ADOPTION OF INTELLIGENT COMPUTATIONAL TECHNIQUES FOR STEAM BOILERS TUBE LEAK TRIP

Firas Basim Ismail[1*], Deshvin Singh[2], Mohammad Shakir Nasif[3]

[1,2]Power Generation Unit, Institute of Power Engineering (IPE), Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

[3] Department of Mechanical Engineering, Universiti Teknologi Petronas, 32610, Seri Iskandar, Perak, Malaysia

Email: Firas@uniten.edu.my[1*](corresponding author), deshvin1992@hotmail.com[2]
mohammad.nasif@utp.edu.my[3]

DOI: https://doi.org/10.22452/mjcs.vol33no2.4

ABSTRACT

Frequent boiler tube trips in coal fired power plants can increase operating cost significantly. An early detection and diagnosis of boiler trips is essential for continuous safe operations in the plant. Several methodologies for the fault diagnosis in a plant have been developed. However these methodologies are difficult to be implemented. In this study, two artificial intelligent monitoring systems specialized in boiler trips have been proposed. The first intelligent monitoring system represents the use of pure artificial neural network system whereas the second intelligent monitoring system represents merging of genetic algorithms and artificial neural networks as a hybrid intelligent system. In the first system using pure artificial neural network, the trip was predicted 5 minutes before the actual trip occurrence. The hybrid intelligent system was able to optimize the selection of the most influencing variables successfully and predict the trip 2 minutes before the actual trip. The first intelligent system performed better than the second one based on the prediction time. The proposed artificial intelligent system could be adopted on-line as a reliable controller of the thermal power plant boiler.

Keywords: Coal-fired power plant, Boiler, Artificial neural networks (ANNs), Genetic algorithms (GAs), Tube leak

1.0 INTRODUCTION

A typical power plant consists of the boiler, condensers, heaters, turbines, and various components that work together continuously, producing energy. Steam boilers represent the main equipment in the power plant. Boiler tubes continuously pass fluids while withstanding extremely high temperatures and pressures of the steam generator [1]. This eventually leads to cracks in boiler tubes. A small leak from a particular tube causes erosion to other surrounding tubes as well causing serious secondary damages. A boiler tube leak can also be caused by one of the following reasons: caustic attack, oxygen pitting, hydrogen damage, acid attack, stress corrosion, waterside corrosion, super heater fireside ash corrosion, short-term overheat, long-term overheat, graphitization, and dissimilar metal weld [2]. Malfunction of boiler tube causes frequent boiler trips which eventually increase the operating costs of a plant [3]. A tube leakage causes abnormal changes in the values of the related boiler parameters. As the leakage continues certain related boiler parameters such as feed water flow rate, super heater steam temperature, economizer inlet pressure and drum leveldeviates from its normal range. Once the sensors detect that a number of influential boiler parameters has exceeded or has reduced from the acceptable range, a trip occurs where the particular system isolates itself from others. The time taken for repair can be significantly long as it involves shutting down the whole operating system, allowing the boiler to cool down, identifying the cause of the tube leak and also the replacement and welding of the tubes. The cost due to damage caused by a serious tube leak could go up to an average of $6 million per leak [4-5]. Eventually, by the time the replacements are carried out, the plant would have undergone a big financial loss.

Conventional detection methods of detecting tube leaks by measuring the make-up water or inspecting is ineffective as it gives information on the existing tube leak only when serious damage has already occurs. Lang et al. [5] developed an on-line method which detects steam generator and heat exchanger tube leaks occurs using the Tube Failure Model. It was concluded that tube leakage represents a major source of unrecognized heat rate degradation. The results show that the impact of tube leakages approximately decreases the boiler efficiency by 0.7% for every 40,000 lbm/hr leakage flow. Therefore, preventing tube leakages through early detection could eventually improve the overall plant efficiency. Consequently, prediction of boiler tube leak trips is crucial to maintain normal and safe operational conditions of the plant [6], so that the operator can be warned earlier before the trip actually occurs and preventive measures can be taken to avoid further damage and an eventual shutdown. Benefits of an early detection of boiler tube leak trip are: to increase operating profit by reducing repair costs and secondary damage, increase safety of the plant, increase availability and tube life, and avoid unplanned outages [7].

Prediction can be achieved by developing intelligent monitoring system which monitors the operating conditions. Using information provided by measurements from several sensors and actuators that are abundantly stored throughout operations, the intelligent monitoring system is able to predict a failure [8]. The system stores the real data of previous faults and will notify the operator when it detects operating conditions which indicate that a trip will occur. The resources for these real data of any operating plant are readily available. From over thousands of available measurements, the scope can be reduced to working fluid components in the boiler in which only variables that have high influence on boiler operations are taken into account. They include values of the temperatures, pressures, flow and levels of boiler components including the economizers, drums, headers, circulating water pumps, risers and super heaters.

One of the most commonly used intelligent systems for early fault detection is the Artificial Neural Network. Messai et al. [9] have applied the ANN for on-line fault prediction of fuel rod temperature measurement sensor in a nuclear core. This study highlighted that ANN is preferable over other intelligent systems because it does not require derivation of the mathematical model of the process.

Muhammad and Halim [10] proposed a predictive ANN model for visualization techniques' classification. The performance of the proposed ANN model been compared with five other classifiers such as k-nearest neighbor, naïve Bayes, decision tree, random forest, and support vector machine. The ANN-based prediction model of automated visualization selection outperformed other classifiers in terms of accuracy and execution time.

Halim et al. [11] developed a predictive ANN-SVM model for driver's profile classification. The trained classifier driving features were proposed as model inputs. The training process was based on clustered data. Comparison between these clusters were performed which showed that the various data sets with different combinations of attributes provided approximately the same number of clusters.

ANNs are designed in layers that include input, output, and hidden layers. Based on a review study on on-line condition monitoring system for high level trip water in steam boilers drum, the researcher has used only one-hidden layer (1HL) ANN structure for fault prediction and diagnosis [12]. However, in another study on pure artificial intelligence monitoring system for coal-fired power plant boiler drum high water level trip, both one hidden layer and two hidden layers cases were tested and compared in which the optimum NN structure with two hidden layers outperformed the one hidden layer structure [13]. Kumar et.al [14] have applied ANN for the prediction of refuse plastic boiler performance. The ANN topologies taken into account to achieve an optimal architecture include the training algorithms, activation functions, number of inputs nodes, hidden layers, hidden nodes, and output nodes. The best network architecture achieved for optimization of the plant operation consisted of 4-12-3 (input-hidden-output) neurons, training count of 50,000, learning rate of 0.1, and momentum coefficient of 0.1. Another study on ANN had adopted the methodology of feed forward back propagation to diagnose a steam boiler based on super heater monitoring, in which the minimisation algorithms were applied [15]. Feed forward networks are faster compared to feedback networks due to the inexistence of repeated loops. Rostek et al. [16] presented an early detection of leaks in fluidized-bed boilers using ANN. It was verified in the study that the performance of ANN using multi-layer perceptron (MLP) was better than radial basis function (RBF) neural network as a detection and prediction tool. In an application of ANN for the prediction of high efficiency boiler steam generation and distribution, the root mean square error was used as a performance indicator for the intelligent system where the mean square error identified the weights that minimises the error [17].

ANN as a stand-alone system is effective. However, there are several other intelligent systems that have been applied together with ANN for early fault detection. Genetic algorithm is an optimization tool which reduces the possibilities of getting into local minimum in the process of finding an optimum solution [18]. A patented invention by Lang and Rafael [19] involved rapid detection of tube failures and their locations without using direct instrumentation in order to prevent serious damage and minimize the repair time of the affected heat exchanger. This method is applicable to Input/Loss method of monitoring fossil-fired thermal plants.

Momeni et al. [20] applied hybrid genetic algorithm-based ANN for the prediction of pile bearing capacity. From this study, it was concluded that using the optimum Genetic Algorithm (GA) parameters, the hybrid model outperformed the conventional ANN model in the prediction process. Genetic algorithm was applied successfully together with ANN as a hybrid system for prediction of flow characteristic in serpentine micro channels. The best network configuration was evaluated using trial and error of different combinations of ANN topologies [21]. Researchers in the past have shown that the GA is an efficient tool in optimization of the selection of variables for various applications [22-24]. A study on the

ANN-GA smart appliance scheduling for optimised energy management in the domestic sector highlighted that implementation of GA significantly reduced the computational time of the whole process [25]. Majid et al. proposed a hybrid system for a real wastewater treatment plant which optimised weights and thresholds of the multi-layer perceptron neural network and radial basis function neural network [26]. Hajnayeb et al. [27] compared the ANN- based feature selection with the genetic algorithm for gearbox fault diagnosis. In this application, new generation of chromosomes is created using previous generation by implementing the selection, crossover and mutation methods.

Halim et al. [28] used evolutionary algorithm (EA) to cluster large probabilistic graphs in order to achieve better results. The proposed approach was compared with two baseline clustering algorithms and two state-of-the-art approaches for uncertain /probabilistic data. The EA based approach showed better performance. Optimum clustering was achieved by having local and global search by enabling the EA. The EA was able to have local and global search capability due to its multiple populations.

Halim and Uzma [29] state that the majority of the research performed on Minimum Spanning Tree-based clustering (MST) focused on forming better cluster formations by partitioning the MST. Therefore, they optimized the MST-based clusters using Evolution Strategy (ES). The proposed work differs from previous approaches as it is not based on any assumption about the underlying data. It was concluded that the proposed solution performed better compared to other clustering approaches.

Halim and Muhammad [30] proposed a set of visualization metrics in order to evaluate visualization techniques. The EA framework which uses tree-maps as a case study was applied. The optimum visualization layout was achieved using the genetic operators and the proposed visualization metrics as an objective function. These evolved visualizations were compared with a randomly created visualization and visualization created using a state-of-the-art treemap visualization tool. The comparison was made using internal and external evaluation metrics. The visualization evolved using the combined fitness function was more effective than the visualizations optimized for effectiveness, expressiveness, readability, and interactivity in isolation.

From all of the existing solutions, there has not been any intelligent monitoring system established particularly for boiler tube leak trip in thermal power plants. Therefore, it is important to develop a reliable intelligent monitoring system for prediction of boiler tube leak trip. Several researches used mathematically simulated data instead of real plant data for training and validation, which is inaccurate for decision making. It can be noted from previous applications that the raw data are used directly for the training process. However, the data are not always complete and the quality is not acceptable. Therefore, a data preparation framework is important for the complex data analysis. Previously, trial and error approach has been applied for the selection of best NN topology combination and the optimization of the boiler operation parameters. The selection of suitable boiler operation variables and the design of the optimal neural network topology can be achieved using additional artificial intelligent system technique such as GA. Based on previous researchers, no direct comparison of performance between a pure intelligent monitoring system (ANN) and a hybrid intelligent monitoring system (GA+ANN) for boiler tube leak trip was made. This is important to understand the prediction capability of both systems in predicting a boiler tube leak trip.

Therefore, the objective of this research is to develop two IMSs for boiler trip prediction by adopting pure artificial neural network system and hybrid intelligent system. The first intelligent monitoring system represent the use of the pure artificial neural network system for boiler trip prediction. The final architecture for this system was explored after investigation of various neural network topology combinations which include two hidden layers, one to ten neurons for each hidden layer, three types of activation function, and four types of multidimensional minimization training algorithms. The second intelligent condition monitoring system is a hybrid intelligent system using the application of artificial neural networks with genetic algorithms. To ease the selection of suitable boiler operation variables and to achieve the optimal neural network topology, encoding and optimization process is performed using genetic algorithms. The main objective of developing the two IMSs, is to determine if the pure system is able to successfully predict a trip and if a hybrid system is able to enhance the capability of the pure system in predicting a trip. Each design process of the intelligent monitoring system goes through preliminary and basic training and validation processes, in which the data is segmented to 60% for training and 40% for validation. The performances of these systems are measured using Root Mean Square Error (RMSE), where output error between the actual output and the expected output is evaluated. It is noted that lower error value indicates better performance. The comparison of the IMSs from this study allows us to understand the capability and performance of each system particularly in predicting a boiler tube leak trip. The feed-forward neural network methodology has been adopted as a major computational intelligent tool in both the systems. Real operational data of a particular boiler tube leak trip obtained from a coal-fired power plant was used in this study.

2.0 DESCRIPTION OF POWER PLANT

In this research, the power plant being considered consists of three sub-critical pressure boilers. Each boiler is a single reheat and controlled circulation type. The boiler is fired with pulverized coal to produce steam for the continuous generation of 700 MW(e). The combustion system consists of a single furnace with balanced draught. Using direct tangential firing located in the furnace corners, the maximum heat input that can be attained is 40% of the Boiler Maximum Continuous Rate (BMCR). Complying with the Malaysian environmental requirements, the plant consists of low nitrogen oxide combustion burner system such as the Over Fire Air (OFA) ports. The sulphur dioxide emission is controlled through the Flue Gas Desulphurisation (FGD) plant and the dust in the flue gas at the boiler outlet is removed by an Electro-Static Precipitator (ESP). Remote controlled fuel oil burners with high energy ignitions are used to start-up the boiler. Pulverized coal is able to undergo combustion at low firing rates using these burners. Apart from three boiler circulating pumps, the main auxiliary equipment includes two forced draft fans, two induced draft fans and two primary air fans which are centrifugal fans with control vanes at the inlet. Other components of the auxiliary equipment include two steam air pre heaters, one piece of soot-blowing equipment, two electrostatic precipitators. The boiler has 7 vertical bowl mills) in one coal milling plant. Fig. 1 shows the schematic diagram of the boiler system.

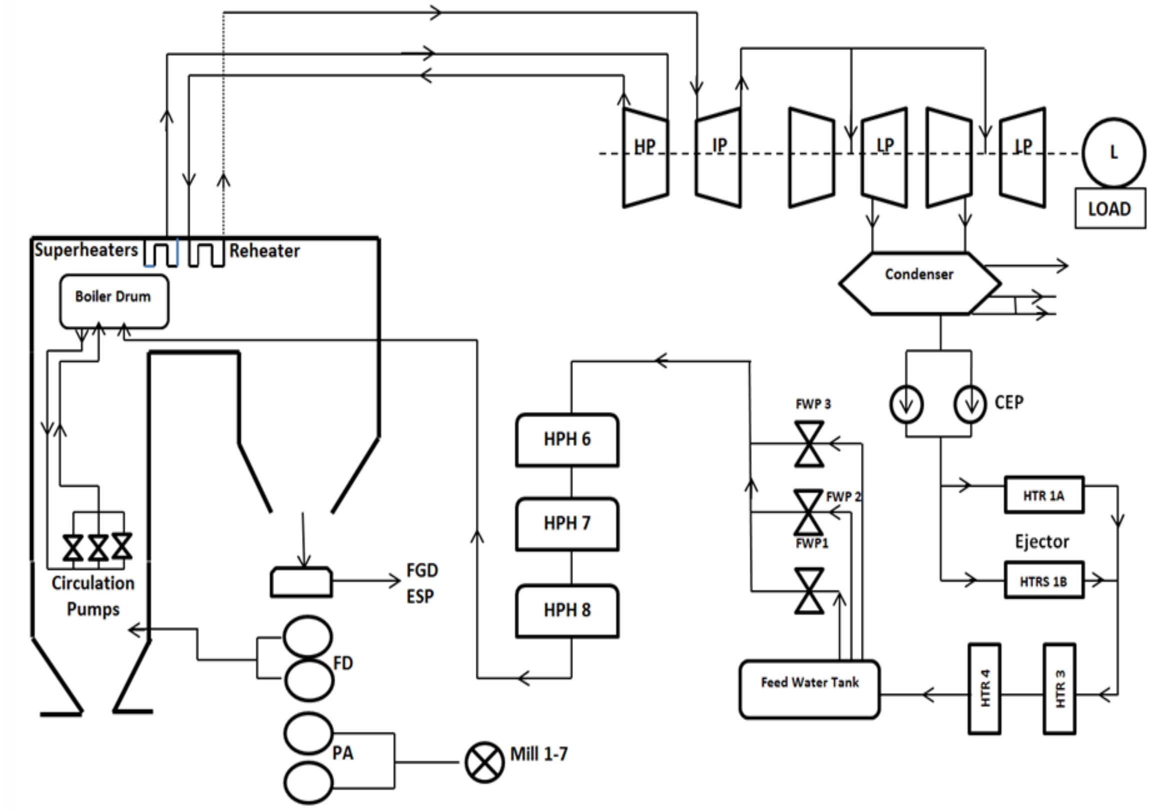

Fig. 1: Schematic diagram of MNJ TPP Boiler.

3.0 DATA PREPARATION FRAMEWORK

The current work includes four related execution phases which are represented in Fig.2. These phases are sequential steps that are identified to execute the entire study.

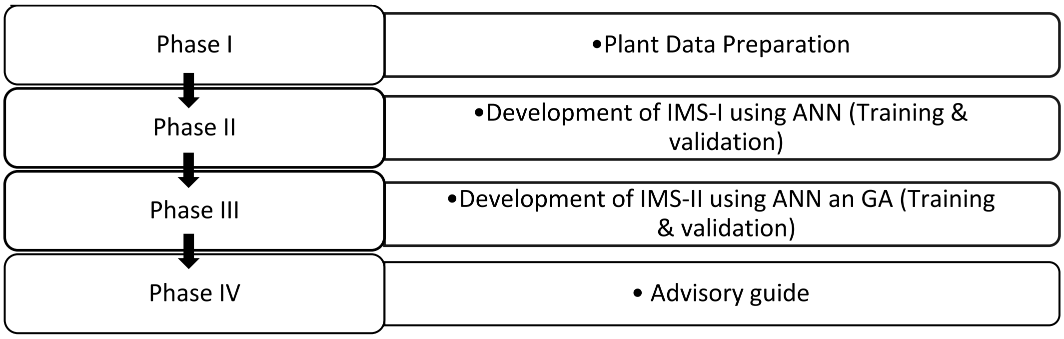

Fig. 2: Execution phases

The first phase was the integrated data preparation which was further divided into three stages as shown in Fig.3. The data preparation stage involved steps which include mathematical and statistical processes.

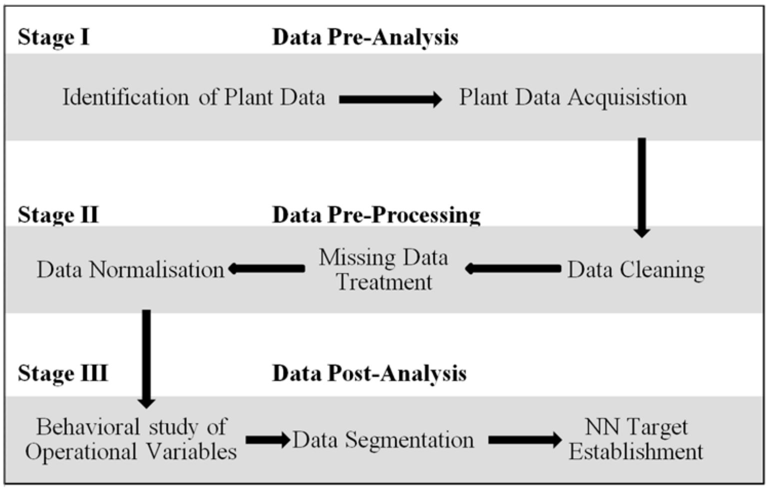

Fig. 3: Plant Data Preparation Stage

3.1 Data Pre-Analysis Stage

This stage of the execution phases focussed on the identification of boiler operational variables and the acquisition of plant data for a specific boiler trip.

3.1.1 Plant Data Identification Step

The on-line plant control system consists of over 1,800 observations of actuators and sensor signals which were identified as the boiler process observations. This large number of observation was reduced to 177 based on the work scope, which focuses on the diagnosis of working fluid components only in the boiler. This includes relevant measurement values of operational variables related to headers, economisers, circulating water pumps, risers, drums and super heaters. The work scope excluded observations of furnace items, induce fans and mills. Plant operator experience was taken into account to identify observations that do not affect boiler trips. These non-effective factors were neglected and the observations were reduced to 93. Many observations were measured by multi-sensors thus, eventually thirty-two influential operation boiler variables were identified and tabulated in Table 1. The average number of signals retrieved for each variable was represented by the 'criteria' column in Table 1. Once the operation boiler variables are identified, the data are acquired as described in the following section.

Table 1: Influential boiler operation variables

Code	Variables	Unit	Criteria	High Alarm (minute)
V1	Total combined Steam flow	t/h	$= 1/1 \sum_{1}^{1} V1_1$	705
V2	Feed water flow	t/h	$= 1/1 \sum_{1}^{1} V2_1$	704
V3	Boiler drum pressure	Barg	$= 1/1 \sum_{1}^{4} V3_4$	704
V4	Superheater steam pressure	Barg	$= 1/1 \sum_{1}^{1} V4_1$	704
V5	Superheater steam temperature	°C	$= 1/1 \sum_{1}^{1} V5_1$	2471
V6	High temperature Re-heater outlet temperature	°C	$= 1/4 \sum_{1}^{4} V4_4$	963
V7	High temperature superheater exchange metal temperature	°C	$= 1/4 \sum_{1}^{4} V7_4$	-
V8	Intermediate temperature (A) superheater exchange metal temperature	°C	$= 1/4 \sum_{1}^{4} V8_4$	2472
V9	High temperature superheater inlet header metal temperature	°C	$= 1/4 \sum_{1}^{4} V9_4$	2471
V10	Final superheater outlet temperature	°C	$= 1/6 \sum_{1}^{6} V10_6$	-
V11	Superheater steam pressure transmitter (control)	bar	$= 1/7 \sum_{1}^{7} V11_7$	2471
V12	Feed water valve station	t/h	$= 1/8 \sum_{1}^{8} V12_8$	704
V13	Feed water control valve position	%	$= 1/4 \sum_{1}^{4} V13_4$	704
V14	Drum level corrected (control)	mm	$= 1/1 \sum_{1}^{1} V14_1$	2214
V15	Drum level compensated (from protection)	mm	$= 1/1 \sum_{1}^{1} V15_1$	704
V16	Feed water flow transmitter	%	$= 1/1 \sum_{1}^{1} V16_1$	-
V17	Boiler circulation pump1 pressure	bar	$= 1/1 \sum_{1}^{1} V17_1$	2031
V18	Boiler circulation pump 2 pressure	bar	$= 1/2 \sum_{1}^{2} V18_2$	1959
V19	Low temperature super heater left wall outlet before super heater dryer	°C	$= 1/4 \sum_{1}^{4} V19_4$	704
V20	Low temperature superheater right wall outlet before super heater dryer	°C	$= 1/2 \sum_{1}^{2} V20_2$	2612
V21	Low temperature super heater left wall after super heater dryer	°C	$= 1/2 \sum_{1}^{2} V21_2$	958
V22	Low temperature super heater right wall exchange metal temperature	°C	$= 1/1 \sum_{1}^{1} V22_1$	2474
V23	Intermediate temperature (B) super heater exchange metal temperature	°C	$= 1/1 \sum_{1}^{1} V23_1$	1948
V24	Intermediate temperature super heater outlet before super heater dryer	°C	$= 1/1 \sum_{1}^{1} V24_1$	1944
V25	Intermediate temperature super heater outlet header metal temperature	°C	$= 1/2 \sum_{1}^{2} V25_2$	2007
V26	High temperature super heater outlet header metal temperature	°C	$= 1/6 \sum_{1}^{6} V26_6$	2480
V27	High temperature Re-heater outlet steam pressure	bar	$= 1/2 \sum_{1}^{2} V27_2$	2477
V28	Superheated steam from intermediate temperatures outlet pressure	bar	$= 1/11 \sum_{1}^{11} V28_{11}$	-

V29	Super heater water injection compensated flow	ton/hr	$= 1/10\sum_{1}^{10} V29_{10}$	2479
V30	Economizer inlet Pressure	bar	$= 1/6\sum_{1}^{6} V30_{6}$	961
V31	Economizer inlet temperature	°C	$= 1/1\sum_{1}^{1} V31_{1}$	-
V32	Economizer outlet temperature	°C	$= 1/1\sum_{1}^{1} V32_{1}$	-

3.1.2 Plant data acquisition phase

The data of the boiler operation variables, which were identified in the previous phase were captured based on groups. There are two groups of data captured from the boiler, in which group A is based on the outage intervals for a particular trip, and group B represents normal boiler operation. These groups were compared to determine the difference in parameters of the variables before and during a trip. The sampling time of one minute was used because short sampling time allows for precise training and validation results. The shutdown period determines the intervals in between the trip data. Trip data duration was 24 hours before and after the trip if the shutdown lasts more than 24 hours. In this case the boiler water wall tube leak trip lasted for 5 days. The acquired data then goes through the data pre-processing stage.

3.2 Data Pre-Processing Stage

After identifying and acquiring the data in the pre-analysis stage, the data goes through data pre-processing stage. The large data collected from the power plant containing thousands of values was filtered and passed through three important steps. They are visual cleaning of data, missing data treatment, and data normalisation.

A large spreadsheet of data has several errors in the form of noise values, which is identified visually through the process of visual cleaning. Missing data treatment involves an automated graphical tool which is used to remove duplicate records caused by variations of the same erroneous listing. In a large set of data, default values are usually assigned to missing data which forms non-numeric errors. A MATLAB code is used as a data analysis tool to input the valid range of values. The number of values for a numeric record was reduced using data smoothing which include the rounding technique and the mean values computing technique. Mean value smoothing was applied when a classifier does not support numerical data. A typical data point was located and removed from the set of data using outlier's removal smoothing. The removal tool used was MATLAB code with 95%confident level. Lost information in the form of missing data can be solved by discarding the missing values records. These missing data is also replaceable with real data using class mean. Missing data treatment also include the application of mathematical forecasting methods such as extrapolation and interpolation. Other similar condition values can also be used to replace the missing values.

A better IMS performance is observed using scaled data compared to non-scaled data. Therefore, data normalisation is used as the data transformation technique to scale the data between 0 and 1 (0 represents running mode and 1 represents faulty). Since the minimum and maximum values are known, the Min-Max data normalisation is applied using the following equation 1.

$$New\ value = \frac{\text{Original value - Min}_{\text{old}}}{\text{Max}_{\text{old}}\text{ - Min }_{\text{old}}} \tag{1}$$

3.3 Data Post Analysis

This stage involved the behavioural study of boiler operational variables, data segmentation into two sets: sub group A and sub group B, and establishment of NN targets. Using the data from group B, which represents the normal operational conditions without a trip, mean value for each variable was determined. To achieve the desired ANN input range of [0, 1], the Min-Max data normalisation transformation was used to transform the mean values for normal and faulty boiler operation to normalised mean values. Data from group A involving a trip was used to investigate behaviour of each variable. Understanding the behaviour of the variables is essential in the analysis of the IMS.

For each variable, the time interval to reach high alarm indicator "1" was illustrated in the "High Alarm" column of Table 1. From the 32 variables, the low temperature super heater left wall outlet variable (V20) was selected as the most effective variable. This is because V20 reached the high alarm indicator at the 2612th minute, which is closest to the shutdown time of 2614th minute. The rest of the effective variables (V5, V8, V9, V11, V14, V17, V18, V22, V23, V24, V25, V26, V27, and V29) were also chosen based on the time interval to reach high alarm indicator.In the training and validation process, only data before the unit shutdown was used. The ANN training was done in two phases: the preliminary training phase and the basic training phase. Data sub group A was divided into 70% for preliminary training and 30%for preliminary validation, whereas data sub group B was divided into 70% for basic training and 30% for basic validation.

The target matrix can be established using the results from the high alarm occurrence table. The fault introduced matrix indicates the time at which each influencing variable reached its maximum value i.e. reaching “1”. The method of the target matrix establishment was repeated by assuming the faulty data with ±5, ±10, ±15, ±20 and ±25 minutes. The analysis showed that the ±20 minutes provided optimum training performance for the ANN system, unlike the ±25 minutes which is negligible because the steady state convergence was achieved. The fault target interval was decided to be within 20 minutes before reaching 1 and 20 minutes after reaching “1”. Therefore, all other values are “0” when normalized and were considered non-faulty values. Thus targets of the trip are assigned in the newly established matrix which was later used for the ANN training process.

4.0 DESIGN AND IMPLEMENTATION OF IMSS

In this study, the feed-forward methodology was adopted using the back-propagation neural network. Activation functions are applied to the weighted sum of its input. The main NN topologies are training algorithms, learning rate, momentum coefficient, activation functions, number of hidden layers, and the number of neurons in each hidden layers. In this study, up to 2 hidden layers were used. An ANN architecture of one hidden layer (1HL) is able to approximate arbitrarily with any system that contains continuous mapping from one finite space to another. However, ANN architecture containing two hidden layers (2HL) can represent a decision boundary with arbitrary accuracy. RMSE indicator was used to investigate the structured IMS-I performance. Based on the best performance of the IMS-I, training and validation is done. The thirty two variables presented before were used as inputs in this IMS-I and the outputs are either “0” which represents normal boiler situation or “1” which represents an abnormal boiler situation.

4.1.1 Preliminary Training

The main objective of this preliminary training is to determine the optimal NN topology combination. The IMS-I was input with faulty data set which represents 70% of sub-group A data. The network input includes all thirty-two operation variables. Fig. 4 shows the data segmentation sub groups for training and validation.

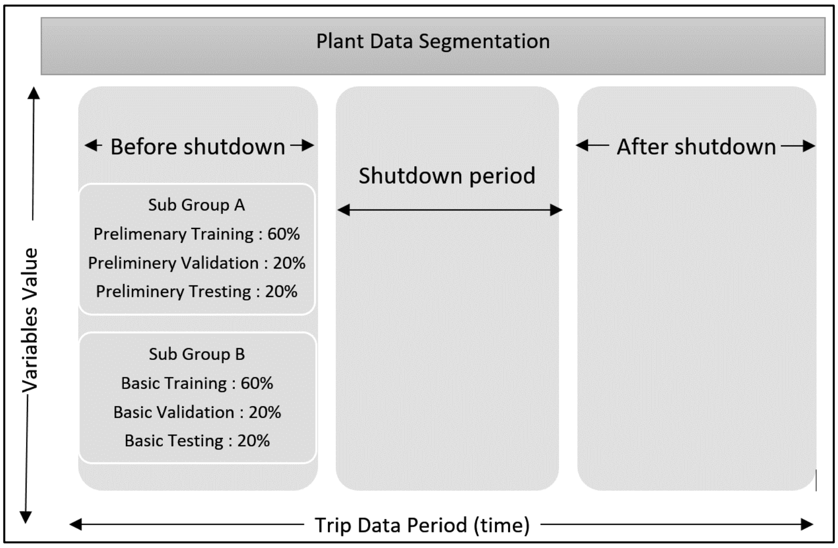

Fig. 4: Data segmentation sub groups for training and validation

4.1.2 Basic Training

The optimal topology combination obtained from the preliminary training results was used for the basic training. Real boiler data set which represents 70% of sub-group B data were presented as the basic training sets for the IMS-I. Based on the proposed system performance indicator, the basic training results were analysed. The calculated RMSE values of the optimal NN topology combination were compared with the RMSE value of the preliminary training process. The aim of this process is to get smaller or equal calculated RMSE values than the preliminary training process.

4.1.3 Preliminary Validation

In the preliminary validation process, the remaining 30% of sub-group A data set was used as a validation real data set. Applying the optimal NN topology combination obtained from the preliminary training process, validation of real data set was used to determine the rapidness of the proposed IMS-I in predicting the specific trip. The performance of the system during normal boiler operation was also validated

4.1.4 Basic Validation

In the basic validation process, the remaining 30% of sub group B data set was used. Another set representing normal boiler operational data was also used in this validation. The performance of the proposed system was examined using the obtained results in which the capability of the proposed IMS-I for the specific trip was explored. Fig. 5 shows the proposed IMS-I code execution flow which were used for the training and validation of the system.

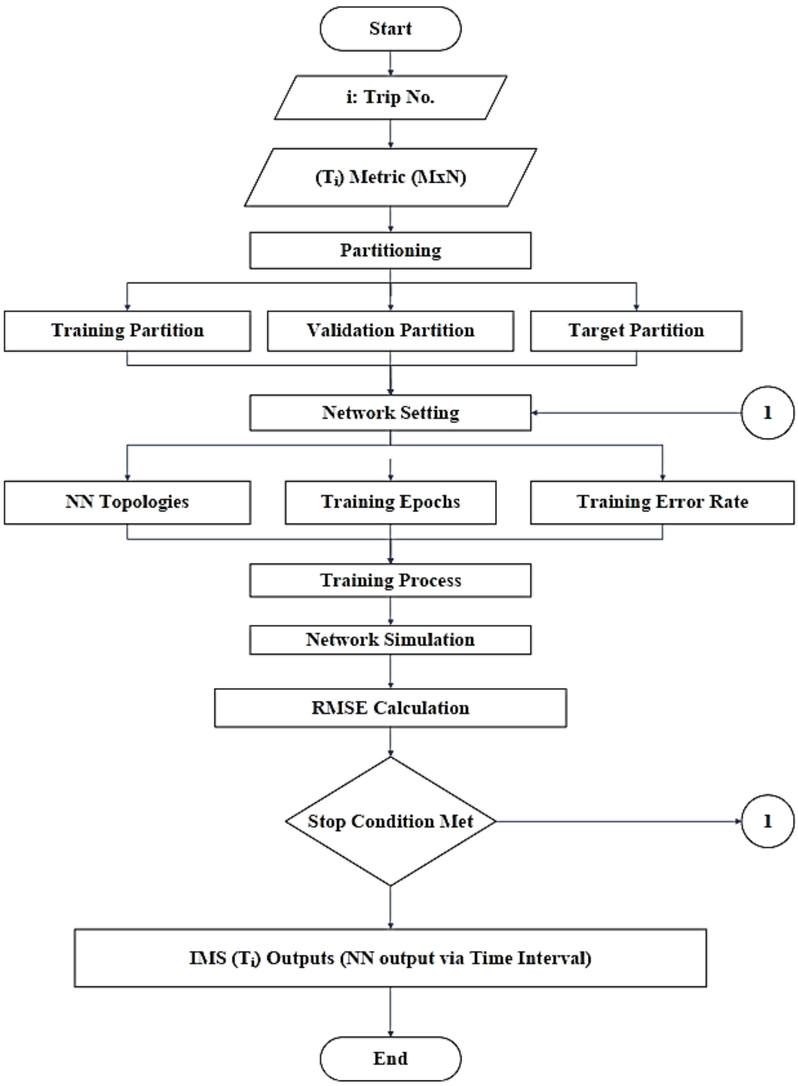

Fig. 5: Execution flow chart of the proposed IMS-I code

4.2 Design of IMS-II (ANN + Hybrid)

In the development of IMS-II, GA was used for automated design and NN parameter optimization. The IMS-II scheme was developed based on GA encoding.

4.2.1 GA Encoding

In this study, NN topologies were considered as the phenotype for the GA application. The NN topologies were encoded together with the NN training parameters into the genes of the GA using Weak Specification Representation (WSR). The proposed GA individual has genotypes which represent every phenotype of ANN topologies together with training parameters. The topologies proposed in this study are NN structure, activation function, back-propagation algorithms, and all the boiler operation variables.

The search space was reduced significantly using the WSR scheme, which is applied for multidimensional minimization algorithm selection, the ANN structure, the optimal ANN input parameters selection and the selection of activation types for the hidden nodes and the output nodes. GA considered all the four multidimensional minimization training algorithms and they are represented by two binary entries, thus forming the first two bits of binary string representation.

The WSR of ANN structure consists of seven binary entries of the string, which represents up to 91 viable ANN structures. WSR of the activation function take into account the three activation functions, which provides nine different combinations for one hidden layer and 27 combinations for two hidden layers. These combinations were included in the genetic representation and were encoded with 5 final binary entries in the bit string. All thirty two variables that were encoded with binary string formed the final thirty two bits of string. The equation below was used to calculate the combinations probability.

$$P_{vc} = 2^{n} \qquad (2)$$

4.2.2 IMS-II Scheme

The IMS-II scheme for optimal NN topology combinations and boiler operation variables selection can be divided into three major elements as shown in Fig. 6. They are the user element, the genetic algorithm optimization element and the ANN training element. I/O processes are handled by the user element. The optimization process includes numerous sub-sections interacting with each other and with the user part to complete the desired procedure.

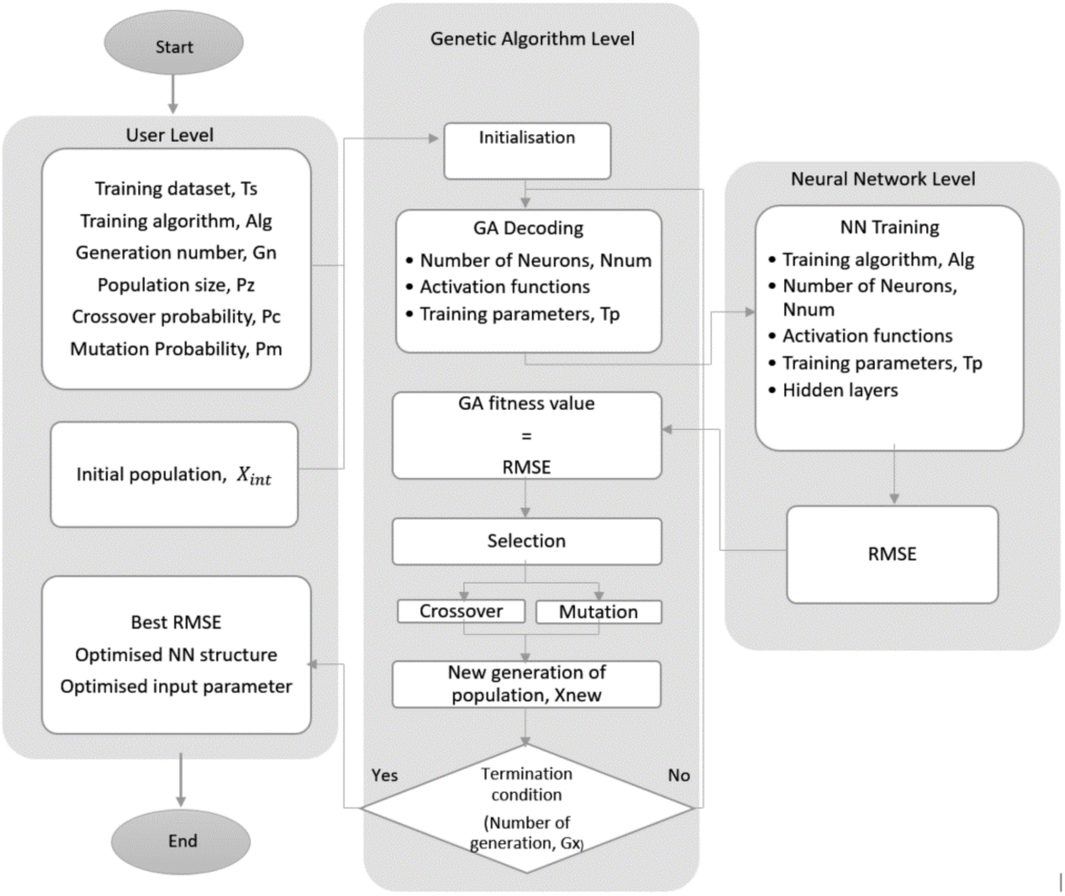

Fig. 6: Schematic representation of proposed IMS-II

At the first element, the user provides, the ANN training set (**Ts**), the parameters set for BBP training algorithms (BPs), the number of generations of the GA (**Gn**), the size of population of the GA (**Pz**), the probabilities of crossover of the GA (**Pc**), and the mutation probability of the GA (**Pm**). Every initial population of several binary strings represents a specific network topology and the NN training parameters set. This initial population is represented by the (**Xint**) in Fig.6. Every user inputs are then passed to the main optimization part.

"GA decoding" as the internal function accepts the population of binary strings. This is to decode each string of binary into explicit information regarding the four parts: Multidimensional Minimization Training algorithms (**Alg**), the network structure (**struct**), the Activation Function type of the network (**AF**) and the NN Training Parameters (**TP**). Training of the optimized parameters is carried out using specific topology. "GA fitness" which is the internal function, receives the calculated RMSE values for each individual after training. The value minimised by the GA is the fitness which is equivalent to the RMSE. A MATLAB code is formed for the implementation of the proposed scheme.

5.0 ASSESSMENT OF RESULTS

5.1 IMS-I Training

For the training using 1HL, the logistic activation function for both the input and output node performed better than the other two activation function. Every activation function probability combination performed proportionally to the NN input value history. The "Resilient Back-Propagation" and "Scaled Conjugate Gradient" algorithms had better performance than others in cases of different number of nodes in the hidden layer.

The performance of each training results was indicated by their RMSE values. The overall best performance (optimal RMSE) in the preliminary training exploration for both 1HL and 2HL architectures for the trip is tabulated in Table 2. The preliminary training results of the 1HL showed better performance for the input node using linear summation function. The best NN combination with minimum error value of 0.433 was achieved using the Scaled Conjugate Gradient (SCG) algorithm and 6 neurons in the hidden layer. For the hidden node and the output node, the logistic activation function performed better than the hyperbolic tangent activation function.

Table 2: The best ANN topology combination for both the 1HL and the 2HL

HL	RMSE	Architecture	Activation Function	Training Algorithm
1 HL	0.433	6HL1	L+L	Scaled Conjugate Gradient
2 HL	0.434	7HL1`-9HL2	P+T+L	Scaled Conjugate Gradient

5.2 IMS-I Validation

The trained output for the proposed IMS-I was “0” for normal boiler operation and “1” for faulty boiler operation. The prediction from the system produced an output range values of between “0” to “1”. Thus lower and upper thresholds were applied, in which the lower thresholds value indicate the point where boiler operation is assumed to be normal, and upper thresholds indicate the point where boiler operation is considered faulty. Based on the alarm boundaries of the selected boiler operation variables, the NN output upper threshold and the NN output lower threshold were set at 0.4 and 0.5 respectively. NN outputs ranging between these values indicates that the operation is neither normal nor faulty. Shift from normal to faulty boiler operation was based on the lower threshold and shift from faulty to normal boiler operation is based on the upper threshold.

The IMS-I output on the real data set is shown in Fig.7. With time step of one minute interval, the total data sampling interval was at the point of 392nd minute before the shutdown. The introduction of fault in the operation occurred at the 225th interval. The intelligent system was able to predict the fault within the 220th interval. The fault indication was considered strong (close to one) with system output of 0.95.

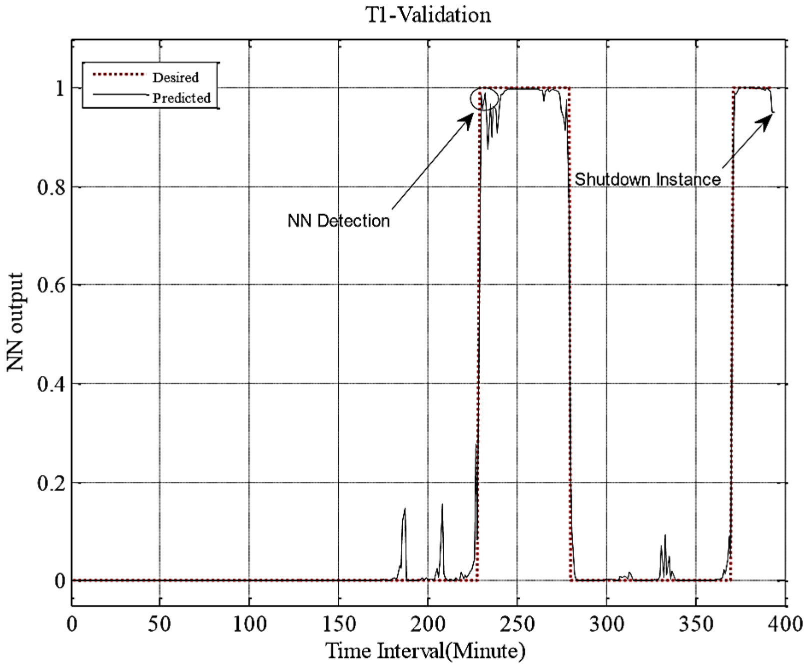

Fig. 7: IMS-I outputs for boiler trip

From the plot it is observed that despite being able to predict the fault, the system output returned to the normal boiler operation value (0.4) at multiple significant intervals. This occurred during the main boiler faulty operation and also happened towards the end of the specific fault (after the 267th minute) where the fault was about to disappear. Therefore the boiler operation is returned to normal again.

The ability of the proposed intelligent system to continuously recognize normal boiler operation was tested using the real data set which contains only normal boiler operation data. Fig.8 shows the IMS-I output which is close to “0” for almost the whole operating period.

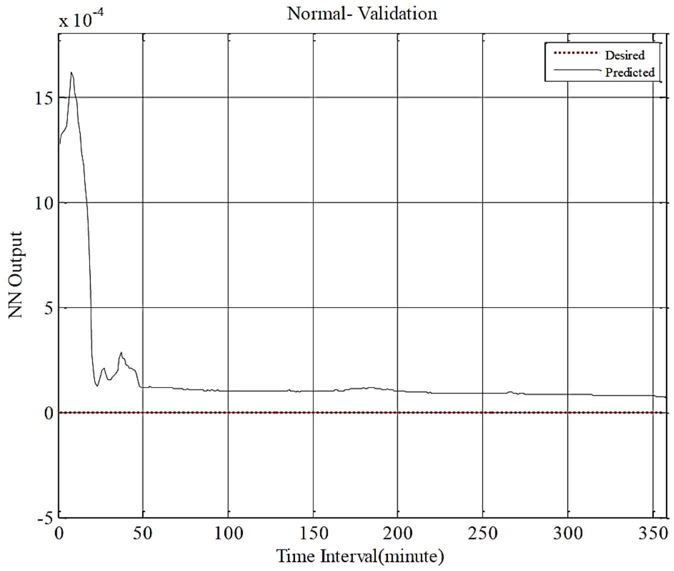

Fig. 8: IMS-I outputs for Normal operation

The specific thermal power plant trip was predicted by the proposed IMS-I approximately 5 minutes before the actual trip which is considered a satisfactory time period.

5.3 Hybrid Intelligent Monitoring System (IMS-II) Result Analysis

In the application of this IMS-II, it is crucial to determine the best values of probability of crossover (P_c) and probability of mutation (P_m). The GA system trials were carried out using population size (P_z) of 46 bit strings. The one point crossover was adopted since it is the most common crossover type. Results were obtained from the IMS-I as a function of P_c for three different values of P_m (0.01, 0.05 and 0.1). The P_c value of 0.6 gave the best IMS-II performance. In Fig. 9 the performances of the probability of mutation were compared. The comparison showed that P_m with the value of 0.05 performed better than the other probability of mutation.

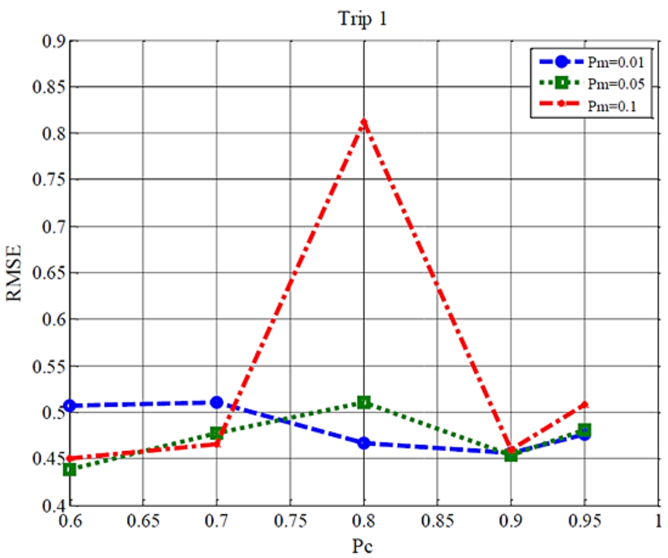

Fig. 9: Performance of the IMS-II using several probabilities of crossover and mutation

To determine the most appropriate population size (P_z), limited population size of L to 2L were investigated, where L represents the length of bit strings. Thus, the population sizes explored were between 46 and 92. The performance of the hybrid system using two different values of generation numbers (15 and 30) is compared in Fig. 10. The comparison showed that P_z value of 76 resulted in the best performance. Fig.11 shows the best RMSEs obtained from each generation during the best run of IMS-II.

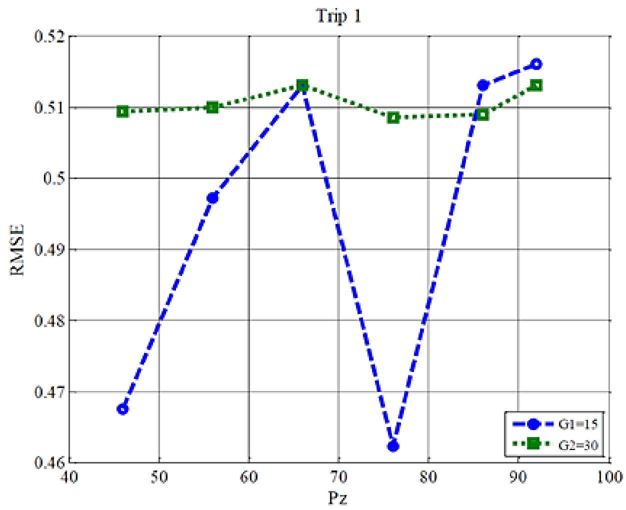

Fig. 10: Performance of the IMS-II using two population size candidates

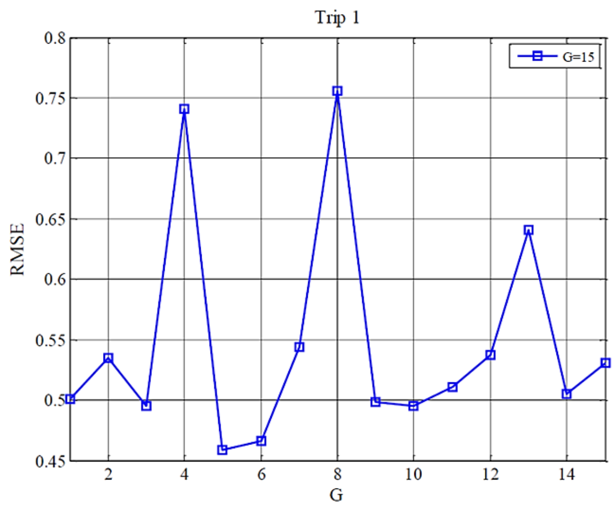

Fig. 11: Best RMSE obtained during the best IMS-II run

The main purpose of designing the GA fitness was to select fewer boiler operation variables. The selection which chooses more than necessary variables or the best NN topology was penalized while the training error for a fixed number of IMS-II iterations were reduced.

The best GA search selections for each generation for a boiler water wall tube leak trip was actually found at generation 10 with the following individual's bit string of [0010010110111011111100010000000111100010000000], which is interpreted as a 2HL with 9 and 3 neurons in the first and second hidden layers respectively, linear summation in the first hidden and output layer neurons, logistic function in the second hidden layer neurons trained with Resilient Back-Propagation training algorithm, 12 selected variables as ANN inputs. This best selection gave a value of RMSE = 0.456.

Table 3: Best GA selection interpretations

T1	RMSE	Selected Input Variables	No. of Inputs	ANN Topologies		
HSG.	Fitness			Training Algorithm	Architecture	Activation Function
G1	0.512541	V[1,4,6,7,9,10,12,13]	8	SCG	6HL1-2HL2	L+L+T
G2	0.510131	V[1,2,7,10,11,12]	6	BFGS	4HL1-4HL2	L+P+T
G3	0.534745	V[1,3,7,10]	4	Rprop	10HL1	L+P
G4	0.489741	V[1,2,3,4,5,6,10]	7	BFGS	3HL1-5HL2	L+T+T
G5	0.485571	V[1,2,3,4,7,10,11,12]	8	BFGS	3HL1-10HL2	L+P+P
G6	0.510384	V[2,7,10]	3	LM	7HL1-4HL2	L+P+L
G7	0.536561	V[2,3,4,5,9]	5	SCG	3HL1-2HL2	T+T+P
G8	0.564870	V[17,20,22,23,24,25,26,27,29,30]	10	BFGS	4HL1-10HL2	T+P+T
G9	0.474437	V[2,7,10,17,18,23,26,27,28]	9	LM	2HL1-7HL2	P+T+T
G10	**0.456288**	**V[1,2,3,4,5,6,10,18,19,20,21,25]**	**12**	**Rprop**	**9HL1-3HL2**	**P+L+P**
G11	0.530030	V[16,17,18,21,24,25,26]	7	LM	6HL1-8HL2	P+P+P
G12	0.486163	V[1,2,23,25,26,27,28,30,31]	9	Rprop	3HL1-8HL2	P+T+P
G13	0.523695	V[1,4,5,6,7,9,10,12,29]	9	SCG	2HL1-9HL2	T+T+T
G14	0.523899	V[11,12,13,14,15,16,19,20,23,31, 32]	11	SCG	2HL1-4HL2	T+P+L
G15	0.471608	V[2,4,7,10]	4	LM	4HL1-5HL2	L+T+L

The best GA selection interpretation is summarised in Table 3, in which the hybrid system generation is represented as "HSG". Based on the table, generation 15 and 9 gave the 2nd and 3rd best GA optimization selection. This shows that the two hidden layer architecture performed better than the one hidden layer architecture. During the GA search for these generations, operation variables 2 and 10 were the global optimal selections.

Using the optimal results obtained from the IMS-II, the validation process was repeated on the IMS-I. The same real boiler validation data sets and the same decision support approach were used to determine the capability of IMS-I to predict the fault. The same threshold range of 0.4-0.5 were used which indicates the previously known condition of the system output. The IMS-II output on the validation data set are shown in Fig.12.

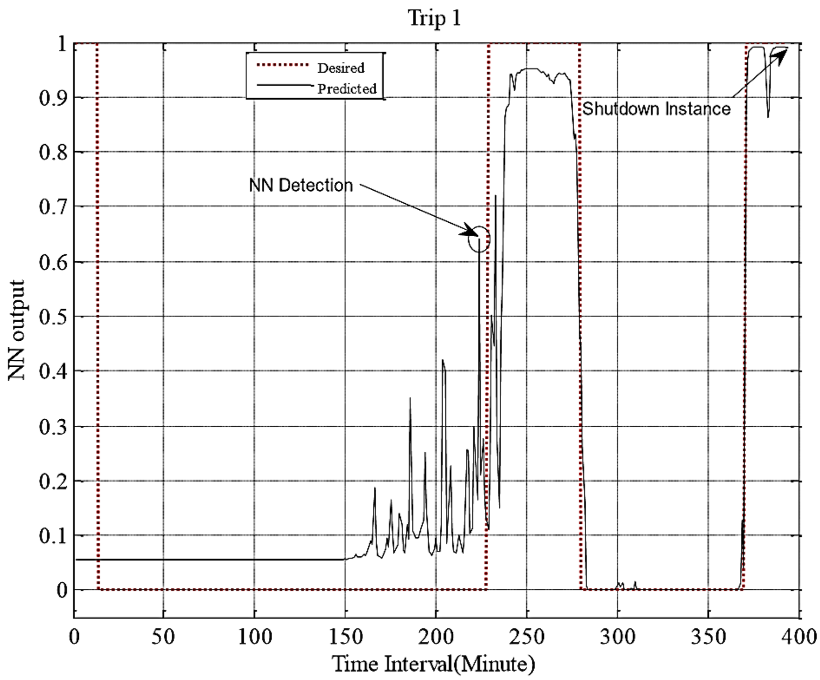

Fig. 12: IMS-II output during the validation

Using a one minute time interval, the boiler tube leak trip was introduced at the 225th minute However the intelligent system was able to predict the fault 2 minutes before the actual trip at the 223th minute. The system output with of 0.65 was considered a strong fault indication (close to one). The results of validation before and after using the IMS-II indicate that the first validation outperformed the second one with time interval difference of 3 minutes.

6.0 COMPARISON BETWEEN IMS-I AND IMS-II PERFORMANCES

The specific boiler operation trip was detected by the proposed IMS-I before the fault occurrence, with slight differences from the second proposed system. The IMS-II was proposed for optimization and automation of the selection of optimal combination of NN topologies and boiler operation variables. The infinite problem space under investigation was narrowed down after some limitations on the selection of the available NN topologies and the number of the most influential boiler operation variables as NN inputs. Using pure NN technique, only a limited portion of this type of problem was covered.

In this study, the problem space consisted of 246 possible topologies combinations which has to be trained several time with various initial conditions. The pure NN technique becomes quite ambiguous due to an exhaustive search, which is almost impractical. Thus the merging of a sophisticated optimization technique with the NN technique as a hybrid intelligent system provides radical processing power to the exhaustive search and the optimal solution results are more achievable compared using the pure NN technique.

Embedding GA with ANN enabled the system to determine the best structure of the ANN topologies and the optimized input parameters. The system was able to establish non-linear relationships between several boiler parameters to determine the behavior of the parameters based on the trip. This allows the system to detect the most influential parameters related to the particular trip. This was performed based on the fitness of the GA, which is equivalent to the RMSE values. The optimization of the ANN topologies structure was determined using GA, which determined the best structure from hundreds of different combinations. Trial and error approach was replaced by GA, which can calculate the RMSE for each combinations of ANN topology and determine the optimized structure.

It was concluded that a direct comparison of results achieved by IMS-I and by IMS-II could give an insight into the efficiency of the IMSs. Therefore, a preliminary training process was performed using a pure NN in order to find the optimal NN topology. The search explorations were essentially focused on 1HL and 2HL architectures. Detailed results that were obtained by the IMS-I, together with the ones achieved by the IMS-II, are shown in Table 4.

Table 4: Optimal solution given by the IMS-I and IMS-II

IMS	RMSE	Architecture	Activation Function	Training Algorithm
IMS-I	**0.0642**	6HL1	L+L	SCG
IMS-II	0.2029	9HL1-3HL2	P+L+P	Rprop

A slightly lower RMSE of 0.0642 was observed in the first system compared to IMS-II, which obtained RMSE of 0.2029. IMS-I was able to predict the trip 5 minutes before the actual trip whereas IMS-II predicted the trip 2 minutes before the actual trip. The global optimal selections by a plant control system and the hybrid IMS-II system are given in Table 5. From this table, it is clear that the IMS-II succeeded in finding the most effective boiler operation. The ability of the system to come up with the most effective boiler operation variable is essential as it allows for an accurate prediction.

Table 5: The most effective boiler variables by plant control system and IMS-II

IMS	Variables
Plant Control System	V[5,8,9,11,14,17,**18**,**20**,22,23,24,**25**,26,27,29]
IMS-II	V[1,2,3,4,5,6,10,**18**,19,**20**,21,**25**]

7.0 ACTIONS ON DETECTING THE BOILERS TRIPS

The main purpose of being able to predict a faulty boiler operation and identify a specific trip before it actually occurs is to allow corrective action to be taken within the available time. A simple corrective action of reading some control component could restore a satisfactory operation. This is similar to the concept of adaptive control. Since the interaction of different components causes different faults that can be traced back to many alternate source, there is no need to trace the real cause of a fault.

When a failure occurs, it is possible to come up with alternative measurements to reduce downtime, which includes determining the missing measurements from general energy and mass balances. Providing substitute measurements is equivalent to providing replacement or standby instruments in case of a failure. This does not include any extra cost other than the computer cost. In case of detection of a malfunction, the process control system is manipulated by the operator to manually control the automatic loop.

A supervisory program using computer control can be devised to prevent control action being taken based on incorrect information. This is done by blocking the paths in redundancy network of the faulty equipment and diverting the information flow to a correctly functioning path. Implementation of this function is possible using a simple computer program, thus improving the plant availability in case of instrument or equipment failure. When a fault is detected, it is important to decide if the problem is in the plant or in the detectors. It is also essential to decide if the fault is permanent or just temporary. Urgent plant shutdown might be required in case of a real failure. However, in case of problems with measuring instruments, the plant can continue its operations while the anomalies are indicated to the operators.

8.0 CONCLUSIONS

In this study, the early prediction of boiler tube leak trip using intelligent monitoring system has been confirmed. IMSs were developed and compared for boiler trip prediction using pure artificial neural network system and hybrid intelligent system. Real plant data containing normal and faulty boiler operations were used instead of mathematically simulated data. Due to the importance of plant data preparation, an integrated framework for boiler tube leak trip was proposed to train, validate and analyse the IMSs. The boiler trip considered in this study was predicted by both systems before the plant control system using the best ANN topology combination which includes Scaled Conjugate Gradient algorithm, logistic activation function for the hidden node and the output node, one hidden layer and 6 neurons in the hidden layer. IMS-I was able to predict the trip 5 minutes before the actual trip, whereas IMS-II predicted the trip 2 minutes before the actual trip. However, IMS-II was able to successfully optimize the most influential operational variable and its final architecture. Therefore, it can be concluded that IMS-II performed better than IMS-I. Further development of the proposed system can be considered in which online real time data is used for prediction of thermal power plant boiler tube leak trip.

REFERENCES

[1] P. K. Nag, *Power Plant Engineering*, ed. II, Tata McGraw-Hill, India, 2002.

[2] Md. Mujibur and A. Kamal, "Failure analysis of high temperature superheater tube of a pulverized coal-fired power station", *International Scientific Conference*, Putra Jaya Malaysia, 14-15 January 2011, pp. 517-522.

[3] T. Zhong, A.T. Alouani, and R.A. Smoak, "The identification of sensitive variables of boiler tube leaks", *Proceedings on the twenty-ninth Southeastern Symposium on System Theory*, Cookeville, TN, USA, 9-11 March 1997, pp. 196-200.

[4] J. P. Herzog, "Kernel-based method for detecting boiler tube leaks", United States Patent: US 8,275,577 B2, 2012.

[5] F. D. Lang, D. A.T. Rodgers, and L. E. Mayer, "Detection of tube leaks and their location using input/loss methods", *in Proceedings of ASME International Joint Power Generation Conference,* Maryland USA, 2004, pp. 143-150.

[6] J. Ma and J. Jiang, "Application of fault detection and diagnosis methods in nuclear power plant: A review", *Progress in Nuclear Energy*, vol. 53, 2011, pp.255-266.

[7] D. Flynn, *Thermal power plant simulation and control*, London, United Kingdom, The Institution of Electrical Engineers, 2003.

[8] L. Fortuna, S. Graziani, A. Rizzo, M.G Xibilia, *Soft sensors for monitoring and control of industrial processes*, London, Springer-Verlag, 2007.

[9] A. Messai, A. Mellit and I. Abdellani, "On-line fault detection of a fuel rod temperature measurement sensor in a nuclear reactor core using ANNs", *Progress in Nuclear Energy,* vol.79, 2015, pp.8-21.

[10] T. Muhammad and Z. Halim, "Employing artificial neural networks for constructing metadata-based model to automatically select an appropriate data visualization technique", *Applied Soft Computing*, vol. 49, 2016, pp.365–384.

[11] Z. Halim, R. Kalsoom and A. R. Baig, "Profiling drivers based on driver dependent vehicle driving features", *Applied Intelligence,* vol.44, 2016, pp.645–664.

[12] F. B. I. Alnaimi, M. A. Ali, H. H. Al-Kayiem, and K. S. M. Sahari, "On-line condition monitoring system for high level trip water in steam boiler's drum", *MATEC Web of Conferences*, vol. 13, 2014.

[13] F. B. I. Alnaimi, "Pure artificial intelligence monitoring system for coal-fired power plant boiler drum high water level trip", *International Conference on Intelligent Computing and Intelligent Systems*, China, 18 November 2011.

[14] S. Kumar, R. Eldon, H. Park, "Performance prediction of a RPF-fired boiler using artificial neural networks", *International Journal of Energy Research*, vol. 38, 2013, pp. 995-1007.

[15] F. B. I. Alnaimi, H. H. Al-Kayiem, "Artificial intelligent system for steam boiler diagnosis based on superheater monitoring", *Journal of Applied Sciences*, vol. 11(9), 2011, pp. 1566-1572.

[16] K. Rostek, L. Morytko and A. Jankowska, "Early detection and prediction of leaks in fluidized-bed boilers using artificial neural networks", *Energy*, vol. 89, 2015, pp. 914-923.

[17] S. Dusan, G. Marjan and A. Jurij, "Artificial neural networking model for the prediction of high efficiency boiler steam generation and distribution", *Simulation Modelling Practice Theory*, vol. 57, 2015, pp. 58-70.

[18] T. Jingwen and G. Meijuan, "Network intrusion detection method based on high speed and precise genetic algorithm neural network", *Conference on Network Security, Wireless Communications and Trusted Computing*, Wuhan, China, 25-26 April 2009.

[19] F. D. Lang and S. Rafael, "Method for detecting heat exchanger tube failures and their location when using input/loss performance monitoring of a power plant", United States Patent: US 6,651,035 B1, 2003.

[20] E. Momeni, R. Nazir, D. Jahed and H. Maizir, "Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN", *Measurement*, vol. 57, 2014, pp.122-131.

[21] M. Rahmi, M. Hajialyani, R. Beighzadeh and A. Abdulaziz, "Application of artificial neural network and genetic algorithm approaches for prediction of flow characteristic in serpentine microchannels", *Chemical Engineering Research and Design*, vol. 98, 2015, pp.147-156.

[22] A. Azadeh, H. Shams, S. M. Asadzadeh, M. Saberi, M. Sheikhalishahi, "A flexible ANN-GA-multivariate algorithm for assessment and optimization of machinery productivity in complex production units", *Journal of Manufacturing Systems,* vol. 35, 2015, pp.46-75.

[23] G. Meijuan, T. Jinwen and J. Xu, "Building logistics cost forecast based on high speed and precise genetic algorithm neural network", *International workshop on Intelligent Systems and Applications,* Wuhan, China, 12 June 2009.

[24] D. S. Nagesh, G.L. Datta, "Genetic algorithmfor optimization of welding variables for height to width ratio and application of ANN for prediction of bead geometry for TIG welding process", *Applied Soft Computing*, vol. 10, 2010, pp.897-907.

[25] B. Yuce, Y. Rezgui and M. Mourshed, "ANN–GA smart appliance scheduling for optimised energy management in the domestic sector", *Energy and Buildings*, vol. 111, 2015, pp. 311-325.

[26] B. Majid, S. Ahmad, B. Zahra and A. Morad, "Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach", *Process Safety and Environment Protection*, vol. 95, 2015, pp. 12-25.

[27] A. Hajnayeb, A. Ghasemloonia, S. E. Khadem and M. H. Moradi, "Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis", *Expert Systems with Applications*, vol. 38, 2011, pp.10205-10209.

[28] Z. Halim, M. Waqas and S. F. Hussain, "Clustering large probabilistic graphs using multi-population evolutionary algorithm", *Information Sciences*, vol. 317, 2015, pp. 78–95.

[29] Z. Halim and Uzma, "Optimizing the minimum spanning tree-based extracted clusters using evolution strategy", *Cluster Computing*, 2017, pp. 1-15.

[30] Z. Halim and Tufail Muhammad, "Quantifying and optimizing visualization: An evolutionary computing-based approach", *Information Sciences*, vol. 385, 2017, pp. 284–313.