of samples transmitted through the neural network) was set at ten. In the building of a neural network, there are no definite and scientifically verified stages to take. However, it is known that the number of layers, number of epochs, number of neurons, and batch size all have an effect on output accuracy and processing time [34]. Therefore, trial and error method (involving multiple iterations of manual adjustment) is used to determine these values until the optimal combination of accuracy and processing time was discovered.

Figure 4.5 shows the plot of predicted values vs actual values. The model resulted in an R^2 value of 0.9562 and an RMSE value of 0.1142.

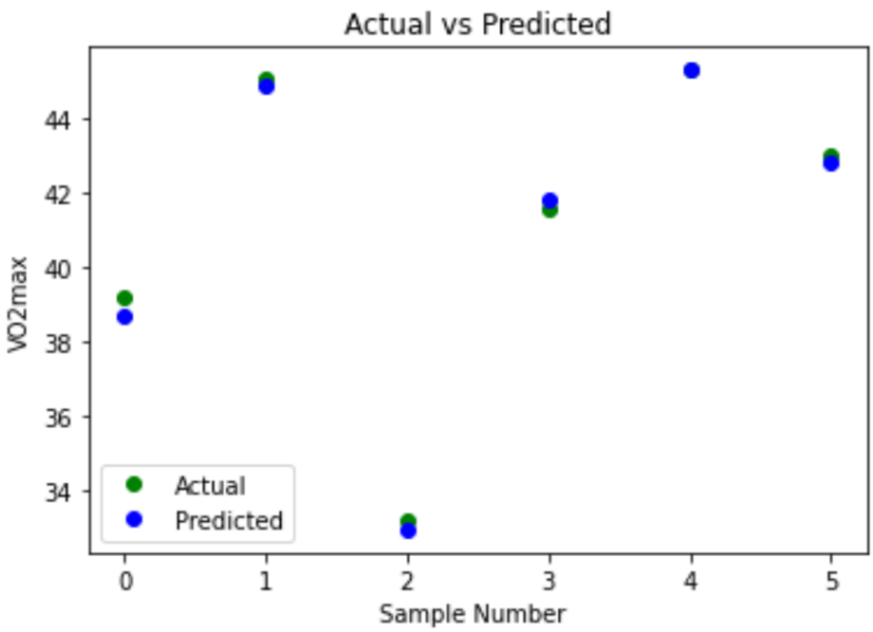


Figure 4.5. *Predicted Vs Actual Values (LSTM)*

4.2.3 Support Vector Machime

Support Vector Machine is based on fundamental concepts from statistical learning theory. The simplicity stems from SVM's use of a basic linear approach to information in a high-dimensional feature space which is not linearly connected to input space. The appeal of the SVM arises from its ease of use as well as its cutting-edge performance on a wide range of learning issues (classification and regression). SVM employs an implicit projection of the training dataset into a high-dimensional feature space created by a kernel function (a function that returns the inner product of two feature space data points).

Because of its high prediction accuracy, SVM is regarded as one of the most promising regression methods, and it is widely used in a variety of application fields. SVMs were created to address classification problems, but they have recently been rebuilt to additionally handle regression problems. SVM creates a hyperplane or group of hyperplanes in a high- or infinite-dimensional space to do a regression analysis. The efficacy of SVM-based models is determined by the kernel function, its parameters, and the regularization parameter C. [44, 45]. The radial basis function kernel, which contains a single optimization parameter gamma, is a popular choice. (γ). Table 4.5 shows the parameter chosen for SVM regressor to predict VO_2. The model resulted in an R^2 value of 0.8231 and an RMSE value of 3.1189.

Parameter	Range
Kernel	Linear
Cost	100
Gamma	Auto

Table 4.5. *Value ranges of the parameters used in the SVM-based model.*

One noteworthy aspect of SVM and comparable kernel-based systems is that, with the right kernel function, one may essentially work in any dimension space without experiencing high processing costs. Another advantage of SVM and kernel approaches is that they allow the creation and application of a kernel for a specific problem, which may be applied straight to data without the requirement for a feature extraction method. This is especially important when the feature extraction approach eliminates a substantial quantity of data structure [44].

4.6 shows the results from SVM.

4.2.4 Multiple Linear Regression

Multiple Linear Regression (MLR) is a technique that is nearly universally employed in sports physiology to study the linear correlations between one or more predictor factors and a single-target variable. Simple linear regression is another name for regression analysis using only one predictor variable [44, 46]. MLR is an expansion of ordinary linear regression in that it may predict a desired target variable using two or more predictor variables in an equation. Figure 4.7 shows the actual and predicted value of VO_2 using MLR. The model resulted in an R^2 value of 0.8499 and an RMSE value of 2.5539.

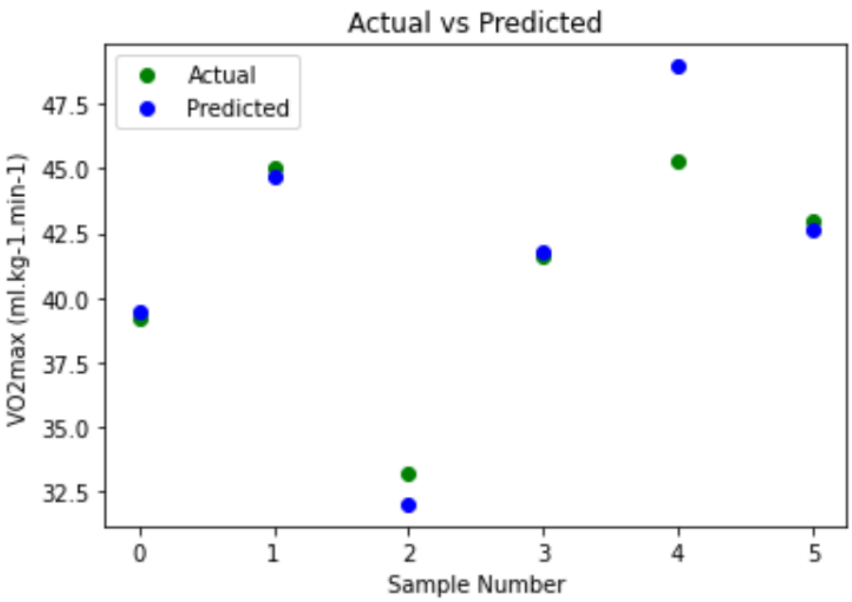


Figure 4.6. *Predicted Vs Actual Values (SVM)*

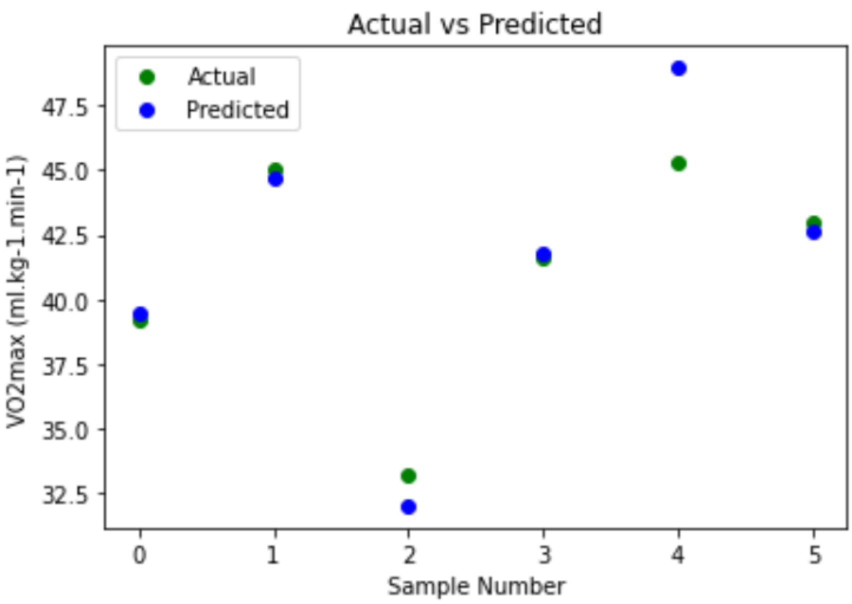


Figure 4.7. *Predicted Vs Actual Values (MLR)*

4.3 Effect of Predictor Variable

Several combinations of predictor factors were used to examine the impact of predictor variables on VO_2 prediction. The model's efficiency was determined by how various predictor variables influenced prediction accuracy while using the same model for each test. Eight VO_2 prediction models were created in total, with the RMSE and R^2 score serving as the assessment measures.

Table 4.6 shows the combinations of predictor variable used for VO_2 prediction and summarises the results of the effect of the predictor variables on the VO_2 prediction.

Model	Predictor Variables	R^2	RMSE
1	Height, body mass, Age, Gender, HR, EE	0.9311	1.7346
2	Height, body mass, Age, Gender, EE	0.8771	2.0928
3	Height, body mass, Age, Gender, HR	0.7382	4.0945
4	Height, body mass, Age, HR, EE	0.9329	1.7034
5	Height, body mass, Gender, HR, EE	0.9291	1.6754
6	Age, Gender, HR, EE	0.9157	1.8325
7	Height, body mass, Age, Gender	0.6782	4.6782
8	Height, body mass, HR	0.7943	3.3702

Table 4.6. *VO_2 Prediction Models*

Figure 4.8 shows the actual vs predicted values of VO_2 for Model 1 containing all predictor variables. Model 1 resulted in an R^2 value of 0.9311 and an RMSE value of 1.734.

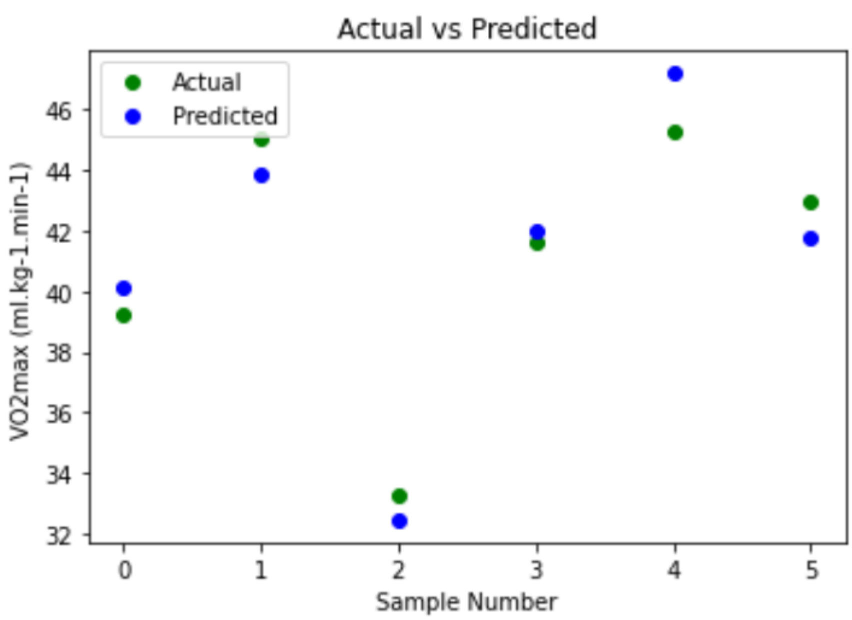


Figure 4.8. *Predicted Vs Actual Values (Model 1)*

Figure 4.9 shows the actual vs predicted values of VO_2 for Model 2 without HR. Model 2 resulted in an R^2 value of 0.87706 and an RMSE value of 2.09277. Model 2 yielded a smaller R^2 value than Model 1, indicating that HR is of significant value in VO_2 prediction.

Figure 4.10 shows the actual vs predicted values of VO_2 for Model 3 without EE. Model 3 resulted in an R^2 value of 0.7382 and an RMSE value of 4.0945. Model 3 yielded a smaller R^2 value than Model 2 and Model 1.

Figure 4.11 shows the actual vs predicted values of VO_2 for Model 4 without gender. Model 4 resulted in an R^2 value of 0.9329 and an RMSE value of 1.7034. Model 4 yielded a similar R^2 and RMSE value as of model 1.

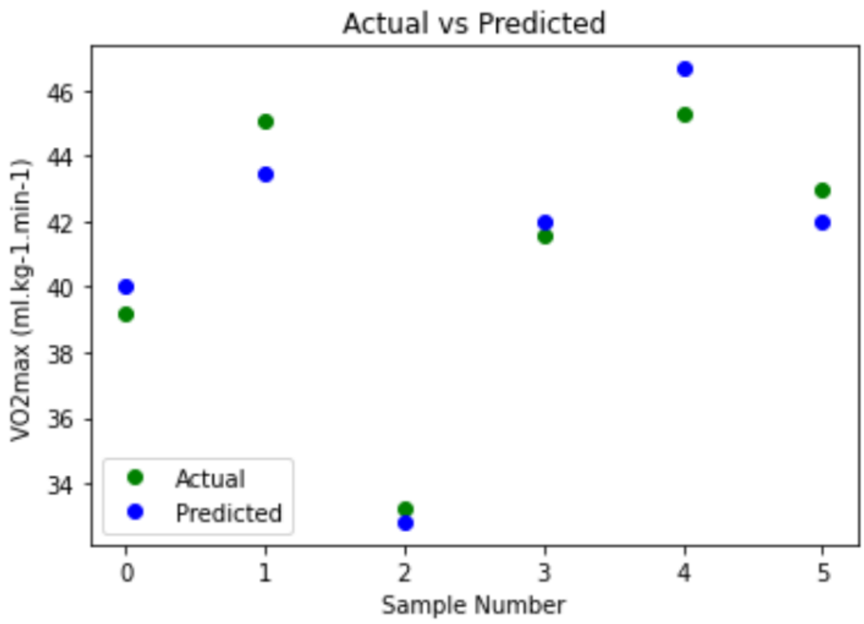


Figure 4.9. *Predicted Vs Actual Values (Model 2)*

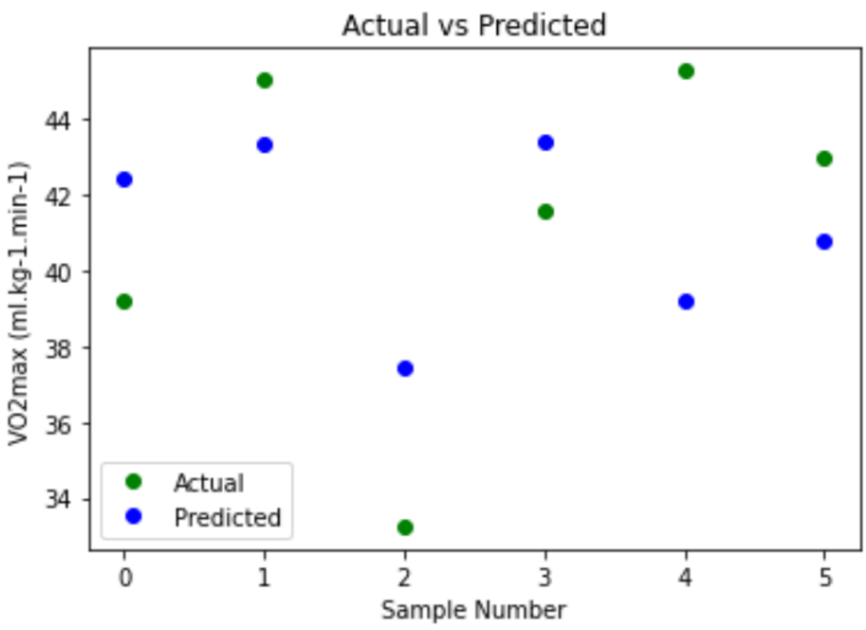


Figure 4.10. *Predicted Vs Actual Values (Model 3)*

Figure 4.12 shows the actual vs predicted values of VO_2 for Model 5 without age. Model 5 resulted in an R^2 value of 0.9291 and an RMSE value of 1.6754. Model 5 yielded a similar R^2 and RMSE value as of model 1 and model 4.

Model 6 (without height and body mass) resulted in an R^2 value of 0.9157 and an RMSE value of 1.8325. Model 6 yielded a similar R^2 and RMSE value as of model 1, model 5 and model 4.

Figure 4.13 shows the actual vs predicted values of VO_2 for Model 7 without HR and EE.

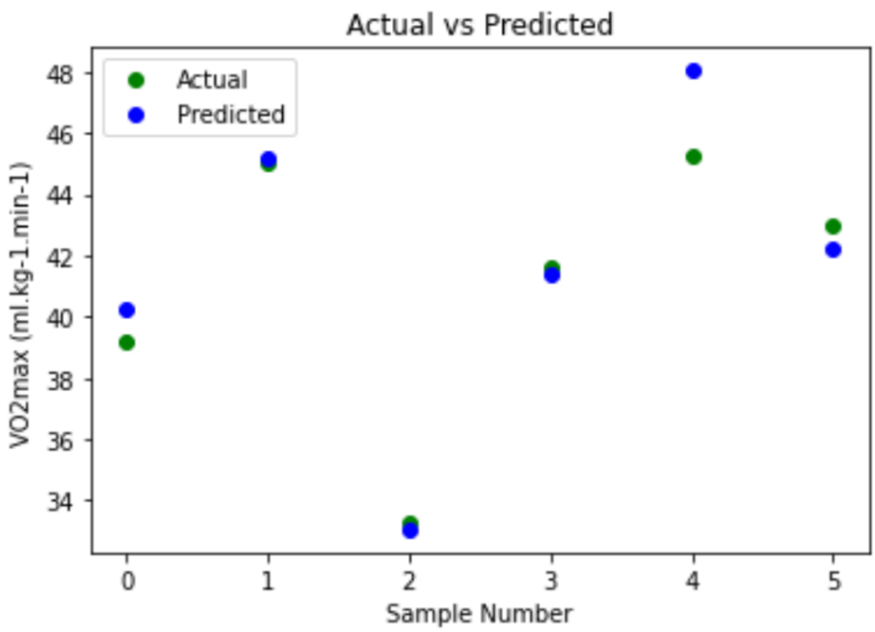


Figure 4.11. *Predicted Vs Actual Values (Model 4)*

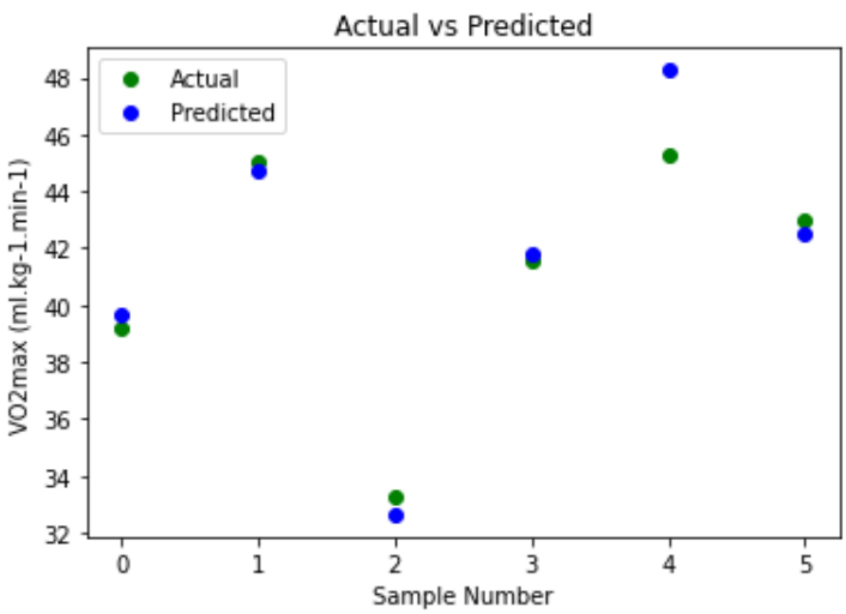


Figure 4.12. *Predicted Vs Actual Values (Model 5)*

Model 7 resulted in an R^2 value of 0.6782 and an RMSE value of 4.6782. Model 7 yielded a lower R^2 and RMSE value as compared to other models.

Figure 4.14 shows the actual vs predicted values of VO_2 for Model 8 containing only height, body mass and HR. Model 8 resulted in an R^2 value of 0.7943 and an RMSE value of 3.3702. Model 8 showed higher RMSE and R^2 values as compared to model 7 but lower values as compared to other models.

From the results, it can be concluded that generally, a higher number of predictor variables

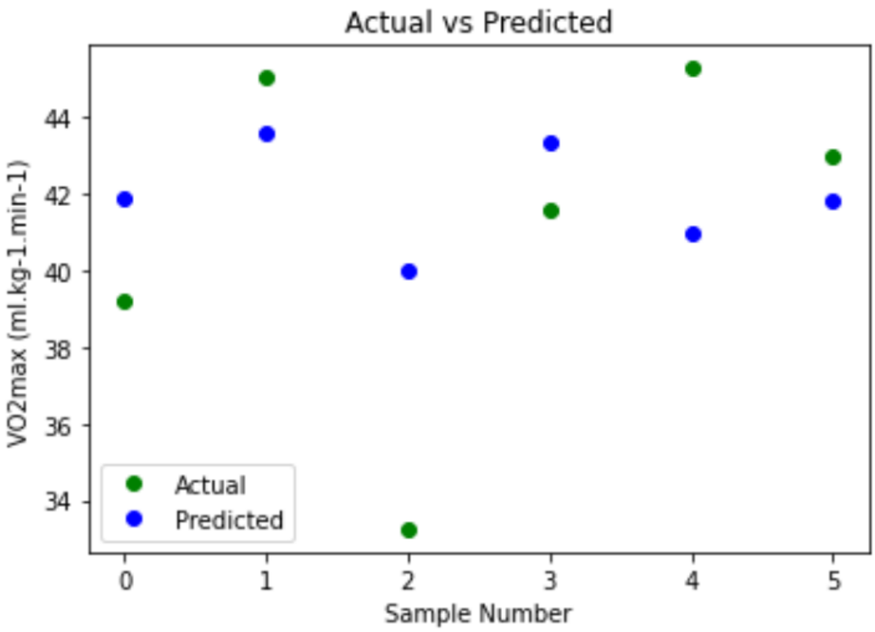


Figure 4.13. *Predicted Vs Actual Values (Model 7)*

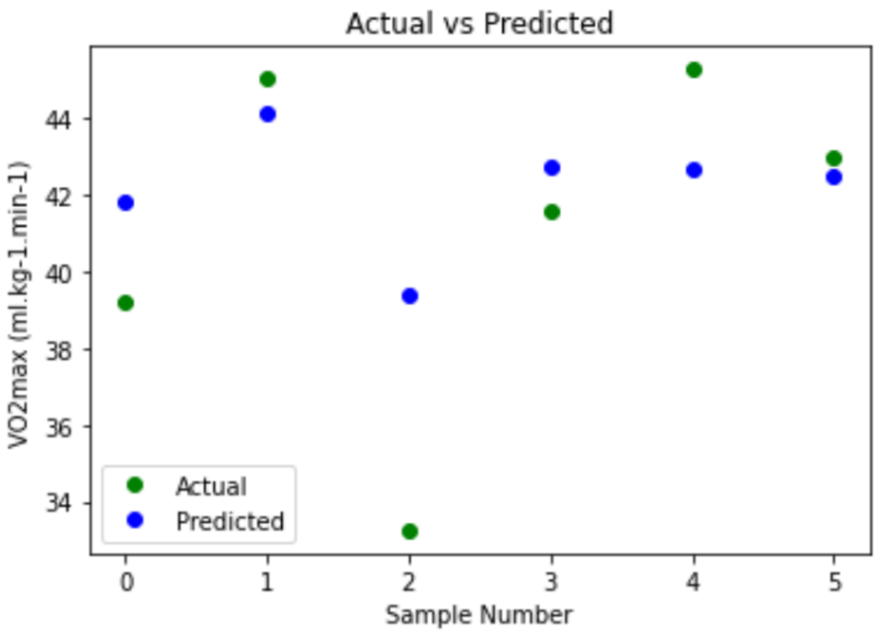


Figure 4.14. *Predicted Vs Actual Values (Model 8)*

result in a higher R^2 value and lower RMSE value. HR is an important predictor variable, as it constitutes for high R^2 value in models 1, 3, 4, 5, 8. Generally, Models without HR as a predictor variable (Model 7, 2) yielded a lower R^2 value as compared to other models. Physiological variables such as age, height, body mass and gender had little effect as compared to HR and EE.

5. LIMITATIONS AND FUTURE RECOMMENDATION

The study's main limitations include its small sample size (n = 18) and the subject pool's limited demographics. As a result, extending the results to other ages and levels of fitness is difficult. Future research should include a broader and more diverse range of topics. Researchers might also experiment with adding new predictor variables to the model (for example, muscle oxygen saturation) or combine a neural network with other methods of machine learning to predict VO_2max.

In this thesis, fewer characteristics have been included in the predictive model's training and testing to minimize its complexity. The acquired data went through a processing phase. Outliers (inaccurate data) and less significant characteristics (those that have no effect on performance) were deleted from the training dataset during this step. In the future, the feature selection method can allow for quicker training of the ML algorithm, simpler interpretation of the prediction model, and, lastly, it can decrease overfitting and time limitation.

Furthermore, the techniques and methodology employed in this thesis may be easily extended to other exercise modes, such as cycling, as long as inputs relating to exercise intensity and cardiopulmonary system response were available.

Accurate VO_2max measurement without the requirement for a maximal cardiopulmonary exercise test would greatly expand VO_2max accessibility and perhaps allow it to become a critical sign.

Further investigation into the ML-guided prediction of VO_2 is still required to develop a more accurate model with greater predictability than those that have already been developed. In the context of forecasting VO_2 values, the pilot research discovered that a recurrent neural network (LSTM layers) can involve large amounts of data from various mechanical and physiological variables (such as heart rate and breathing frequency), along with historical input values, to produce good VO_2max forecasts. For this purpose, the current model can further be developed into a recurrent neural network. The time-based data available in the OpenKin folder can be fed into the recurrent neural network for VO_2max prediction. However, the factor of generalisation should be kept in mind.

This algorithm has the potential to be implemented in a portable device in the future and to enable real-time assessment of individual VO_2 during training. This can form another thesis topic. According to my knowledge, this work has not been done previously and the integration of a neural network-based model in a wearable device can be a novel advancement in sports science.

6. CONCLUSION

VO_2max is often recognized as the most reliable predictor of overall health and fitness. It is a great predictor of the risk of cardiovascular disease. This thesis explored several approaches related to oxygen uptake (VO_2) prediction and the use of machine learning to predict VO_2 during physical activities. Successful prediction models are developed using ML algorithms to determine VO_2 values without the need for the completed procedures associated with direct approaches. These models are frequently built using data from exercise, non-exercise, or hybrid models.

From the statistical analysis presented in Section 4.1, it was seen that speed, VO_2 and energy expenditure have a direct relationship hence proving Hypothesis 1. Males generally, show high energy produced as compared to females. This can be due to a number of factors such as fitness and haemoglobin levels etc.

Using predictor variables such as HR and physiological data, this research developed a simple and robust neural network model for VO_2 prediction and demonstrated that an ANN can reliably predict VO_2 responses while walking or running at various sub-maximal intensities. This approach has the potential to be developed in the future to predict VO_2 responses using more predictor variables. The ANN model was compared with LSTM based model. It was seen that an LSTM based model works well for time series data and can give more accurate performance as compared to a simple neural network. The ANN-based model is the best VO_2 prediction model was compared to SVM and MLR. Hence, proving Hypothesis 2 true. Different SVM kernel functions were used and, it was discovered that the RBF kernel function produces more accurate results than the linear kernel function. The neural network-based models outperformed SVM and MLR based models in terms of VO_2 prediction values.

Eight different combinations of predictor variables were used to observe the effect of the number of predictor variables on the performance of ANN. It was seen that a higher number of predictor variables resulted in higher R^2 and RMSE value, hence, proving Hypothesis 3.

According to the findings of this thesis on the prediction of the VO_2, Using the ML approach, it is possible to achieve decent predictability while taking into account a number of factors. Data pre-processing is required to identify potential characteristics (predictor variables). To prevent overfitting, a dropout layer was included. The effect of predictor variables was also examined, It was seen that a higher number of predictor variables resulted in high RMSE and R^2 values. Certain variables such as HR are important for VO_2 prediction.

REFERENCES

[1] Shandhi, M. M. H., Bartlett, W. H., Heller, J. A., Etemadi, M., Young, A., Plötz, T. and Inan, O. T. Estimation of Instantaneous Oxygen Uptake During Exercise and Daily Activities Using a Wearable Cardio-Electromechanical and Environmental Sensor. *IEEE Journal of Biomedical and Health Informatics* 25.3 (2021), pp. 634–646. DOI: 10.1109/JBHI.2020.3009903.

[2] Ross, R., Blair, S. N., Arena, R., Church, T. S., Després, J.-P., Franklin, B. A., Haskell, W. L., Kaminsky, L. A., Levine, B. D., Lavie, C. J., Myers, J., Niebauer, J., Sallis, R., Sawada, S. S., Sui, X. and Wisløff, U. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* 134.24 (2016), e653–e699. DOI: 10.1161/CIR.0000000000000461.

[3] Barbara Strasser, M. B. Survival of the fittest: VO2max, a key predictor of longevity?: *FBL* 23.8 (2018), pp. 1505–1516. DOI: 10.2741/4657.

[4] Buttar, K. K., Saboo, N. and Kacker, S. A review: Maximal oxygen uptake (VO2 max) and its estimation methods. *IJPESH* 6 (2019), pp. 24–32.

[5] Leibetseder, V., Ekmekcioglu, C. and Haber, P. A simple running test to estimate cardiorespiratory fitness. *Journal of Exercise Physiology* 5.3 (2002), pp. 6–13.

[6] Abut, F., Akay, M. F. and George, J. Developing new VO2max prediction models from maximal, submaximal and questionnaire variables using support vector machines combined with feature selection. *Computers in biology and medicine* 79 (2016), pp. 182–192.

[7] Abut, F., Akay, M. and GEORGE, J. A robust ensemble feature selector based on rank aggregation for developing new VO2max prediction models using support vector machines. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING COMPUTER SCIENCES* 27 (Sept. 2019), pp. 3648–3664. DOI: 10.3906/elk-1808-138.

[8] Ashfaq, A., Cronin, N. and Müller, P. Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction: A review. *Informatics in Medicine Unlocked* 28 (2022), p. 100863. ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2022.100863. URL: https://www.sciencedirect.com/science/article/pii/S235291482200017X.

[9] Comana, F. *The value of vo2 – health measure or performance marker?* URL: https://blog.nasm.org/sports-performance/the-value-of-vo2-health-measure-or-performance-marker.

[10] Daley, J. *Vo2 and VO2max*. URL: http://www.shapesense.com/fitness-exercise/articles/vo2-and-vo2max.shtml.

[11] Firstbeat. *What's a good vo2max for me? fitness, age, men and women.* Mar. 2022. URL: https://www.firstbeat.com/en/blog/whats-a-good-vo2max-for-me-fitness-age-men-and-women/.

[12] Deka, P., Pozehl, B. J., Pathak, D., Williams, M., Norman, J. F., Alonso, W. W. and Jaarsma, T. Predicting maximal oxygen uptake from the 6 min walk test in patients with heart failure. *ESC heart failure* 8.1 (2021), pp. 47–54.

[13] Leitner, J. From Vision to Actions: Towards Adaptive and Autonomous Humanoid Robots. PhD thesis. May 2015.

[14] Hu, Y. H. and Hwang, J.-N. *Handbook of neural network signal processing*. 2002.

[15] SheetalSharma and SheetalSharma. *Artificial Neural Network (ANN) in Machine Learning*. Aug. 2017. URL: https://www.datasciencecentral.com/artificial-neural-network-ann-in-machine-learning/.

[16] Borror, A., Mazzoleni, M., Coppock, J., Jensen, B. C., Wood, W. A., Mann, B. and Battaglini, C. L. Predicting oxygen uptake responses during cycling at varied intensities using an artificial neural network. *Biomedical Human Kinetics* 11.1 (2019), pp. 60–68.

[17] Benardos, P. and Vosniakos, G.-C. Optimizing feedforward artificial neural network architecture. *Engineering applications of artificial intelligence* 20.3 (2007), pp. 365–382.

[18] London and Fountas, Z. Imperial College Spiking Neural Networks for Human-like Avatar Control in a Simulated Environment. (Mar. 2022).

[19] Nabi, J. *Recurrent neural networks (rnns)*. July 2019. URL: https://towardsdatascience.com/recurrent-neural-networks-rnns-3f06d7653a85.

[20] Yin, C., Rosendahl, L. and Luo, Z. Methods to improve prediction performance of ANN models. *Simulation Modelling Practice and Theory* 11.3-4 (2003), pp. 211–222.

[21] Rathod, J. *Underfitting, overfitting, and regularization*. Sept. 2021. URL: https://jashrathod.github.io/2021-09-30-underfitting-overfitting-and-regularization/.

[22] Basheer, I. A. and Hajmeer, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods* 43.1 (2000), pp. 3–31.

[23] Brownlee, J. *How to configure the learning rate when training deep learning neural networks*. Aug. 2019. URL: https://machinelearningmastery.com/learning-rate-for-deep-learning-neural-networks/.

[24] Uhle, K. *Predicting stock market liquidity using neural networks*. Nov. 2020. URL: https://trepo.tuni.fi/handle/10024/123706.

[25] Budhiraja, A. *Learning less to learn better-dropout in (deep) machine learning*. Mar. 2018. URL: https://medium.com/@amarbudhiraja/https-medium-com-

amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5.

[26] Team, K. *Keras Documentation: Earlystopping*. URL: https://keras.io/api/callbacks/early_stopping/.

[27] Akay, M. F., Zayid, E. I. M., Aktürk, E. and George, J. D. Artificial neural network-based model for predicting VO2max from a submaximal exercise test. *Expert Systems with Applications* 38.3 (2011), pp. 2007–2010.

[28] Jensen, K., Frydkjær, M., Jensen, N. M., Bannerholt, L. M. and Gam, S. A maximal rowing ergometer protocol to predict maximal oxygen uptake. *International Journal of Sports Physiology and Performance* 16.3 (2021), pp. 382–386.

[29] Park, H.-Y., Jung, H., Lee, S., Kim, J.-W., Cho, H.-L. and Nam, S.-S. Estimated Artificial Neural Network Modeling of Maximal Oxygen Uptake Based on Multistage 10-m Shuttle Run Test in Healthy Adults. *International Journal of Environmental Research and Public Health* 18.16 (2021). ISSN: 1660-4601. DOI: 10.3390/ijerph18168510. URL: https://www.mdpi.com/1660-4601/18/16/8510.

[30] Beltrame, T. Prediction of Oxygen Uptake and its Dynamics by Wearable Sensors During Activities of Daily Living. (2016).

[31] Zignoli, A., Fornasiero, A., Ragni, M., Pellegrini, B., Schena, F., Biral, F. and Laursen, P. B. Estimating an individual's oxygen uptake during cycling exercise with a recurrent neural network trained from easy-to-obtain inputs: A pilot study. *Plos one* 15.3 (2020), e0229466.

[32] Alzamer, H., Abuhmed, T. and Hamad, K. A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications. *Electronics* 10.16 (2021), p. 1956.

[33] De Brabandere, A., Op De Beéck, T., Schütte, K. H., Meert, W., Vanwanseele, B. and Davis, J. Data fusion of body-worn accelerometers and heart rate to predict VO2max during submaximal running. *PloS one* 13.6 (2018), e0199509.

[34] De Brabandere, A., Op De Beéck, T., Schütte, K. H., Meert, W., Vanwanseele, B. and Davis, J. Data fusion of body-worn accelerometers and heart rate to predict vo2max during submaximal running. *PLOS ONE* 13.6 (2018). DOI: 10.1371/journal.pone.0199509.

[35] Bradshaw, D. I., George, J. D., Hyde, A., LaMonte, M. J., Vehrs, P. R., Hager, R. L. and Yanowitz, F. G. An accurate VO2max nonexercise regression model for 18–65-year-old adults. *Research quarterly for exercise and sport* 76.4 (2005), pp. 426–432.

[36] Akay, M. F., Inan, C., Bradshaw, D. I. and George, J. D. Support vector regression and multilayer feed forward neural networks for non-exercise prediction of VO2max. *Expert Systems with Applications* 36.6 (2009), pp. 10112–10119.

[37] Ashfaq, A., Cronin, N. and Müller, P. *Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction : A Review*. Jan. 1970. URL: https:

//jyx.jyu.fi/handle/123456789/79610?show=full&locale-attribute=en.

[38] Sharma, D. *Application of machine learning methods for human gait analysis*. Aug. 2019. DOI: 10.13140/RG.2.2.33265.40805.

[39] Ando, T., Piaggi, P., Bogardus, C. and Krakoff, J. VO2max is associated with measures of energy expenditure in sedentary condition but does not predict weight change. *Metabolism* 90 (2019), pp. 44–51.

[40] D'silva, L., Cardew, A., Qasem, L., Wilson, R. and Lewis, M. Relationships between oxygen uptake, dynamic body acceleration and heart rate in humans. *The Journal of Sports Medicine and Physical Fitness* 55.10 (2014), pp. 1049–1057.

[41] Kalyanshetti, S. and Veluru, S. A cross-sectional study of association of body mass index and VO2 max by nonexercise test in medical students. *National Journal of Physiology, Pharmacy and Pharmacology* 7 (Jan. 2016), p. 1. DOI: 10.5455/njppp.2017.7.0825804092016.

[42] L. B. P. Bara, C. *Journal of Exercise Physiology Online*. URL: https://www.asep.org/asep/asep/JEPonlineFEBRUARY2019_Bara.pdf.

[43] Eneh, L. *Neural networks in household investor behavior prediction*. May 2020. URL: https://trepo.tuni.fi/handle/10024/122081.

[44] Akay, F. and Abut, F. Machine learning and statistical methods for the prediction of maximal oxygen uptake: Recent advances. *Medical Devices: Evidence and Research* (2015), p. 369. DOI: 10.2147/mder.s57281.

[45] Akay, M. F., Çetin, E., Yarım, İ., Bozkurt, Ö. and Özçiloğlu, M. M. development of novel maximal oxygen uptake prediction models for Turkish college students using machine learning and exercise data. *2017 9th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE. 2017, pp. 186–189.

[46] Akay, M., Cetin, E., YARIM, İ. and ÖZÇİLOĞLU, M. New prediction models for the maximal oxygen uptake of college-aged students using non-exercise data. *New Trends and Issues Proceedings on Humanities and Social Sciences* 4.4 (2017).