Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 45 No. 1 (2024)

Selection of Most Relevant Input Parameters Using WEKA for Artificial Neural Network Based Oar-Based Water Turbine Harvester Hydrokinetic Efficiency Prediction Models

Mr. V. Prabhu [1*], Dr. V.B.M. Sayana[2]

[1] Associate Professor & Head, Department of Civil Engineering, Selvam College of Technology, Namakkal – 637003, Tamil Nadu, India, snvprabhu@gmail.com, +91 9842370854*

[2]Professor & Head, Department of Civil Engineering, St.Peters Insitute of Higher Education & Research, Avadi, Chennai – 600077, Tamil Nadu, India, vbms19@gmail.com, +91 9444288899

Abstract:- Accurate hydrokinetic efficiency prediction plays a crucial role in renewable energy research, particularly in assessing the effectiveness of small water turbine harvesters. Various models, including both conventional and Artificial Neural Network (ANN) approaches are being utilized for hydrokinetic efficiency prediction. External factors and geographical variables significantly influences these predictions, underscoring the importance of identifying relevant variables for precise forecasts. To address this, the Waikato Environment for Knowledge Analysis (WEKA) software analyzed 25 diverse water sources with distinct environmental factors to pinpoint influential input parameters for ANN-based hydrokinetic efficiency prediction. Key parameters identified included turbine hydraulic power, shaft mechanical power, turbine efficiency, the ratio of blade length to blade width, blade axle length, and blade inclination angle. Three ANN models (ANN-1, ANN-2, and ANN-3) were developed, with maximum Mean Absolute Percentage Errors (MAPE) of 19.36%, 11.29%, and 8.31%, respectively. Impressively, the ANN-3 model, incorporating specific input variables, demonstrated a 12.67% enhanced prediction accuracy compared to ANN-1 and ANN-2. WEKA identified blade width, blade axle length, blade inclination angle, and turbine efficiency as the most relevant input variables, culminating in a robust hydrokinetic efficiency prediction of 29.6%.

Keywords*: hydrokinetic efficiency, ANN, ore based water harvester, prediction models.*

1. Introduction

Hydropower is an environmentally friendly source of energy that gets its supply from the sun-driven, perpetually renewing water cycle. It is an environmentally beneficial energy source because of its dependency on water. The conventional method of using hydropower entails building large reservoirs and dams to create a significant water head that allows for the production of electricity with efficiency. Nonetheless, the possibility of small-scale hydropower systems is becoming increasingly recognized, especially when considering minor flow canals and rivers. This realization stems from the desire to extend the advantages of hydropower to more rural or economically disadvantaged areas, as well as environmental and social problems related to large-scale dams.

Due to a number of benefits, small hydrokinetic turbines intended for use in rivers and canals with moderate flow are becoming more and more popular as an alternative to fossil fuels. Small-scale hydrokinetic turbines present an environmentally benign and more cost-effective alternative to huge dams, which can be costly and destructive to the environment. huge dams also require extensive infrastructure. The unhindered flow of fish and other aquatic species is ensured by their operation without impounding water, which is a vital advantage over traditional hydropower plants where dams can obstruct fish migration pathways and have negative effects on local aquatic ecosystems.

Because hydrokinetic turbines can run in partially submerged locations, they provide a variety of installation possibilities. Turbines of this design can run effectively above and below the water's surface. In rivers or tidal environments, partial submersion improves flexibility to changing levels of water and flow conditions. By harnessing the kinetic energy of flowing water, the turbines minimise their negative environmental effects while producing electricity. Because of their adaptable deployment, they may be used in a variety of aquatic habitats and contribute to decentralised, sustainable energy solutions that cause less disturbance to the environment and the visual arts.

For effective electricity generation from flowing water, small-scale hydrokinetic turbines rely on a number of crucial characteristics for proper operation. One important aspect that affects the amount of kinetic energy that can be extracted is the water's flow velocity. For turbines to function efficiently, they have to run within certain velocity ranges. The amount of energy that can be captured is directly impacted by water density, another crucial factor. The efficiency of a turbine is largely dependent on its design, which includes the form and arrangement of its blades. This allows the turbine to convert kinetic energy into mechanical power. The area that the revolving blades cover, or the swept area, affects how much water is caught and how much energy is extracted. The turbine's operational boundaries are defined by its cut-in and cut-out speeds, guaranteeing safe and efficient operation.

The overall efficiency of the system depends on the generator efficiency, which quantifies the transformation of mechanical power into electrical power. To survive exposure to water, materials used in the construction of turbines must be robust and resistant to corrosion. Fish-friendly designs facilitate safe fish passage by taking the effects on aquatic life into account. When designing and implementing small-scale hydrokinetic turbines, deployment depth, modularity, scalability, environmental effect, operational maintenance demands, and integration with power electronics are all important factors to take into account. For energy solutions to be sustainable, effective, and ecologically responsible, these factors must be balanced.

It is essential to research hydrokinetic turbine parameters in order to create effective, long-term energy solutions. Optimising design and performance requires an understanding of variables like flow velocity, turbine efficiency, and environmental effect. By guiding turbine adaptation to changing water conditions, these parameters maximise energy extraction and foster reliability. Making turbines that are both efficient and environmentally benign requires balancing factors including material durability, fish-friendly design, and operational requirements. In order to maximise ecological impact, minimise the viability of hydrokinetic energy, and improve the overall effectiveness of small-scale hydrokinetic systems, a thorough understanding of these characteristics is essential. Numerous studies have been conducted as a result of the expanding significance of hydrokinetic small-scale turbine harvesters.

In order to determine an appropriate power prediction technique for horizontal axis hydrokinetic turbines, Andrei et al. (2017) carried out research. Using an open-source framework for modelling and design, the study made use of operational data from literature, laboratory testing, and simulation. The best blade geometry for tiny rotor models was found through the investigation, and these models were evaluated in an experimental bench specifically designed for axial hydraulic turbine models. The study modified the performance curve power coefficient vs. tip speed ratio to forecast the extracted power for a range of hydrokinetic turbines with varying diameters but comparable geometry using the data and findings gathered.

A numerical and experimental investigation on a scale-model Horizontal Axis Hydrokinetic Turbine (HAHT) was carried out by Javaherchi et al. (2017). The turbine's modified blade shape was intended to reproduce the RM1's planned Cp–TSR performance curve at particular Reynolds numbers. It was based on the U.S. Department of Energy Reference Model 1 (RM1). Planar particle image velocimetry was used to get new insights, and a load cell and magnetic angular encoder were used to measure wake structure and performance. The study used numerical solutions of the RANS equations to examine the wake evolution and the rotor flow field. Insights into characterising HAHT and recommendations for verifying numerical tools with experimental data were obtained by comparing the results of experiments and numerical calculations.

In order to forecast scour caused by marine hydrokinetic turbines on erodible bed surfaces in fluvial or tidal environments, Musa et al. (2018) created a modelling framework. The paper describes flow structures affecting equilibrium scour depth conditions at the turbine base by applying the phenomenological theory of turbulence

and utilising recent developments in bridge scour formulation. The theoretical model uses dimensionless parameters to estimate scour depth by connecting turbine operating conditions to flow structures and scour depth through drag force. Different sediment mobility regimes, turbine configurations, hydraulic settings, bed materials, and migrating bedform types are all included in the validation process at the laboratory scale. This work provides a physics-based prediction formula for the local scour depth, which is useful for flow energy conversion technologies' system design anchoring studies and feasibility assessments.

A horizontal axis micro-hydrokinetic river turbine (HAMHRT) was created by Wang et al. (2019) for use in regional renewable energy projects. In order to develop a 2 m diameter, three-bladed HAMHRT, the research involved choosing a hydrofoil shape, evaluating hydrodynamic and cavitation characteristics, and maximising chord length and twist angle for various blade placements. After that, hydrodynamic analysis under several operating conditions was conducted using numerical computer modelling, which revealed a maximum efficiency of 25.2% at a river current speed of 0.8 m/s, 4° pitch angle, and a tip speed ratio of 6. Despite variations in tip speed ratio from design values and varying current speeds, the rotor exhibited consistent power production and efficiency.

Riglin et al. (2021) constructed and tested a preliminary hydrokinetic turbine prototype intended for use in rivers. The prototype was tested in a circulating water channel after extensive blade characterization and optimization investigations using computational fluid dynamics (CFD) simulations. Channel flow rates were varied between 1.0 and 1.7 m/s during the tests, while generator loading was manually changed. In conjunction with the ground renewable energy system (GREENS), a solar charging device replicated the operation of a turbine. During manual generator loading, validation using CFD predictions produced a peak power coefficient of 0.37 at a tip speed ratio of 2.50%. With a relative inaccuracy of less than 3.0% when taking into consideration different component losses in numerical forecasts, the prototype showed that it could be successfully integrated with GREENS for portable applications.

Studies on the Savonius Hydro-Kinetic Turbine (SHKT), renowned for its affordable design and possibility for local manufacture, were carried out by Rengma et al. (2022). Through 3D CFD simulations, artificial neural network (ANN)-augmented optimisation, and testing, an optimised semicircular SHKT geometry was developed, positioning it as a feasible option for off-grid power generation in remote and hilly places. After identifying the factors impacting the power coefficient (Cp) through CFD research, the ANN was trained to optimise the blade parameter. The optimised blade was verified by experiments, which showed that better performance was suggested by aspect ratios between 1.4 and 2.0 and overlap ratios between 0.15 and 0.2. A maximum Cp of 0.194 at a TSR of 0.8 was obtained with a blade arc angle of 166°.

The Savonius hydrokinetic turbine, a vertical axis turbine used to generate electricity in rivers and canals, was explored by Khani et al. (2023). The power coefficient (CP) was used to gauge the turbine's performance. The study used a number of soft computing techniques, taking into account a variety of input variables, including aspect ratio, overlap ratio, blockage ratio, number of blades, blade arc angle, blade form factor, twist angle, Reynolds number, and tip speed ratio (TSR). When compared to other statistical indices, CatBoost proved to be the most effective strategy for estimating the power coefficient. Aspect ratio was found to be the most important design parameter by sensitive analysis, which suggests CatBoost for precise forecasts taking into account a variety of input variables in Savonius hydrokinetic turbines.

The focus of Paturi et al. (2022) was on hydrokinetic turbines, which are known for their exceptional efficiency in hydropower applications. A hydrokinetic turbine's physical dimensions and operating characteristics affect its efficiency. In order to evaluate an Archimedes screw turbine (AST), the study used a standalone artificial neural network (ANN) with a graphical user interface (GUI) due to the intricate and nonlinear interaction between geometric configuration and output. Axle length, blade stride, blade angle, and diameter ratio were the input variables, while the only output was the power coefficient (Cp). The 4-3-1 design showed the lowest average error (0.0211) and root mean square error (RMSE) (0.0008) among different neural network topologies. The predictions showed good agreement with data from second-order regression models (SORM) and computational fluid dynamics (CFD). The effect of AST variables on the efficiency of hydropower production was measured using a virtual hydropower system, with diameter ratio being the most important parameter (84%). The

relationship between AST's geometric configuration and hydropower production efficiency was accurately analysed by the created model.

Blade Element Momentum (BEM) theory, which is frequently used to forecast the performance of horizontal axis conversion systems like wind and water turbines, was assessed by Abutunis et al. (2019). Though conceptually straightforward, BEM theory has a number of convergence problems. To get over these convergence problems, the study proposed multilayer perceptron (MLP) neural networks (NNs) as a computational intelligence method. In order to improve BEM accuracy in determining lift and drag coefficients over a range of local Reynolds numbers, NNs were also used for multivariate interpolation. The BEM-NNs model outperformed models with a constant Reynolds number at greater tip speed ratios without experiencing convergence issues, providing better power prediction.

Wind power, a well-liked renewable energy source, has an extractable value that is restricted to 0.593, also referred to as the Betz-Joukowsky limit. Siavash et al. (2021) investigated this topic. Given that wind power is directly proportional to cubic wind speed, a small increase in wind speed results in a noticeable rise in power output. The research went beyond the Betz limit by introducing a controlled duct with adjustable shrouding angles. Due to the time and financial limits of wind tunnel research, the study evaluated just four duct opening angles. Multiple linear regression and an artificial neural network were used to forecast turbine performance. For turbine power generation and rotor angular speed under various wind velocities and duct opening angles, both models were recommended. When calculating the turbine power curve and rotor speed, the created neural network performed better than the regression model. The paper proposed the neural network model as a lookup table for regulating turbines equipped with the novel mechanism, emphasising the intense importance of shrouding angle on turbine performance at higher wind velocities.

After reviewing previous literature and conducting significant investigations, it was observed that there is a scarcity of studies focusing on the development of Artificial Neural Network (ANN) based prediction models specifically for predicting the overall efficiency of ore-based small-scale water turbines. The lack of literature in this specific domain prompted the undertaking of this research. The primary objective here was to construct ANN-based prediction models tailored to the unique characteristics of Ore-Based Water Turbine Harvesters.

2 Materials & Methods

2.1 Source of data

The study utilized 25 distinct small water flowing regions situated across various climatic zones in Tamil Nadu, India, for both training and testing purposes in Artificial Neural Network (ANN) models, as outlined in Table 1. Figure 1 illustrates the configuration of the Oar-Based Water Turbine Harvester employed in the investigation.

Table 1. Details of regions selected for training and testing of the Oar-Based Water Turbine Harvester prediction ANN models

Data No	District	Region – Canals with Small water flow	Latitude	Longitude	Average Flow rate (AFR) (m^3/s)
1	Tanjore	Pookara Grand Anaikut	10°46'21.7"N	79°08'52.5"E	0.015 – 0.215
2	Tanjore	Othaipalam	10°46'50.2"N	79°07'44.5"E	0.026 – 0.194
3	Tanjore	Seappanavari Palam	10°46'50.6"N	79°07'29.4"E	0.019 – 0.206
4	Tanjore	Padithurai	10°46'52.7"N	79°07'01.1"E	0.012 – 0.216
5	Tanjore	Peacock River View Cross	10°45'41.9"N	79°04'30.0"E	0.032 – 0.211
6	Tanjore	Alakudi Cross	10°45'04.5"N	79°03'13.2"E	0.029 – 0.231
7	Nagapattinam	Vettaru Bridge	10°49'21.0"N	79°50'17.1"E	0.019 – 0.184
8	Nagapattinam	Narimanam Small Canal	10°48'55.2"N	79°47'41.7"E	0.022 – 0.174
9	Nagapattinam	Thundam Bridge Canal	10°48'25.8"N	79°45'59.6"E	0.026 – 0.156
10	Nagapattinam	Anaimangalam Kohur Canal	10°47'55.5"N	79°44'21.2"E	0.029 – 0.178
11	Nagapattinam	Poolangudi Poovai Bridge	10°47'20.1"N	79°42'58.1"E	0.017 – 0.157
12	Nagapattinam	Palayavalam Vettaru Bridge	10°47'39.8"N	79°40'24.3"E	0.022 – 0.211
13	Tiruchirappalli	Kambarasampettai Dam	10°50'59.3"N	78°41'02.6"E	0.028 – 0.245
14	Tiruchirappalli	Melur Bridge	10°52'01.6"N	78°39'38.6"E	0.011 – 0.231

15	Tiruchirappalli	Putru Kovil	10°52'34.4"N	78°37'44.0"E	0.014 – 0.213
16	Tiruchirappalli	Panayakuruchi Palam	10°50'19.1"N	78°43'27.6"E	0.017 – 0.226
17	Tiruchirappalli	Uthamarseeli	10°49'48.0"N	78°46'52.3"E	0.019 – 0.168
18	Tiruchirappalli	Kowtharasanllur Palam	10°50'06.8"N	78°48'08.4"E	0.026 – 0.174
19	Tiruchirappalli	Thogur Canal	10°50'02.6"N	78°48'46.1"E	0.022 – 0.156
20	Tiruchirappalli	Mykulumi Outlet	10°49'42.6"N	78°48'57.4"E	0.023 – 0.144
21	Karur	Thirukaduthurai Palam	11°05'31.1"N	77°59'14.7"E	0.028 – 0.152
22	Karur	Kutchipalayam Canal	11°05'19.4"N	77°58'41.6"E	0.024 – 0.195
23	Karur	S Kondalam	11°04'07.6"N	77°55'59.7"E	0.016 – 0.221
24	Kumbakonam	Kabisthalam	10°56'39.9"N	79°16'12.4"E	0.022 – 0.185
25	Kumbakonam	Kudamuriti Melauthamannallur	10°56'12.6"N	79°16'53.1"E	0.031 – 0.221

This approach ensures a diverse representation of environmental conditions, enhancing the robustness and applicability of the developed ANN models for predicting the performance of the water turbine harvester across different settings. The details regarding the water bodies and rivers flowing in different parts of Tamil Nadu, has been taken from the open source data available in Tamil Nadu State Council for Science, Technology & Environment, hosted by Ministry of environment, Forests & Climate Change, Government of India. The temperature, rainfall, minimum flow of water and maximum flow rate of water in these small rivers and canals were obtained from the open source data available in the repository of Meteorological Department, Government of India [11]. From the meteorological data available for the past five consecutive years, the average of them has been taken for consideration.

Figure 1. Photograph of Oar-Based Water Turbine Harvester (OBWTH)

Conducting experiments under carefully specified flow conditions, the blades were immersed at 25% for an extensive two-month duration spanning from September to October. This time frame was intentionally chosen to coincide with a period of increased rainfall compared to other months. The resulting dataset is robust, consisting of 60 readings meticulously recorded during the experiments.

Various parameters were diligently observed and documented at the experimental location, each offering key insights into the turbine's performance:

1. Daily Maximum Temperature (T_{max}): This metric represents the greatest temperature that was measured each day, giving insight into the thermal conditions that prevailed during the experiment.

2. Daily Minimum Temperature (T_{min}): Tmin, the lowest temperature ever measured in a given day, adds to our knowledge of the range of temperature variations and how they could affect a turbine.

3. Average Rainfall (AR): This parameter provides important information on the availability of water resources and how they affect turbine operations by quantifying the mean amount of rainfall that was received throughout the experimental period.

4. Wind Speed (WS): WS, which indicates wind speed, is essential for assessing the possible effects of wind forces on turbine blades and overall system performance.

5. Maximum Fluctuation in Height (MFH): This parameter provides information about the dynamic changes in the hydraulic circumstances that the turbine is subjected to by measuring the highest variation in water level.

6. Gross Height Drop (GHD): GHD provides information about the gross height energy available for the turbine by reflecting the difference in water level between the accumulation point (point 0) and the turbine exit (H).

7. Net Height Drop (H_n): Hn, which measures the turbine's available potential energy while taking the net height drop into account, is essential to comprehending the turbine's overall effectiveness in turning water potential into electricity.

To replicate practical operating conditions, the studies were carried out with the blades half submerged. The selected period, which is marked by increased precipitation, provides important information about how well the turbine functions in these kinds of conditions. The resulting dataset, which includes variations in these indicated characteristics, makes a substantial contribution to a thorough comprehension of the behaviour of the system during the given time frame. The information gathered for the location-based parameters over a 60-day period (October to December) is displayed in Figure 2.

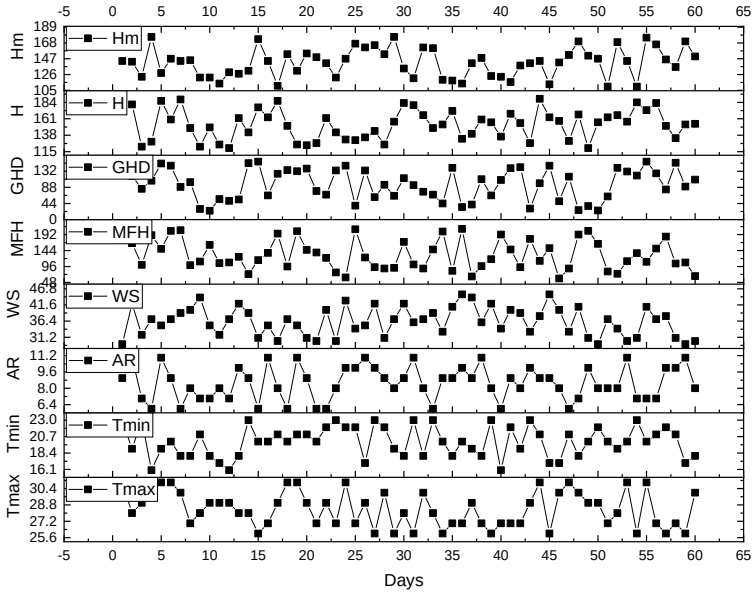

Figure 2. Data collected during the 60 day span (October to December) for the location based parameters

The information about the Oar-Based Water Turbine Harvester is as follows. Hydraulic energy loss equation H_{loss} is displayed beneath.

$$H_{loss} = A \times Q^2 \text{ ------- Eq. 1}$$

The turbine's coefficient, A, and the volume of water passing through it, Q (m^3/s), are given in the equation above.

The following formula is used to determine the turbine's hydraulic power [12].

$$HP = \rho \times Q \times g \times H_n \text{ ------Eq. 2}$$

The density in the equation above is denoted by ρ.

G stands for gravitational acceleration.

The net height drop, or Hn, is a measure of the turbine's potential energy that is available (m).

On the turbine shaft, mechanical power (MP) [13] is displayed as

$$MP = T \times \omega = T \times \frac{\pi n}{30} = T \times \frac{n}{9.549}$$ ---------Eq. 3

The turbine's efficiency can be calculated by dividing its mechanical power by its hydraulic power.

$$\eta_T = \frac{MP}{HP}$$ --------- Eq. 4

Specific speed based on the power element - SSP

The turbine SSP's rotating speed in revolutions per minute (min^{-1}) is defined as a height drop of one metre and a power level of one kilowatt.

$$SSP = n \times \frac{\sqrt{P}}{H_n^{5/4}}$$ ----- Eq. 4

The benefit of this word is that the power P includes the utilization rate.

The specific speed as seen by the flow aspect - SSQ

It is defined as the turbine SSQ's rotations per minute (min^{-1}) at a one-meter height drop and a one-cubic-meter-per-second flow rate.

$$SSQ = n \times \frac{\sqrt{Q}}{H_n^{5/4}}$$ --------- Eq. 5

The following illustrates the link between SSP and SSQ.

$$SSP = 3.65 \times \sqrt{\eta} \times SSQ$$ ---- Eq. 6

P stands for turbine power and ω for angular velocity (rad/s) in the equations above.

Thus, when developing the ANN model, the following parameters were included.

Loss of hydraulic energy - H_{loss}

HP is hydraulic power.

Equipment mechanical power - MP

Efficiency of turbines - ηT

Specific speed while considering power SSP

The specific velocity as seen by the flow aspect – SSQ

There are a number of important factors that affect the Oar-Based Water Turbine Harvester's performance. These parameters are described below with thorough explanations [14].

1. Blade Length to Blade Width Ratio (RBLBW): This ratio, which is expressed as bl/bw, is essential for figuring out how the turbine blades are arranged geometrically. It represents the proportionate relationship between the breadth and length of the blades, which affects how well the turbine captures water flow to produce electricity.

2. Axle Length (AL): The Axle Length (AL) of a turbine is the distance between its pivotal points, expressed in millimetres (mm). This parameter affects the turbine's stability and overall structural integrity by influencing its mechanical features.

3. Blade Inclination Angle (BIA): The Blade Inclination Angle (BIA), which is measured in degrees (°), indicates the tilt or angle at which the turbine blades are positioned with respect to the water flow. This angle is essential for maximising the interaction between the water flowing through the blades, which affects how well the turbine converts water energy into power..

An extensive evaluation that mirrored the geographic environment and lasted 60 days, from October to November, was conducted to determine how well the Oar-Based Water Turbine Harvester performed in relation to these factors. The resulting data, as shown in Figure 3, sheds important light on how changes in RBLBW, AL, and BIA affect the turbine's performance during the given time frame. This investigation advances our knowledge of how design factors interact with the efficiency of the turbine in capturing water energy.

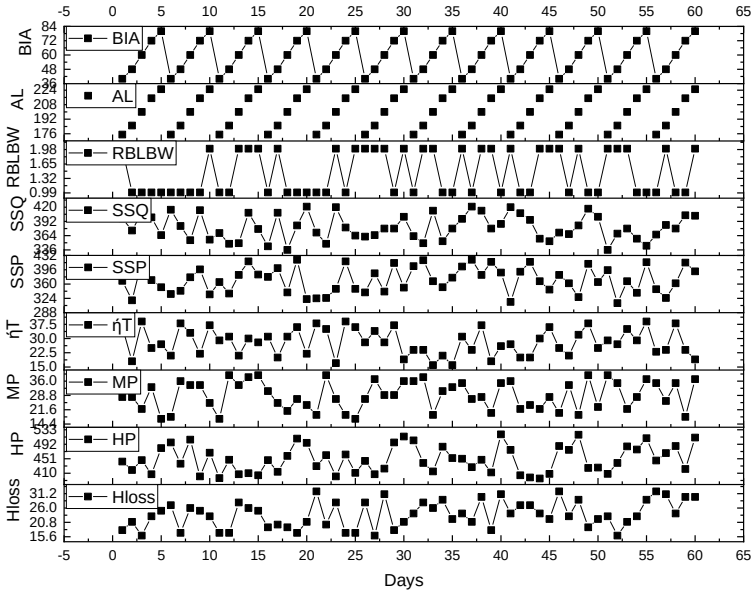

Figure 3. Data related to the parameters of Oar-Based Water Turbine Harvester evaluated for 60 days

2.2. Input variables selection using WEKA

Picking the right input variables is a crucial first step in building an Artificial Neural Network (ANN) model. In particular, seventeen input process variables were selected as the training set for the Oar-Based Water Turbine Harvester (OBWTH) efficiency prediction models. A critical evaluation is conducted as part of the variable selection procedure to determine which input factors are most important in determining the OBWTH's efficiency estimate.

The study makes use of WEKA [15], a software suite that the government of New Zealand developed in 1993, for this purpose. WEKA is a flexible tool used in many fields, such as machine learning, business analytics, and data mining. The J48 algorithm, which is an implementation of the c4.5 algorithm, is commonly used in the WEKA framework to build Decision Trees. In this case, decision trees aid in identifying and comprehending the critical factors impacting the effectiveness of the OBWTH. Decision trees are useful in identifying links within intricate datasets. The creation of precise and useful efficiency prediction models for the Oar-Based Water Turbine Harvester is dependent on this methodical approach to input variable selection.

A decision tree, which represents knowledge in a structure like to a tree, functions as a categorization rule. Using the Decision Tree approach, relevant variables are chosen in order to estimate the Oar-Based Water

Turbine Harvester's (OBWTH) efficiency. A standard Decision Tree produced by the c4.5 algorithm (or possibly ID3 or c5.0) consists of several branches, one root, numerous nodes, and numerous leaves [16].

In this structure, a branch is represented by a sequence of nodes that go from the root to the leaf, with each node representing a distinct variable. When a variable is added to the tree, it offers important information about how important it is in influencing the efficiency predictions for the Oar-Based Water Turbine Harvester (OBWTH). As a result, the design and organisation of the Decision Tree provide an information-rich and visually intuitive representation that highlights the relationships and significance of the many variables in the context of predicting the effectiveness of the OBWTH.

WEKA explorer was used to demonstrate how to use WEKA to choose pertinent variables from an input vector X that is generated from input variables. Using twenty-five sets of data samples from various geographic locations—small water canals, in particular—this video attempts to illustrate the process of choosing relevant input variables.

Choosing a search strategy and attribute evaluator in WEKA is part of the process. Next, all of the variable ranks that result are noted. Table 2 gives a detailed summary of the ranks that WEKA allocated to each input variable in relation to solar radiation prediction. The variables that have the lowest rankings are highlighted and labelled as secondary. When Artificial Neural Network (ANN) models are developed later, process parameters that show less impact on the output prediction accuracy are replaced with parameters that have a greater positive effect on the hydrokinetic efficiency of the Oar-Based Water Turbine Harvester (OBWTH).

The ranks that the WEKA algorithm allocated to the input variables in order to forecast the hydrokinetic efficiency of OBWTH are shown in Tables 2 and 3. With the use of this methodical technique, it is ensured that the variables chosen will best contribute to the precision and efficacy of the hydrokinetic efficiency estimates for the Oar-Based Water Turbine Harvester.

Table 2 - Rank of Input Variables (Tmax, Tmin, AR, WS, MFH, GHD, H, Hm, & Hloss)

Location	Tmax	Tmin	AR	WS	MFH	GHD	H	Hm	Hloss
1	0.11036	0.2915	0.09829	0.11	0.07529	0.03973	0.09677	0.15189	0.12376
2	0.18876	0.4664	0.07045	0.14021	0.05996	0.12138	0.02776	0.15298	0.12564
3	0.12821	0.0858	0.13361	0.02725	0.12701	0.06026	0.07248	0.09364	0.0618
4	0.1014	0.5058	0.07428	0.10778	0.03676	0.0252	0.0787	0.0248	0.12568
5	0.14538	0.445	0.066	0.14028	0.15605	0.05822	0.03468	0.06219	0.01174
6	0.17855	0.0848	0.17556	0.06761	0.0549	0.08747	0.09884	0.11794	0.01599
7	0.0584	0.3189	0.12768	0.08818	0.1032	0.14946	0.08065	0.15189	0.05956
8	0.18296	0.0618	0.12815	0.05526	0.03379	0.14683	0.05842	0.0286	0.01422
9	0.04619	0.4698	0.05825	0.04491	0.09224	0.09765	0.00856	0.07462	0.09462
10	0.18141	0.4682	0.19675	0.09307	0.00883	0.09469	0.11153	0.08805	0.02348
11	0.19312	0.0746	0.21203	0.05019	0.02089	0.12191	0.0272	0.06615	0.09209
12	0.04948	0.1505	0.20327	0.07492	0.12088	0.10005	0.08417	0.01891	0.06504
13	0.05618	0.4117	0.07433	0.12686	0.11468	0.08106	0.00734	0.00656	0.10512

14	0.09936	0.284	0.17542	0.12819	0.08747	0.06314	0.1532	0.11399	0.0249
15	0.14213	0.5348	0.11371	0.14819	0.08774	0.12347	0.08727	0.15258	0.15827
16	0.12894	0.3477	0.17826	0.04983	0.05443	0.15507	0.15263	0.01845	0.04456
17	0.20971	0.3328	0.23283	0.10396	0.0717	0.02572	0.01159	0.022	0.07774
18	0.17207	0.2603	0.15837	0.00795	0.11751	0.15748	0.09246	0.09024	0.04103
19	0.16913	0.1577	0.19325	0.13751	0.07305	0.11023	0.07649	0.03458	0.03073
20	0.0885	0.3821	0.06028	0.1197	0.02773	0.02726	0.06485	0.06436	0.14993
21	0.04056	0.1612	0.1469	0.14243	0.13436	0.15322	0.12549	0.13361	0.09206
22	0.09832	0.4878	0.07841	0.14049	0.14264	0.06489	0.07669	0.13532	0.1096
23	0.13158	0.0458	0.10393	0.14018	0.00926	0.05116	0.05705	0.10284	0.0467
24	0.03568	0.2348	0.1843	0.08427	0.00513	0.09617	0.15468	0.07399	0.057
25	0.13511	0.116	0.16533	0.12029	0.01955	0.03302	0.01394	0.10046	0.11111

Table 3 - Rank of Input Variables (HP, MP, ήT, SSP, SSQ, RBLBW, AL, BIA)

Location	HP	MP	ήT	SSP	SSQ	RBLBW	AL	BIA
1	0.11778	0.13616	0.04656	0.13434	0.04744	0.0842	0.0644	0.04248
2	0.02598	0.12491	0.141728	0.12757	0.02997	0.06024	0.11247	0.01742
3	0.01738	0.05231	0.167648	0.09885	0.08872	0.0904	0.07718	0.07575
4	0.0205	0.01471	0.278272	0.10674	0.03568	0.03265	0.01324	0.10175
5	0.14986	0.07955	0.247392	0.07521	0.08726	0.0067	0.08517	0.10983
6	0.1471	0.00697	0.080736	0.14385	0.06597	0.0659	0.01878	0.12813
7	0.07349	0.14301	0.053376	0.05202	0.12134	0.01236	0.15395	0.11376
8	0.0087	0.05026	0.040352	0.10973	0.03863	0.13507	0.04507	0.07923
9	0.10044	0.05702	0.067328	0.04307	0.14853	0.01972	0.1146	0.06744
10	0.00916	0.07603	0.073376	0.01061	0.01215	0.14001	0.08394	0.00499
11	0.10924	0.05649	0.344416	0.0603	0.03056	0.14399	0.08948	0.02087
12	0.08675	0.08938	0.22592	0.09605	0.11596	0.15125	0.05028	0.05729
13	0.02363	0.02775	0.45984	0.14	0.10995	0.14334	0.01639	0.14621
14	0.09119	0.05721	0.4848	0.1525	0.15764	0.05659	0.06483	0.04617
15	0.13508	0.131	0.441088	0.15153	0.1397	0.04153	0.08206	0.0515

16	0.12575	0.05139	0.063392	0.15659	0.13613	0.13967	0.00644	0.06435
17	0.0168	0.05865	0.3544	0.07506	0.05358	0.13749	0.09331	0.08117
18	0.0377	0.15432	0.049792	0.07945	0.05631	0.07677	0.04662	0.11266
19	0.09519	0.04796	0.119168	0.15151	0.0259	0.08282	0.04492	0.03391
20	0.12642	0.09462	0.234496	0.09849	0.10331	0.0909	0.02183	0.05245
21	0.03386	0.10653	0.074464	0.14605	0.09916	0.13387	0.13769	0.11869
22	0.06213	0.02372	0.194624	0.13043	0.15414	0.10604	0.08036	0.10154
23	0.09708	0.03881	0.202624	0.02554	0.108	0.01926	0.10432	0.15463
24	0.1244	0.06733	0.29136	0.14374	0.15486	0.03268	0.12648	0.12657
25	0.15035	0.13232	0.435456	0.09458	0.10164	0.10415	0.02258	0.03814

3 Results & Discussions

3.1. OBWTH hydrokinetic efficiency prediction models prediction models with selected inputs

Artificial neural network fitting tool (nftool) is utilised in the development of the ANN models (ANN-1, ANN-2, and ANN-3), which are used for prediction. The Levenberg-Marquardt (LM) algorithm-trained standard two-layer feed forward neural network that powers the nftool is appropriate for static fitting challenges. Scaled conjugate gradient [18] is automatically used for training. Sixty percent, twenty percent, and twenty percent of the randomly divided data are utilised for training, testing, and validation, respectively. The input and target values are automatically mapped in the range of -1 to 1. The ANN models are trained using the LM method on the training data. The testing data offer an impartial gauge of network performance both during and after training, and they have no bearing on training.

The validation data are used to gauge the network's capacity for generalisation and are stopped during training when generalisation reaches its limit. Plotting the mean square error (MSE) against the number of epochs represents the performance. The epochs consist of a single training, testing, and validation sweep. The MSE in training, testing, and validation data is displayed in the performance plot. The MSE plot displays a higher curve in the validation data set and a lower curve in the training data. The trained ANN model is the network with the lowest validation mean square error (MSE) [17]. When an increase in the mean square error (MSE) of validation data samples indicates that validation error is no longer improving, the training automatically ends. Several training sessions will yield distinct outcomes because connection weights are randomly initialised and because initial conditions vary. Figure 4 shows the step-by-step instructions for implementing nftool, while Figure 5 shows the suggested methodology and Appendix A contains the Matlab code.

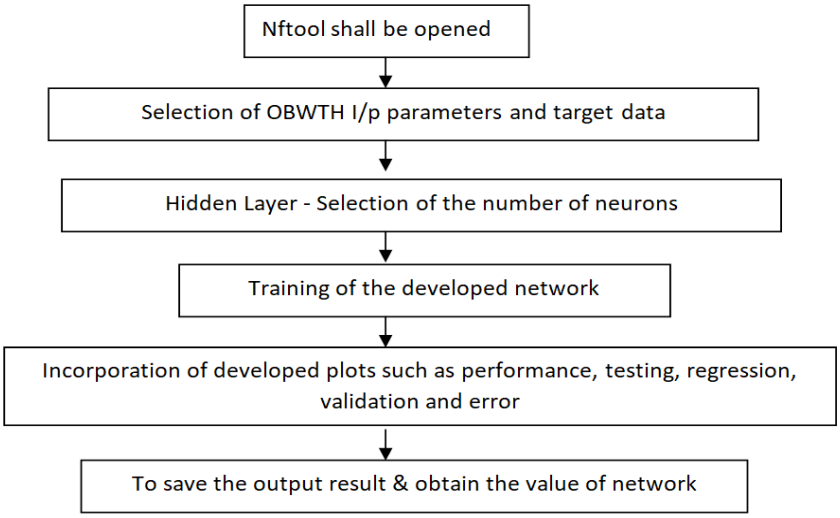

Figure 4. Stepwise sequence for implementation of nftool in WEKA

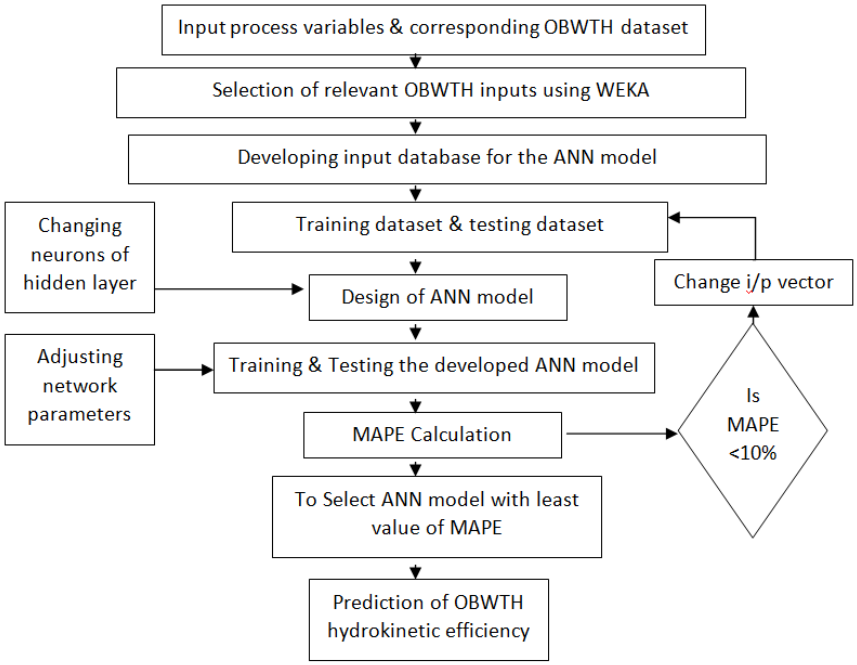

Figure 5. Algorithm for developing WEKA based ANN model

Equation below calculates the number of neurons in the hidden layer; HLn and DSn are the number of neurons in the hidden layer and the number of data samples used in the artificial neural network (ANN) model, respectively; Ni and No stand for the number of input and output parameters.

$$HL_n = \frac{Ni + No}{2} + \sqrt{DS}_n \quad \text{---------- Eq. 7}$$

By measuring the change in prediction error (MAPE) when the number of hidden layer neurons is altered + 5 from the hidden layer neurons estimated by Eq. (1) (above), the sensitivity test is carried out to validate the number of hidden layer neurons. Tables 4–6 display the results of the sensitivity analysis of hidden layer neurons for ANN models. Table 4 displays the statistical error evaluation for the ANN-1 model. Table 5

displays the statistical error evaluation for the ANN-2 model. Table 6 displays the statistical error evaluation for the ANN-3 model.

The ANN architecture with the least MAPE is utilised to forecast the OBWTH hydrokinetic efficiency. The MAPE is provided by Eq. (2).

$$MAPE = \left(\frac{1}{n} \sum_{i=1}^{n} \left| \frac{SR_{iANN} - SR_{Iactual}}{SR_{Iactual}} \right| \right) \times 100 \quad \text{..... Eq. 8}$$

Table 4 - Statistical Error Evaluation in ANN-1 model

Hidden layer neurons sensitivity	MLP structure	Training R value	Max. MAPE during testing	ANN selection
For 8 inputs, Avg. η for 2 months, 25 samples, HLn varied from 9 to 19	8 – 9 – 1	90.81	21.22	ANN architecture with 8 inputs, 10 hidden layers and 1 neuron in o/p layer was selected
	8 – 10 – 1	**92.22**	**19.32**	
	8 – 11 – 1	89.32	19.98	
	8 – 12 – 1	87.41	21.25	
	8 – 13 – 1	96.52	19.98	
	8 – 14 – 1	88.69	24.21	
	8 – 15 – 1	87.41	23.41	
	8 – 16 – 1	85.26	21.69	
	8 – 17 – 1	89.63	20.89	
	8 – 18 - 1	90.02	21.22	

Table 4 - Statistical Error Evaluation in ANN-2 model

Hidden layer neurons sensitivity	MLP structure	Training R value	Max. MAPE during testing	ANN selection
For 7 inputs, Avg. η for 2 months, 25 samples, HLn varied from 9 to 18	7 – 9 – 1	89.68	19.89	ANN architecture with 7 inputs, 11 hidden layers and 1 neuron in o/p layer was selected
	7 – 10 – 1	90.89	21.23	
	7 – 11 – 1	**91.86**	**18.69**	
	7 – 12 – 1	87.41	22.41	
	7 – 13 – 1	88.32	21.52	
	7 – 14 – 1	89.54	20.87	
	7 – 15 – 1	90.41	22.41	
	7 – 16 – 1	91.22	23.54	
	7 – 17 – 1	87.63	21.74	
	7 – 18 - 1	88.41	20.89	

Table 4 - Statistical Error Evaluation in ANN-3 model

Hidden layer neurons sensitivity	MLP structure	Training R value	Max. MAPE during testing	ANN selection
For 4 inputs, Avg. η for 2 months, 25 samples, HLn varied from 9 to 18	4 – 9 – 1	90.23	21.32	ANN architecture with 4 inputs, 11 hidden layers and 1 neuron in o/p layer was selected
	4 – 10 – 1	91.82	20.21	
	4 – 11 – 1	**92.18**	**18.83**	
	4 – 12 – 1	89.51	22.32	
	4 – 13 – 1	90.21	24.56	
	4 – 14 – 1	88.69	25.12	
	4 – 15 – 1	87.21	21.41	
	4 – 16 – 1	89.52	22.23	
	4 – 17 – 1	88.63	23.21	
	4 – 18 - 1	89.91	22.12	

Lewis's MAPE is used to assess prediction accuracy [49]. A prediction accuracy of 10% < MAPE < 20% suggests a decent prediction, a fair prediction of 20% < MAPE < 50% indicates a reasonable prediction, and a forecasting accuracy of 10% < MAPE < 50% indicates erroneous forecasting. The ANN-1, ANN-2, and ANN-3 models' maximum MAPE of testing sites are determined to be 19.36%, 11.29%, and 8.31%, respectively. This indicates that the ANN-3 model's prediction accuracy increases by 12.67% upon deleting less influential characteristics. Consequently, WEKA can be utilised to determine pertinent input parameters for the prediction of solar radiation. ANN-3's MAPE is lower than ANN-2's, indicating that the parameters were more crucial.

We therefore lower the training data when the variables selection challenge is solved. Three neural network (ANN) models (ANN-1, ANN-2, and ANN-3) were created to determine prediction accuracy in order to validate the authenticity of WEKA. Every process variable is incorporated into the ANN model. More significant variables from the first analysis, provided by WEKA, were used in the ANN-2 model. Just five important input variables were used by the ANN-3 model: MP, ΔT, RBLBW, AL, and BIA.

Figure 6 displays the ANN models' performance graphs. The training, test, and validation Mean Square Error were plotted over 11 epochs.

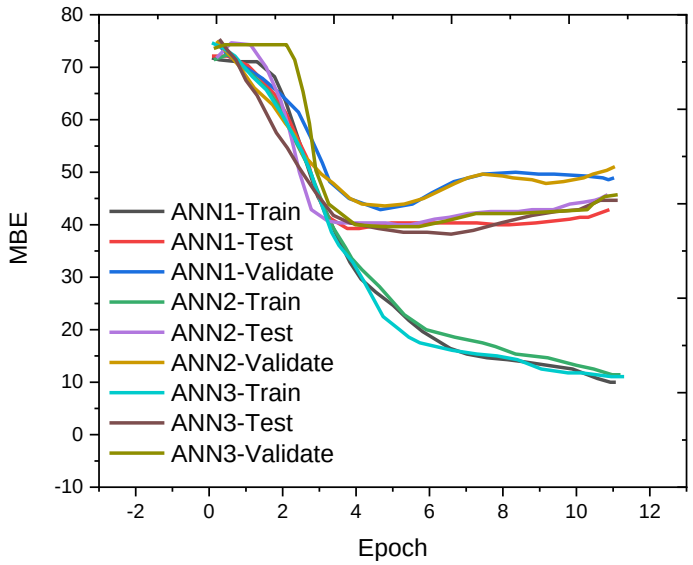

Figure 6. MBE variations for training, testing and validation of ANN1, ANN2 and ANN3.

In the context of the ANN-3 model, the correlation coefficient (R-value) is a critical metric for evaluating the relationship between the goal values and projected outputs. Three Artificial Neural Network (ANN) models—ANN-1, ANN-2, and ANN-3—were created and put through a comparison study in order to assess the predicting performance. ANN-3 significantly outperformed ANN-1 and ANN-2 in terms of prediction accuracy, showing an impressive 12.67% improvement. Blade width, blade axle length, blade inclination angle, and turbine efficiency were found to be the most significant input factors in the feature selection procedure carried out using WEKA.

These factors were found to be important contributors to the prediction of hydrokinetic efficiency overall. An overall hydrokinetic efficiency prediction of 29.6% was obtained by combining the integration of the blade width, blade axle length, blade inclination angle, and turbine efficiency to create a strong prediction model. This shows that, given the chosen input variables, the model can predict hydrokinetic turbine efficiency with reasonable accuracy.

Conclusions

- The study highlights the importance of specific input parameters—hydraulic power, mechanical power, turbine efficiency, blade dimensions, and inclination angle— in predicting hydrokinetic efficiency.
- The comparison of ANN models reveals that ANN-3, incorporating the ratio of blade length to blade width, blade axle length, blade inclination angle, and turbine efficiency, outperforms ANN-1 and ANN-2, showing a significant improvement in prediction accuracy (18.83% MAPE).
- WEKA identifies blade width, blade axle length, blade inclination angle, and turbine efficiency as the most relevant variables, contributing significantly to accurate hydrokinetic efficiency predictions.
- The developed ANN models, particularly ANN-3, demonstrate a good level of prediction accuracy with a 29.6% success rate in forecasting hydrokinetic efficiency.
- The study underscores the significance of specific input parameters and the use of ANN models, particularly ANN-3, in accurately predicting hydrokinetic efficiency for small water turbine harvesters in renewable energy applications.

References

[1] Rares-Andrei, C., Florentina, B., Gabriela, O., & Lucia-Andreea, E. L. (2017, October). Power prediction method applicable to horizontal axis hydrokinetic turbines. In *2017 International Conference on Energy and Environment (CIEM)* (pp. 221-225). IEEE.

[2] Javaherchi, T., Stelzenmuller, N., & Aliseda, A. (2017). Experimental and numerical analysis of the performance and wake of a scale–model horizontal axis marine hydrokinetic turbine. *Journal of Renewable and Sustainable Energy*, *9*(4).

[3] Musa, M., Heisel, M., & Guala, M. (2018). Predictive model for local scour downstream of hydrokinetic turbines in erodible channels. *Physical Review Fluids*, *3*(2), 024606.

[4] Wang, W. Q., Yin, R., & Yan, Y. (2019). Design and prediction hydrodynamic performance of horizontal axis micro-hydrokinetic river turbine. *Renewable energy*, *133*, 91-102.

[5] Riglin, J., Carter III, F., Oblas, N., Schleicher, W. C., Daskiran, C., & Oztekin, A. (2016). Experimental and numerical characterization of a full-scale portable hydrokinetic turbine prototype for river applications. *Renewable Energy*, *99*, 772-783.

[6] Rengma, T. S., & Subbarao, P. M. V. (2022). Optimization of semicircular blade profile of Savonius hydrokinetic turbine using artificial neural network. *Renewable Energy*, *200*, 658-673.

[7] Khani, M. S., Shahsavani, Y., Mehraein, M., & Kisi, O. (2023). Performance evaluation of the savonius hydrokinetic turbine using soft computing techniques. *Renewable Energy*, 118906.

[8] Paturi, U. M. R., Cheruku, S., & Reddy, N. S. (2022). Artificial neural networks modelling for power coefficient of Archimedes screw turbine for hydropower applications. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, *44*(10), 447.

[9] Abutunis, A., Hussein, R., & Chandrashekhara, K. (2019). A neural network approach to enhance blade element momentum theory performance for horizontal axis hydrokinetic turbine application. *Renewable Energy*, *136*, 1281-1293.

[10] Siavash, N. K., Ghobadian, B., Najafi, G., Rohani, A., Tavakoli, T., Mahmoodi, E., & Mamat, R. (2021). Prediction of power generation and rotor angular speed of a small wind turbine equipped to a controllable duct using artificial neural network and multiple linear regression. *Environmental research*, *196*, 110434.

[11] Chuphal, D. S., & Mishra, V. (2023). Hydrological model-based streamflow reconstruction for Indian sub-continental river basins, 1951–2021. Scientific Data, 10(1), 717.

[12] Arndt, R. E., & Chamorro, L. P. (2017). Hydraulic turbines. In Energy Conversion (pp. 257-279). CRC Press.

[13] Fathabadi, H. (2016). Maximum mechanical power extraction from wind turbines using novel proposed high accuracy single-sensor-based maximum power point tracking technique. Energy, 113, 1219-1230.

[14] Ferro, L. M. C., Gato, L. M. C., & Falcão, A. F. O. (2011). Design of the rotor blades of a mini hydraulic bulb-turbine. Renewable Energy, 36(9), 2395-2403.

[15] Aarthi, E., Daniel, J. D., Suba, G. M., Dharani, N. P., & Devi, C. P. (2024). A Naive Bayes Approach for Improving Heart Disease Detection on Healthcare Monitoring Through IoT and WSN. International Journal of Intelligent Systems and Applications in Engineering, 12(2s), 553-570.

[16] Owie-Obazee, E. (2020). Development of a Predictive Model for the Classification of the Survival of Hepatitis patients using Decision Trees Algorithm.

[17] bin Othman, M. F., & Yau, T. M. S. (2007). Comparison of different classification techniques using WEKA for breast cancer. In 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006: Biomed 2006, 11–14 December 2006 Kuala Lumpur, Malaysia (pp. 520-523). Springer Berlin Heidelberg.

[18] Babani, L., Jadhav, S., & Chaudhari, B. (2016). Scaled conjugate gradient based adaptive ANN control for SVM-DTC induction motor drive. In Artificial Intelligence Applications and Innovations: 12th IFIP WG 12.5 International Conference and Workshops, AIAI 2016, Thessaloniki, Greece, September 16-18, 2016, Proceedings 12 (pp. 384-395). Springer International Publishing.

.