ALTINBAŞ UNIVERSITY

Graduated School of Science and Engineering

Electrical and Computer Engineering

THYROID DISORDERS PREDICTION USING LONG SHORT TERM MEMORY (LSTM) TECHNIQUE WITH NON DOMINATED SORTING GENETIC ALGORITHM (NSGA-II) AS RISK FACTOR FEATURE DETERMINATION

Sahar Jasim Mohammed

Master Thesis

Supervised By

Asst. Prof. Dr. Sefer KURNAZ

Istanbul, 2019

THYROID DISORDERS PREDICTION USING LONG SHORT TERM MEMORY (LSTM) TECHNIQUE WITH NON DOMINATED SORTING GENETIC ALGORITHM (NSGA-II) AS RISK FACTOR FEATURE DETERMINATION

by

Sahar Jasim Mohammed

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2019

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Sefer KURNAZ

Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Prof. Dr. Osman Nuri UÇAN.	School of Engineering and Natural Science, Altınbaş University	________________
Prof. Dr. Mesut Razbonyali	School of Engineering and Natural Science, Maltepe University	________________
Asst. Prof. Dr. Sefer KURNAZ.	School of Engineering and Natural Science, Altınbaş University	________________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Çağatay AYDIN

Head of Department

Prof. Dr. Oğuz BAYAT

Director

Approval Date of Graduate School of Science and Engineering: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Sahar Jasim Mohammed

DEDICATION

This thesis is dedicated to: The sake of Allah, my Creator and my Master, My great teacher and messenger, Mohammed (May Allah bless and grant him), who taught us the purpose of life, My homeland Iraq, the warmest womb; Diyala University; my second magnificent home; My great parents, who never stop giving of themselves in countless ways, My dearest Husband, who leads me through the valley of darkness with light of hope and support, My beloved brothers, who stands by me when things look bleak. To all my family, the symbol of love and giving, My friends who encourage and support me, All the people in my life who touch my heart, I dedicate this research.

ACKNOWLEDGEMENTS

It would not have been possible to attain this level of education without the help of a multitude of individuals. My profound thanks go to my supervisor Asst. Prof. Dr. Sefer KURNAZ, for his assistance, guidance, valuable comments and constructive criticisms. I want also to send my thanks and grateful to other lecturers of the Electrical and Computer Engineering Department of Altinbas University for their help and support me.

I am forever grateful to all members of my family especially my father for his faith in me, my mother for her loved and caring to me and my husband who supported me during this hard time. My sincere thanks also go to my brothers, for their love, support and encouragement during the entire M.Sc. program. I also appreciate the support and loved that i got from my husband’s family.

ABSTRACT

THYROID DISORDERS PREDICTION USING LONG SHORT TERM MEMORY (LSTM) TECHNIQUE WITH NON DOMINATED SORTING GENETIC ALGORITHM (NSGA-II) AS RISK FACTOR FEATURE DETERMINATION

MOHAMMED, Sahar Jasim,

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst. Prof. Dr. Sefer KURNAZ

Date: 09/2019

Pages:62

Data science is evolutional field that gained extensive attention of other technology and engineering manufacturers. One of the most popular task of data science is extraction of knowledge by analyzing a big data, it is so called as data mining. It uncovers the hidden patterns from big data that are impossible to be known using traditional calculus. Medical application is not behind these developments, data mining facilities are deployed for medical purpose as well and drawn a noticeable performance in mining the big data and revealing important information such as disease prediction, patients clustering, drugs classification, future demand prediction, etc. This project is employing of data mining algorithms for prediction of thyroid infection. Projects included essential requisites such as dataset preparation, a 2800 medical records of different patients are resourced from online portal and preprocessing are begin with substituting the missing values and encrypting the text information into numerical format. Feed Forward Neural Network (FFNN) is made to learn from this dataset for predicting the disease occurrence. Furthermore, deep learning is exploited to predict the disease occurrence from the same dataset, more likely Long Short-Term Memory Neural Network (LSTM NN) is used to learning the same. The results are made by comparison the performance of training between the two techniques. Performance is judge base on the Mean Square Error (MSE) and number of epochs taken by each to reach the best approximation. The best cost gained from using the FFNN was as (0.09659583) whereas the LSTM has produced the minimal performance after multiple iterations

of its setting, more likely a (0.03916952) mean square error is the minimal error obtained using the LSTM.

Keywords: Thyroid Disorder, Long-Short Term Memory, Feed Forward Neural Network, Non-Dominated Sorting Genetic Algorithm.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

LIST OF ABBREVIATIONS

DM	Data Mining
ML	Machine Learning
FFNN	Feed Forward Neural Network
LSTM NN	Long Short Time Memory Neural Network
ANN	Artificial Neural Network
SVM	Support Vector Machine
K-NN	K-Nearest Neighbors Algorithm
MOEA	Multi-Objective Evolutionary Algorithm
GA	Genetic Algorithm
NSGA	Non-Dominated Sorting Genetic Algorithm
T3	Triodothyronine
TSH	Thyroid Simulation Hormone
TT4	Total Thyroxin
FT4	Free Total Thyroxin
FIT	Thyroid Function Test
F_o	Objective Function
MSE	Mean Square Error

1. INTRODUCTION

1.1 OVERVIEW

Data Mining (DM) is one of evolving field in today's technology due to its contribution in large technology and engineering fields. this field of computer sciences applies sophisticated tools to predict, extract, interpret unknown information from large data. With light of DM theories, information can be derived from large unclassified data sets. The essential tools in DM approaches are lying on the machine learning (ML) where ML algorithms such as neural networks are being applied on the big datasets in order to extract a useful knowledge from data corpus. The internal process of machine learning is statistical calculations and mathematical approaches that aims to interpret the hidden relationships between the data elements and hence uncover that information for the purpose of conclusions comprehensibly defining. Data mining involves another infrastructure which are useful for data preparation. In order to predict the hidden patterns of data, it should be with known classes. However, classification of data is another task of data mining and here data must be prepared in such way that make it suitable for process of data mining. Data pre-processing is one major and important step of mining the knowledge from data. This project is keen on one a specious application of data mining that interleaved with human life which is known as disease prediction. It is a medical exploitation of data science in order to find a way with hug records patients in hospitals. More specifically, thyroid disease prediction is the main agenda of this project. The medical records of various patients are executed at the hospitals, records contained different test that performed in favor of diagnosing the thyroid disease of large number of patients. In different time slots, the data of medical test and their particular diagnosis information are made in form of dataset that includes large number of patients with their test information along with the decision of doctors of every case. Recording this data for long time made it difficult to interpret some characteristics statistically in the manual classic method. The main goal of this project is development of smart paradigm that capable to conduct disease prediction on the biases of the received data. In order to implement this stat of art, arterial intelligences adopted, more specifically, the neural network is considered as corner stone of this work. Two types of machine learning are used with efforts to elect the minimum loss theory that is yielding the best cost. The first paradigm is designed using the Feed Forward Neural Network (FFNN) and hereinafter deep learning paradigm is developed using the Long Short Time Memory Neural Network (LSTM NN). Data are resourced out from university of California data system. These sections are hereby detailing the methodology that relating the deep learning algorithms and data base preparation strategy.

1.2 PROBLEM STATEMENT

Literature contained large scale of approaches in favor of Artificial neural network in various field of technology. The medical application that based on neural network is demanding enhanced performance in the sense of error at the output. Artificial neural network deployment in technological and scientific application is still under evolution and more care should be taken to implement the neural network that is capable to solve complex engineering and scientific problems. The problem of classical neural network is their deficiency of solving time related tasks as the output of layer is not affecting by the previous gates output, in other word, classical neural network is poor of memory so it can’t be utilized for time tasks. The more advanced version of neural network that made capable to cope with time formulas such as Recurrent Neural Network is shown deficiency in learning long term dependencies.

1.3 OBJECTIVES

For the purpose of development best classifiers and predictors with neural network, this project is aimed to perform the following:

1. Performing of efficient preprocessing steps to fulfill the dataset analysis requirement with the purpose of reduction of error at the outcomes.
2. Testing of classical neural network in learning problems with light of big dataset.
3. Using Forward Neural Network to Learn the Thyroid disease existence in large dataset of medical records.
4. Employment of deep learning techniques to predict the thyroid disease in big data base.
5. Modification of deep learning algorithm in order to optimize the performance of learning.
6. Highlighting the effect of deploying different numbers of nodes at the neural network layers on the outcomes performance.
7. Highlighting the impact of varying the number of layers of neural network on the outcomes performance.
8. Conduction of statistical calculations to compare the performance of different experiments for drawing the conclusion.

1.4 PROPOSE METHODOLOGY

In this project we propose using a deep learning technique more likely Long Short-Term Memory neural network to conduct prediction of thyroid disease base on large medical records. As dataset is being downloaded, preprocessing must be conducted to derive a suitable format of the dataset that can optimize the results of prediction in neural network. In order to compare the results of deep learning (LSTM) neural network with the classical neural network, a Feed Forward Neural Network (FFNN) can be used to perform the prediction of the disease using the same dataset. It was understood from the history of using a neural network that data normalization is must for deriving the results and to avoid running through problems. So to say, dataset is to be normalized such that Boolean expressions and other alphabetic cells contents is to be converted into numerical format (encryption). The missing values in the dataset is to be replaced by the mean of the total elements in its particular column. Ultimately, results to be judge using the performance metrics such as mean square error and epochs number.

1.5 THESIS ORGANIZATION

This dissertation is consisted of five technical chapters where the details of the thyroid disease prediction are made available.

- The first chapter entitled as "Introduction" is contained the outline overview of the project and the details of problem statement with thesis objectives.
- The second chapter entitled as "Literature survey" is contained the previous trials made in favor of data mining using deep learning and classical neural network.
- The third chapter entitled as "Methodology" which keen on explaining the methods used to establish this infrastructure.
- The fourth chapter entitled as "Results and discussion" that describe the outcomes obtained after completion of projects steps.
- The fifth chapter entitled as "Conclusion" which conclude the facts obtained after completion of project.
- Last section is enlisting the references used to construct this dissertation.

2. LITERATURE SURVEY

2.1 ARTIFICIAL NEURAL NETWORK ANN

Neural network come to image when application of cognitive nature exists; since demand is raised to produce future event prediction technologies, the deal with neural network become more feasible especially when large neural network libraries are integrated to several platforms alike Python and Matlab. In this chapter we are going to derive through using neural network to tackle time series inputs that is prescribing a nonlinear phenomenon. Large efforts are paid to define the optimum configurations of the so-called network for approaching a better performance. In sections of hereafter, long short-term memory neural network is been emphasized for timing applications. In meanwhile, problem of delay and computational cost is raised. It is noteworthy that delay problem is tackled as well using segmentation techniques so to say:

At [1], in order to modulate time variant event, time variant system is required, more likely, an event changes with time such as system of water quality prediction. Traditional neural network may not server the purpose of this interest. A long-short term memory neural network is exploited in time variant systems. In this study, water quality prediction model is established using LSTM NN, the model consists of two sections: the training phase and the testing phase more likely, seems as traditional structure of neural network classifier. The training set used here is big data set from lake, dated from 2000 through 2006 which reflects the level of phosphorus and oxygen dissolving rate in the water., the readings of those parameters are monthly observed and recorded for six years, the same is used to train the model of the predictor. Technically, LSTM NN is recurrent neural network with addon feature in the hidden layers. In order decide the predictor parameters such as number of neurons and iterations number, author mentioned that several iterations were executed by varying the above parameters and however, a twenty-five epochs with fifteen neurons are shortlisted. The key measure of performance in this project was the number of errors that termed as “MaxError”.

At [2], another application of LSTM NN is demonstrated in this study where computer autonomous detection system is implemented and is based on LSTM NN. The experiment begins with using deep neural network as said by author; the same was reported not suitable for time variant data more likely computer network unforeseen data. Another experiment was conducted using recurrent neural network by freezing the length of data so to say, data was converged to fix length multiple block equivalent to variable length of data sequence. The procedure to detect intrusions in network data can be described by passing the data through a sequential auto-encoder, the use of twofold encoder is for feature extraction and data size reduction. The data of this study is seven days payload of computer network. It involves a two million of connections related to HHTP, SMTP, IMAP, FTP; this payload is diagnosed with five autonomous connections. The connection is said to be none-legitimate if the data packets are not found for this particular connection in either the source port or destination port or in both. The autonomous occurrence rate is expected to be similar after using the proposed technique. This experiment is totally relied on the sequential character of the payload at the source and destination; so, this character is a hexadecimal identifier of 46 digit, the same is passed through the encoder to be converted into decimal series where it can be split into training set and testing set (LSTM NN candidates). Data is trained using Adam training method and mean square error MSR is the deployed objective function.

At [3], due to expansion of social network, target sentiment classification has become insisting demand of community. Deep learning neural network is corner stone of sentiment detection technology. This application is termed as time variant event where the sentiment detection is target dependent process. This study revealed that long short-term memory is not a good choice for target sentiment detection system since it is focusing of single target sentiment process and fail to work with multiple target at the time, furthermore, training time of LSTM NN is bigger for applications of this nature. Author mentioned that such problem is tackled using a combined Regional LSTM neural network with convolutional neural network so the training time can be reduced. The reduction in training time is termed to the segmentation of the data which is done using Regional LSTM neural network and segments features can be expressed using convolutional neural network to prevent any bad segmentation that might be happened. Segmentation is all about preserving the important features in the data and ignoring the rest so that classifier will neglect this data at the time of training and testing. The results obtained from this experiment are said as enhanced results as compared to other available neural network such as support vector machine.

At [4], another application of LSTM neural network is studied over this study where this classifier is used to analysis of radar information more likely, to study over HRRP high resolution radar range profile, this profile can be set after analyzing the backscattered electromagnetic wave reflected from target, it comprises of four columns and large number of rows, the row in every column represents the target from particular angle (since target is mobilized, angular data may get varied with time) so to say, it comes under (x,y,z) coordination and another fourth variable that represents the speed of target. LSTM itself can save the information about the inputs for long time and prevent same error replication, the use of memory in this kind of neural network make it capable to perform memory operations such as writing, reading and deleting, memory operations are achievable by setting the gate of LSTM neural network. The gate LSTM neural network can be instructed to write or read a data even delete them if required. Microsoft network computational tool kits is used as simulator in this study because it's provide the algorithms to meet system requirements in the simulation.

At [5], LSTM neural network have been used as predictor of faults in rotating machinery by analyzing the vibration signal information, data is out sourced from a bearing data set and used in this study with objective to analysis the bearing, rolling and failures. However, study revealed that recurrent neural network is more adaptable with time sequential information and many experiments had reported the deficiency of recurrent neural network when gradient or disappearance of input data are increased. To tackle the problem of data gradient, author reported using long short-term memory neural network. Feature extraction of mechanical contents can be performed using of deep belief network, the same is combined with LSTM neural network to achieve better performance where data features are extracted first using of deep belief network. LSTM NN of three layers is being used in this study and enabled to perform smooth prediction of error occurrence on the basis of selected features from previous step, network is made to run two hundred epochs with 40 percent of input as testing data and remaining 60 percent as training date. Deep belief network uses restricted Boltzmann machine to calculate the feature, it consists of hidden neuron layers and visible neuron layers, those layers are not connected to each other and the inter layer neurons are even not relevant so that, deep belief network can calculate the expected values easily.

At [6], neural network is used to construct time related applications, in here it used to form a language model in speech recognition system. Language model is essential unit of speech recognition system where the uttered speech can be rectified. The wrong word utterance is major problem in automatic speech recognition system. In real life, many speakers are found with plenty of ways of pronunciation and in top of that, wrong utterance and shortcuts are more likely seen within speech recognition field. Language model is an cognitive sector where the real meaning of the utterance are made base on whatever pronunciation acoustic input. This approach i.e. [6], proposed comparing the performance of three neural network approaches more likely LSTM NN, recurrent NN and feed forward NN. The key point of performance metric is the rate of error and perplexity. The resultant perplexity was increased with enlarging of hidden layer neurons however, with six hundred neurons a performance recorded as 112.5, 108.1 and 92.2 in case of feed forward neural network, recurrent neural network, long short-term memory neural network respectively where each particular was used to build up the language model. It is noteworthy to say that more neurons resulting a large perplexity which is undesired parameter in automatic speech recognition.

At [7], weather forecasting is vital to human life wherein natural disasters can be predicted and required precautions to be initiated accordingly, on top of that, essential industries and agricultural sectors are explicitly dependent on weather forecasting information. This application is about accurate calculation of weather parameters and accurate prediction of upcoming parameters. This study analyzed of six weather parameters to predict three parameters from the other hence more likely, temperature, pressure and humidity by learning over existed: wind speed, temperature, pressure, precipitation, humidity and visibility. Long short-term neural network is used to predict the weather future status. System proposed in this study seems to share same structure of most traditional LSTM NN systems where data is enquired from 2007 through 2017 more likely a ten years of weather record is used. The results of this hierarchy are matched with previously made trails to forecast weather more likely, artificial neural network and recurrent neural network, author reveals that he obtained enhanced outcomes as compared with the mentioned techniques. No numerical parameters were revealed in the study alike hidden layer neurons, epochs and performance measure.

At [8], an approach to monetarize the unmanned aerial vehicle attitudes is made, this system is termed of its complexity due to participation of large parameters to perform its model. In order to analysis of unmanned aerial vehicle behaviors are time series data with nonlinear relationship so that, LSTM NN is used to predict the attitude of this system based on current data selected while flying phase of quadcopter. The selection of LSTM NN is done by referring the previous studies that emphasized on their capability to modulate time series information, however, varying of LSTM NN configuration may lead to some performance measure; so to say several iterations (experiments) with different parameters are executed to reach a conclusion. Furthermore, data collection of flying element such as quadcopter can be collected using one of following three methods: cosine method (consumes big calculation power which makes it unsuitable for Realtime applications), Euler angle method (rotation of particle can be estimate by this method but it cannot evaluate the particle full attitude) and quaternion method (it is suitable for realm applications demand with reasonable capability to calculate the attitude of particle). Eventually, parameters of predictor had varied from ten through twenty five epochs and from 4 through 9 seconds of training time and for 10 through twenty neurons in the hidden layer, however, performance found optimum at fifteen neurons, twenty epochs and 7.2497 of training time. Results measured as mean square error of (pitch, yaw and roll).

At [9], approach to convert none expressive music format into expressive music format using a LSTM neural network. The performance testing of this approach is achieved by comparing the resultant expressive music generated by LSTM NN with natural music produced by human. Outcomes of this approach is about establishment of LSTM NN based music generation system approachable to human performance. System is expected to produce a vector of to variables: local tempo of music signal (which means the speed of musician that particular time) and loud amount of the music signal. This vector is well producible after accurate training of the model, author mentioned that he had referred a previous study employing the kernel method and Canonical Correlation Analysis method; the tuning between local tempo and loud of signal is required in such way maximum correlation between the both can be obtained, hence, the above methods have been reported as best method to obtain that level of correlation, however, an adequate correlation is obtained by doing the same as revealed by the author. Data training is performed by firstly feed the data into system from large music database hereafter, training of the model is perform using a python-based library called Keras deep learning.

At [10], a comprehensive approach to state the differences of standard long short term memory neural network and same neural network with varying the number of hidden neurons and bias. In other word, three LSTM NNs are made more likely $LSTM_a$, $LSTM_b$ and $LSTM_c$ each has own configurations. This study made for validating LSTM NN performance by testing the output of LSTM NN by changing the gate signal. Gate is usually combination of input signals and some parameters called as associative parameters. The experiment involves formation of three gated LSTM neural networks by changing the gate equation as follow: gate1 is changed by removing the input signal form the formula, gate2 is formed by elimination of bias and input signal from the gate formula, gate3 is formed by elimination of hidden unit parameter and input signal form gate formula. Four experiments were executed which involved the standard LSTM NN and the previously mentioned three verities of LSTM NN, the performance metric used in this study was the learning rate coefficient which is compared in total four cases. Outcomes of the study were like: coefficient was 0986 in case of standard LSTM NN where as 0.961, 0.752, 0.424 in $LSTM_a$, $LSTM_b$ and $LSTM_c$ respectively. Know the modification of gate combination in above, we can conclude that gate signal must combine all thee parameters more likely involving the input signal, hidden layer information and bias information, that is reflected in the obtained results which reaves maximum learning rate coefficient in case of un modified gate LSTM NN.

At [11], LSTM is used in stock transactions to predict future stock sale and however, the same is conducted using data set of garments transaction content such as price at the end of time, minimum price, price of transaction the stock down limit, minimum price and up lint. This data is labeled by data so that predictor can work for future data base on the provided back dates. This study reveals that LSTM neural network of hidden layer higher than five is consuming big calculation power and hence producing a delay proportional to number of hidden layers. The results of this study obtained by finding the sample accuracy rate which means the rate of producing a predicted data encountered for error rate. An accuracy rate of sample of 0.66 is reported for one-layer LSTM neural network, whereas, three layers LSTM neural network is produced accuracy rate of sample equal to 0.78. That means, accuracy rate of samples can be increased by increasing the neural network hidden layers with respect to the fact of computational power consumption with large number of hidden layers. Author mentioned that 72 percent is the general accuracy of the proposed model where this accuracy is not actually up to mark and more modifications are possible to enhance the accuracy.

At [12], mechanical features are extracted from vibration signal (mechanical related terminology) and processed further by long short-term neural network for predicting the mechanical status of motor. The vibration signal is obtained from special sensors of the motor corpus, such signal is including a large none stationary data related to none linear variation of mechanical parameters which is difficult task with traditional neural networks. LSTM neural network model is established to cope with such time series signal and however to predict of future status of motor. The data is firstly decomposed as a preprocessing step before using this data to train LSTM neural network. There are various methods to decompose the time signals and decomposition of signal means to break it down into sum of intrinsic units and residual units. So, techniques alike Fourier transform can decompose the time signals into frequency components where that is not required in mechanical application like this whereas another technique is employed to perform decomposition which is called Empirical Model Decomposition. Furthermore, a two predictor is designed to generate the results: support vector regression model and LSTM neural network. Results are obtained with mean square error preferences which is found twofold in LSTM as compared to support vector regression machine. It is noteworthy to mention that three hidden layers are used in this model (LSTM NN model).

At [13], identity initialized recurrent neural network is another approach that expected to overcome the problem of gradient in training. This alternative is reported as far from expected performance when negative data is inputted at training phase, however, the training of negative data may stop the back propagation of gradient and then resulting error and degrade the performance. The study of here is proposed to classify images from big data set; hence, data used for training phase is 50000 images and data used for testing phase is 10000 images and another 10000 images are used for validation. The proposal involves using of two layers recurrent neural network and separation of those layers so every layer will run individually so that performance of recurrent neural network can be enhanced. The results are compared with standard recurrent neural network and the method of proposed is drawn a noticeable performance; the simulation is run for 100 batch size and 500 epochs.

At [14], a long short-term memory neural network is stated to have average delay propagation due to hidden layer structure. However, another model can be used to substitute the LSTM neural network with convolutional neural network by developing a supper configuration in it.

this method can serve the propose in some application and not every application since dealing with none stationary signals is required strong model that tolerate time variation and none linearity nature of data. More likely, date is segregated so that larger portion is reserved for training and below that portion is usually used for validation and testing, it usually happened to have same data length in both validation and testing stages; so this study dedicated of 10000 sample for testing, 10000 for validation and 50000 samples for training. Obtained results have shown that accuracy of training can be enhanced by increasing the epoch numbers; in this approach author used 1, 10, 100 and 300 epochs respectively to monitor the accuracy fluctuations and however, it found to be 89,4 percent with 300 epochs and 27.06 percent with single epoch so to say, the more epochs the more accuracy.

2.2 FEATURE SELECTION ALGORITHM

Different data mining techniques are used and described as mentioned for diseases prediction/ classification but needed more improvement and enhancement to their classification ability. Feature selection algorithms are also introduced as a main part of these data mining improvements and represented more with a researcher works as in [15]. It showed that, an improvement on SVM technique are occurred with a feature selection algorithm compared to traditional SVM or even to an improved version of SVM with differential evaluation assistant. In [16], assumption of genetic algorithm as attributes selection are identified with a Bayesian combined algorithm for referring and assigning the missed / ignored possible individuals. It showed that, the classification model performance with accuracy calculations are increased using these schema as algorithm for selecting feature compared to the traditional technique. In this paper a small size of 8 data sets are used with accuracy and missing rates calculation to improve that based algorithm.

Implementing such an algorithm are also taken more interest in multimodal biometric data. In [17], an extracted data related to physical or behavioral data of persons are related with three different type for feature selections. The proposed algorithm of a hybrid schema shown a soft and solid in the classification process compared to other techniques. Feature selection needs a classes label or classes identification are difficult to be applied on both supervised and unsupervised learning. In [18], the author assumed a new algorithm to be combatable with these two learning algorithms by testing it on a various data sets. The testing process was for classification and clustering as well, showed a better reduction of data without affecting on the goodness of the using process.

Combinational of data mining techniques and feature selection algorithm are suffering from a deadlock of local optimization and a lack in prediction time as a disadvantage state. In [19] and to achieve a higher performance for data subsets a whale optimization are suggested for solving such a problem. The assumption binary algorithm improved the overall accuracy and the speed as well by comparing it with a three traditional techniques.

Another algorithms are produced and studied for solving the problem of the classification a new classes which that many supervised learning algorithms such as K-NN are suffering from. In [20], a selected diabetes data are applied to the information gain process. It showed that one attributes can be deled with as a main feature for all type of classifiers. The selection process was among 8 features and this paper claimed that the glucose level are the most affective feature for all the classifiers.

A trained data are interested with a semi- supervised learning and applying a feature selection to the capability of classification as in [21]. In this paper, this model are introduced for drug activities prediction due to its cost and plenty of unknown classes as well. A developing analyzing system are studied for 7 features with a multi- objective algorithm for selecting the important and affected features as in [22]. The performances of this model are different with a various of datasets and solved by using a grading system. Second step of the proposed model is to handled these proposed and obtained data to an ANN to be deled with it. These assistant algorithm improved a good performances for other application not only for disease prediction as in [23]. In this paper, an assumed selection algorithm are proposed for text classification by combining with linear SVM. Other researchers preferred combination of more than one algorithm for classes prediction as in [24]. It proposed two method for selected a good feature to diagnose a cancer disease and showed a better results than the comparing techniques.

2.3 MULTI-OBJECTIVE EVOLUTIONARY ALGORITHM (MOEA)

Data mining techniques guided by Multi-Objective Evolutionary Algorithm (MOEA) has provided more accurate, reliable and interesting measurement. The strengths and changes of combining data mining with a MOEA as a confident rule are presented and explained more in [25]. It showed that, the measurement rules which obtained by MOEA are more significant and cannot be obtained using data mining techniques only. For extracting more information to built any technique in disease prediction a MOEA are used and studied as in [26]. It showed that in spite of a good results of GA with bi-clustering data, but NSGA gave more improvement with cross over and mutation precise selection.

Many fields are now interested and more described with NSGA and especially the second version of it due to its fast optimization and accuracy. This techniques recently used for a different types of problem with a various objective as minimizing cost , maximizing rating score for a special types of facility layout problems as in [27]. Another optimization of problems as parametric process are produced for more than one parameters and for minimizing losses, cost and maximizing efficiency of the selected design as in [28]. Improvement of selected objective to be applied with NSGA-II are more and more increased for more fields and problems as image watermarking optimization, determine the kernel parameters of data mining techniques as in [29][30].

3. METHODOLOGY

3.1 PREFACE

This chapter is detailing the setup of two predictors based on Feed Forward Neural Network and Long Short Time Memory Neural Network for predicting the existence of thyroid disease. The process may begin with data preparation and preprocessing such as encryption and missing values processing and then training the neural network to predict the disease in various modifications. The last section of this chapter is discussing the metrics of performance measurement in order to compare the results from different experiments.

3.2 DATASET PREPARATION

The data is about patients records of thyroid medical investigations along with the patient's name and information. This data is available for public in the database of university of California [31] so, it has been downloaded and prepared for further process. The records shown that 2800 subjects were participated in this dataset, the details of the participants are listed in Table 3.1.

Table 3.1: The dataset Content Explanation.

Particles	Value
Number of candidates	2800
Average age of candidates	51 years
Number of male candidates	1940
Number of female candidates	860

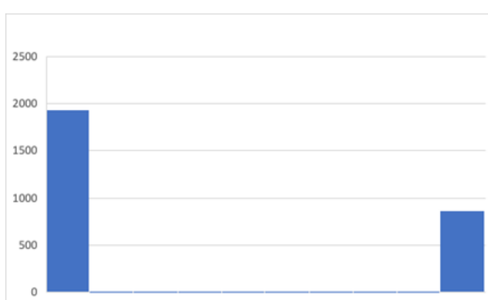

Figure 3.1: Male and Female candidate distribution.

The common attributes of the said dataset can be listed as below:

1. the top most attribute that dominate almost whole data is in Boolean format in which tells about some information alike "whither patient was undergone any medication of antithyroid, thyroxine information availability, whither the surgical intervention was done before for thyroid treatment, status of pregnancy, status of other sickness if available and finally the I131 medication availability", such queries are clarified by using Boolean format more likely true and false.
2. the other information more likely test of some hormones alike T3, TSH, TT4, T4U, TBG and FTI are made available by numerical values.
3. all patients are allotted with special ID in order to recognize them within the pool of data so, individual column has been created to accommodate the candidate's ID.
4. it was realized that some values are missing from the data sets in all columns and the missing values were replaced with question mark symbol.
5. the data in whole is contained from thirty-one column and two thousand and eight hundred rows. Some the columns were omitted from the process such that those columns which are holding unusual information such as patient ID, age and unknown information.
6. the first step of dataset preparation is conversion of data from comma separated values in to cells.
7. second step is the data preprocessing in which undertakes conversion the data elements which are holding a non-numerical information into numerical information. This step involves converting the Boolean data in to zeros and ones instead of true and false. Also, male is encoded by 1 and female is encoded by zero.
8. the cluster head data that reflects the decision-making information are found alike "negative" and "positive" that indicate the occurrence of the disease or not. However, such information are encoded as ones for the positives and zeros for negatives.
9. in order to achieve higher possible accuracy in data training, the mission values are replaced by the average value of the column elements.

The process of preparing the dataset for the training are demonstrated in the hereafter Figure.

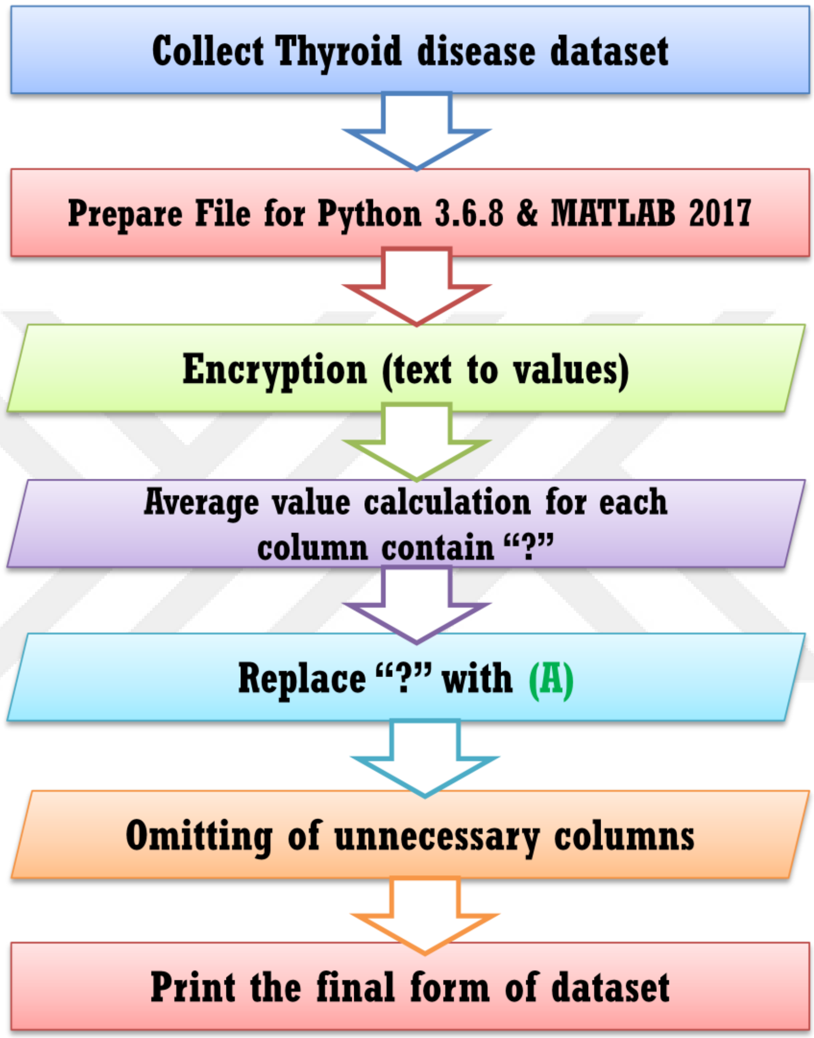

Figure 3.2: Functional steps of data pre-processing.

3.3 NON DOMINATED SORTING GENETIC ALGORITHM NSGA-II

The main steps of thyroid disorders predication using MOEA for rows and features selection are:-

1- Collecting thyroid patient's data after preprocessing.
2- Dividing data into two sets for training and for testing the classification performances.
3- Transforming data into a form compatible with MATLAB program .
4- Applying NSGA-II for rows and attributes measurement performance.
5- Applying different algorithm for mining data as thyroid predication system. In this Thesis FFNN and LSTM are mainly used for this prediction system.
6- Calculating classification performances using some parameters as accuracy, sensitivity , …etc.
7- Comparing the classification training data with the testing one to get a better performances.

Twenty eight of attributes for each patients are used five of these attributes are double numbers. NSGA-II as a fast and elitism MOEA because it could find a wide and spread number of solutions near to the classes. NSGA-II are presented and modeled with determining the non dominated solutions and the calculating of crowding distance between individuals as in [32]. The main NSGA-II procedure are shown in Fig. 3.3, in this figure the patient's data are divided into parents (P) and children (C) for the first generations. It is also shown the different number of attributes (An) selected for the model with Pareto-Front specification to predict the 8 classes.

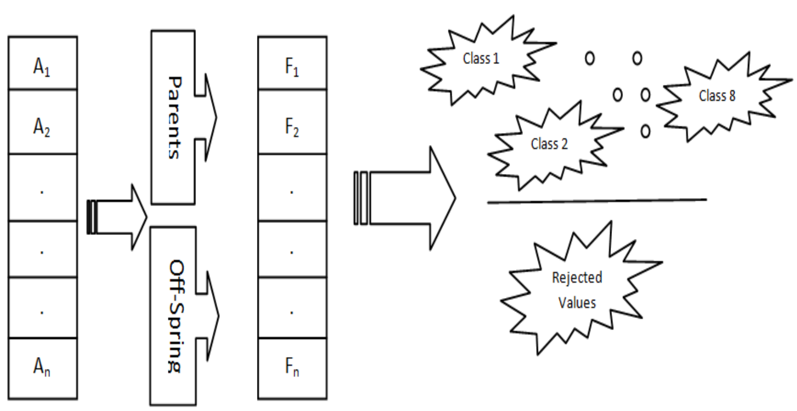

Figure 3.3: The Proposed NSGA-II for Features Selection with Determination of Pareto-Front Individuals for 8 classes

The flow chat of NSGA-II with data mining techniques are show in Fig. 3.4, after collecting and transforming data, NSGA-II own parameters are justified as no. of generation =500 , cross over = 0.4 and the mutation probability = 0.1. First generation are selected randomly form the patient's data , then determining the Non-sorting individuals. Calculating of crowding distance between these patient's are measure as a second step. Combining the first generation with the second as a new generation, then and as a last step calculating the performances of model and comparing it with the previous one until getting a higher and better performances.

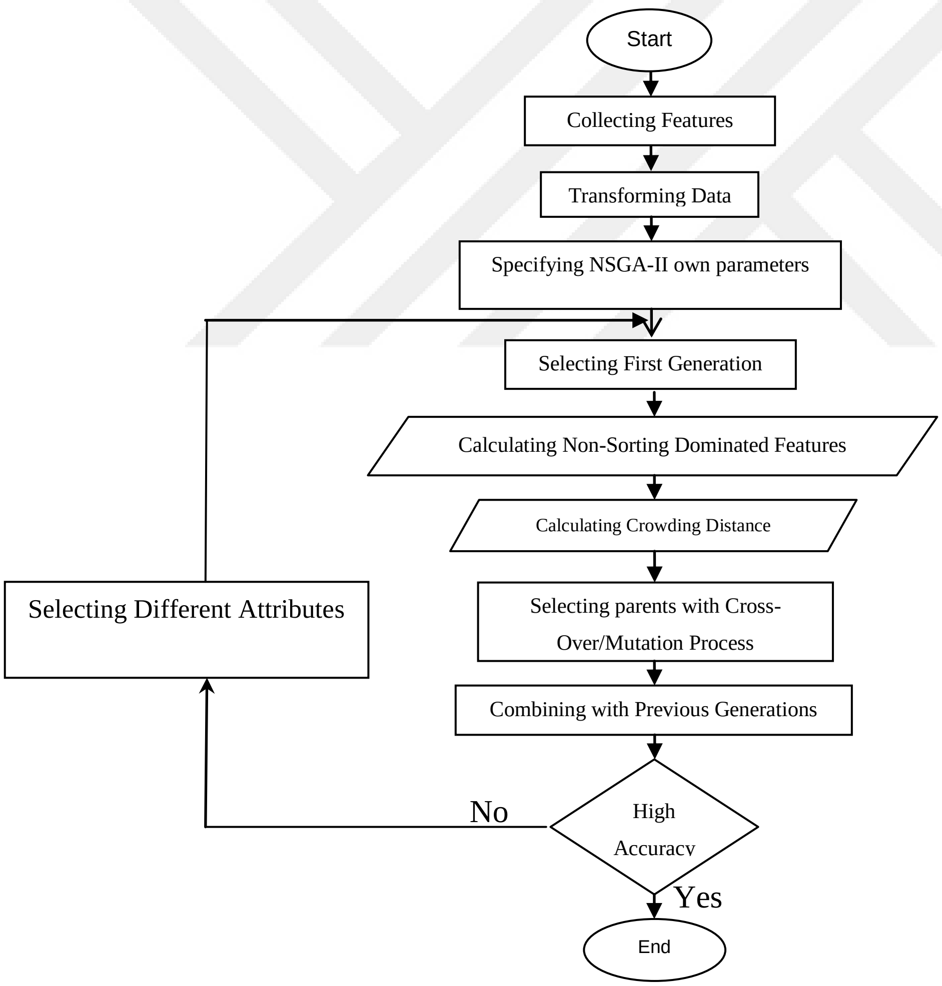

Figure 3.4: NSGA-II Feature selection algorithm for Thyroid Disease Prediction and Classification

The sum weighted model of NSGA-II used as a mathematical model for the feature selection as in Eq.(3.1).

$$F_o = \{ \textstyle\sum_{i=1}^{n} A_i T_i \quad \xrightarrow{Compare} \quad \sum_{j=1}^{n} B_j Q_j \} \tag{3.1}$$

Where F_o are the objective function which represents four classes, A and B are constant and larger than zero for unifying between different parameters. T and Q are the selected patient's attributes, while n is the number of total attributes. The mathematical model was building according to the range of each attributes by mainly focusing on five attributes. The range of five parameters are Thyroxin (T4) , triodothyronine (T3) , Thyroid simulation hormone (TSH), free total thyroxin (FT4) and thyroid function test (FIT) ranges according to collected data are shown in Table 3.2.

Table 3.2: Selected Thyroid Ranges For The Proposed Algorithm And According To The Collected Data

Parameter	Range
TSH	0.4-4.0
T3	0.7-1.8
TT4	80-220
T4U	5.4-11.5
FTI	260-480

After minimizing and detecting the risk factors(affected factors) on the classes results for optimizing and enhancing the used techniques, NSGA-II abstracted featured will be followed by the data mining techniques used as shown in Figure 3.5.

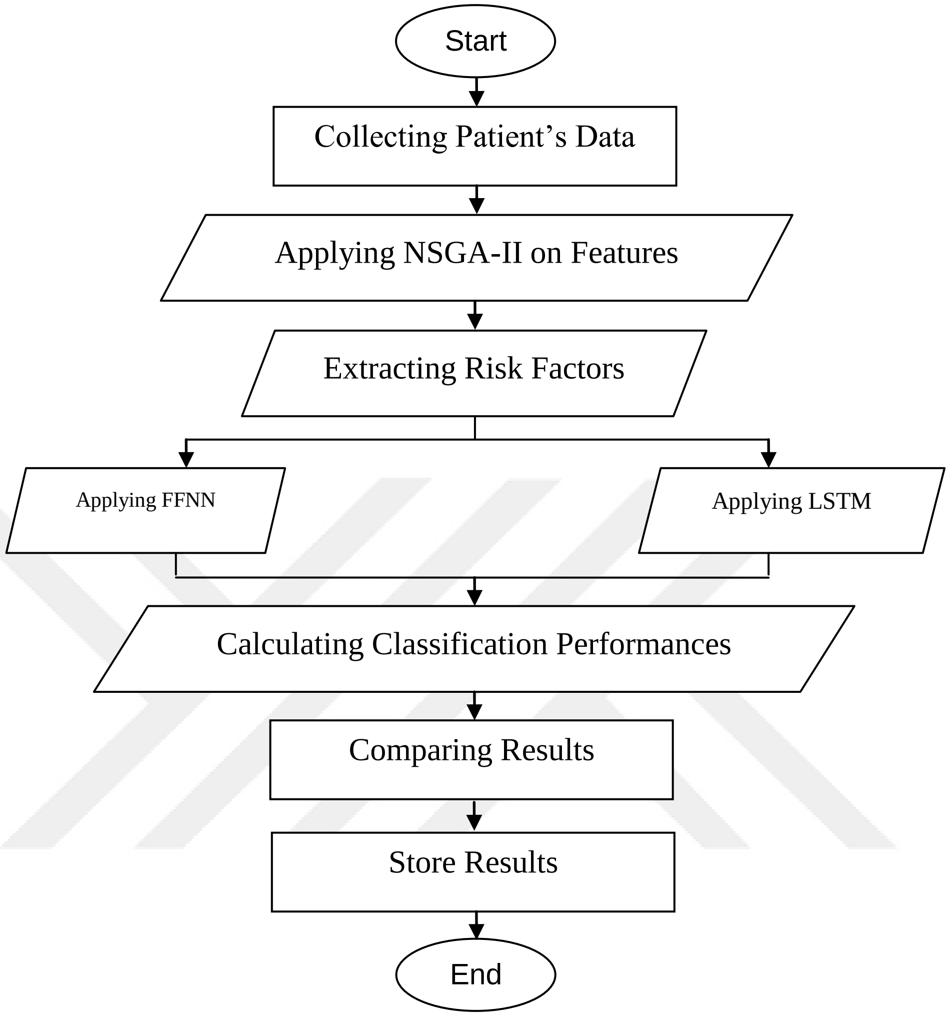

Figure 3.5: Thyroid Disease Diagnosis system by using a Different data mining techniques based on NSGA-II as a feature selection.

3.4 FEED FORWARD NEURAL NETWORK

A kind of artificial intelligence, neural network is located of the top of efficient machine learning technologies with create ability of learning from big data. The feed forward neural networks intake the data at the input layer and process it throughout the network across the hidden layers until it reaches the output layer. The data values are travelling across the network from one layer to another layer with help of neurons that mathematically hold a two values more likely bias and weight. The neurons are connecting the layers nodes with each other and hence with the next layer. In Feed Forward Neural Network, no connections are existed between the (n+1) layer and (n) layer; in other word, feedback of the layers are omitted and data keep are mapped in the forward direction without reverting the previous layer about the data status in the current layer. Figure 3.6 depicts the structural concept of the so-called Feed Forward Neural Network.

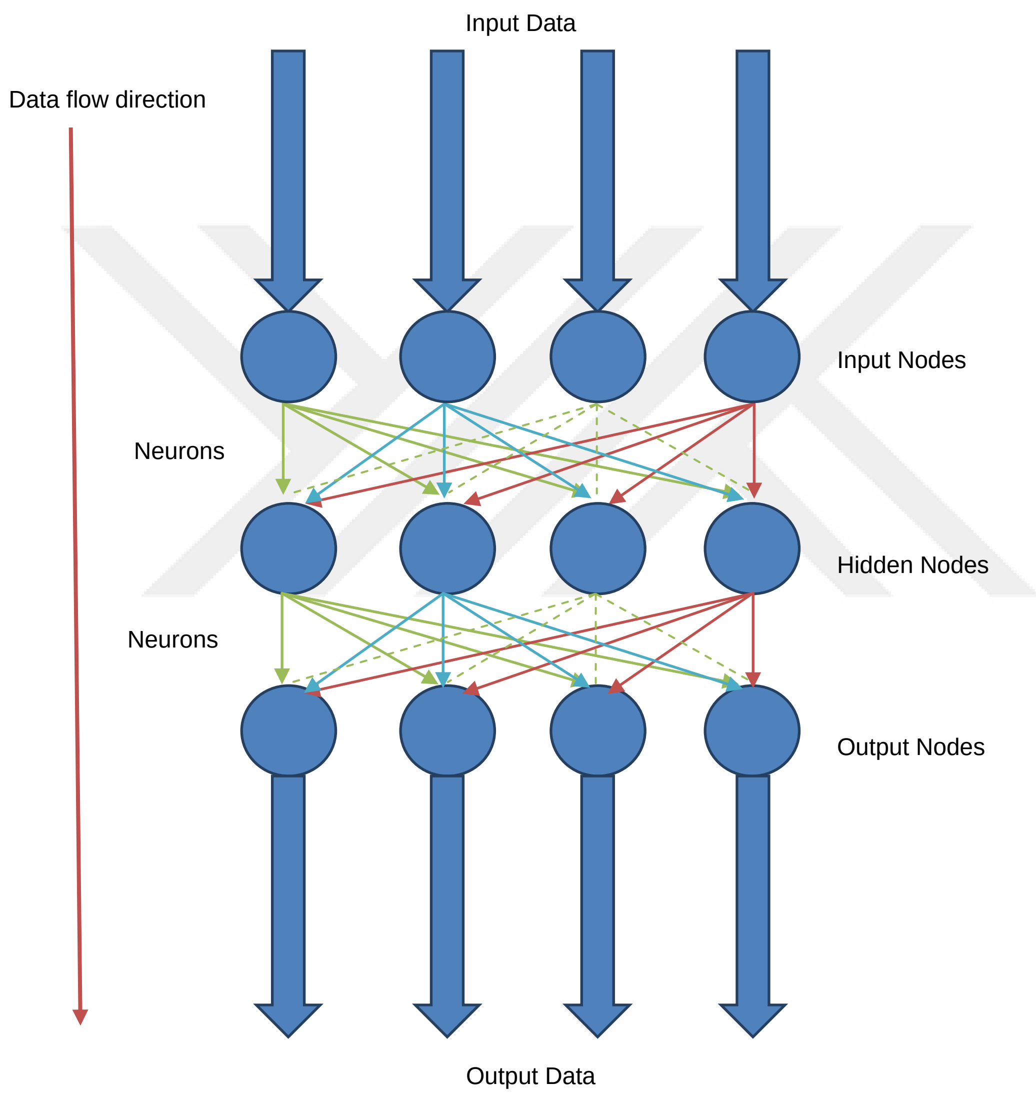

Figure 3.6: The Feed forward neural network structure.

The nodes are equivalent to a transfer function that perform some approximation of the input data and produce the input with particular changes. The main task of this neural network is to map the input data to the particular correct class with minimal possible rate of error. Let the input vector to be X_n then following formula can be written to express the process of the data.

$$Y_n = W_i \cdot X_n \tag{3.2}$$

$$Y_n = (W_i \cdot X_n) + b \tag{3.3}$$

Where:

Y_n: is the output of the node

W_i: is the weight that connects the current node with input node

n: represent the node number

i: represent the neuron number

b: is the bias of neural network

The practical model in implemented with light of the Table 3.3. that enlist the model information:

Table 3.3: Feed forward neural network model information

Particles	Details
Number of hidden layers	2
Number of input neurons	200
Number of output neurons	1
Learning algorithm	LM
Weight generation method	Random
Training data ration	70 %
Testing data ration	30 %

3.5 LONG SHORT TIME MEMORY NEURAL NETWORK

LSTM neural network is another concept of the available neural networks that defers from the FFNN as input of n^{th} layer is directly proportional to the previous input of the same layer. This kind if neural network develops same concept of memory where time is main player of its functions. Long Short Time Memory neural network is likely keen on time problems and time relationship in the input data so to say, it has gained great attention in deep learning applications. The core concept of LSTM is the feedback propagation or the memory of knowledge which means it can approach from the human way of thinking. Well, if human is reading a text, he is going to understand whole idea of that text base on reading the whole document, in other word, human is understanding the information of particular section of document by remembering the information of previous sections. This is totally different from the way of learning in the classical neural network. The method of LSTM neural network in basically derived from a recurrent neural network that simply demonstrated in Figure 3.7.

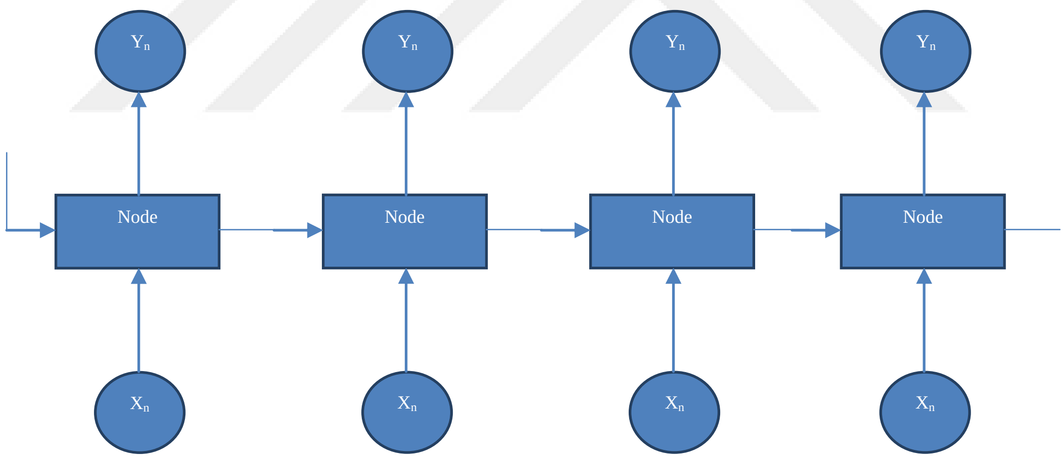

Figure 3.7: Recurrent neural network structure.

The above figure depicts four gates (layers) recurrent neural network the output of each gate is impacted by the previous gate output. For example, the second gate (2) output is accounted for the second input X_2 and the gate (1), same way, the output of every gate is generated with consideration of previous gate output. This is called memory approach of learning and this way is the best learning method for time series problems where the input is changed in accordance with time. The Figure 3.8 is demonstrating the outlook of recurrent neural network.

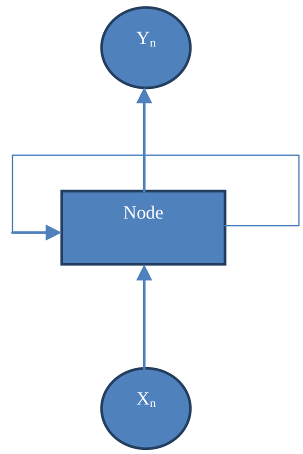

Figure 3.8: Recurrent neural network outlook.

In here, it is noteworthy that long short time memory neural network is a special modified version of the recurrent neural network.

3.5.1 Long Term Dependency

Recurrent neural network is the best option for learning the problems of time related matters, it is however can learn through a data in order to reach some settlement. The long-term dependency is more likely a learning challenge in the way of recurrent neural network. As an example of long term and short term dependency problem, a language model of automatic speech recognition project that aims to predict the last told word in the sentence a like "honey is production of bees" so, if we are willing to use a recurrent neural network to predict the last word of this sentence, the result will be pretty satisfactory as network will predict the last word as "bee", the same is termed as short term dependency. In other cases, such as the perdition of the last word in the sequence alike "I love to learn language alike English", so to say, in order to predict which language is meant by the speaker, here recurrent neural network may not be suitable for mapping the same in to English, such is termed as long-term dependency. An approach was produced to tackle the long-term dependency more likely to develop a neural network that is capable to use particular information from the previous gates in order to process the current input. The same is promoted by using the Long Short Time Memory Neural Network in order to define. LSTM neural network can refer a particular knowledge from previous gates for supporting the current input decision. Such approach is proven efficiency in tackling the long-term dependency problem.

3.6 PERFORMANCE MEASURE

The thyroid dataset is pre-processed and made ready for analysis within the proposed methods more likely, the analysis will begin with using the Feed Forward Neural Network and then Long Short Time Memory Neural Network. In order to derive the performance between the both proposed methods, some metrics are defined as hereinafter:

1. Mean Square Error: which can be determined from the testing outcomes and ideal outcomes i.e. the target (cluster head). As an example of the mean square error is the process of predicting the next numbers in the sequence X= [1, 3, 5, 7]. The perfect (target) of this predictor must be Y= [2, 4, 6, 8], so if the predictor resulted the Y as in Y= [2, 5, 6, 2]. The predictor outcome is showing that two elements are mistakenly predicted, so the mean square error can be calculated from Eq.(3.4) as follow:

$$MSE = \frac{1}{M}\sum(Y_a - Y_b) \tag{3.4}$$

Where:

M: Total Error number

Y_a: Is the ideal output vector (target)

Y_b: Is the actual output vector resulted from the predictor.

The MSE is important metric to evaluate the performance of the training in every classifier.

2. Time: measured in seconds and stands for the time taken by the classifier to reach particular approximation.
3. Epochs: number of epochs that are taken by the classifier to reach the particular results.

The mentioned metrics are applied in all the experiments that are conducted to evaluate the performance of that experiment and to compare the outcomes with the previous outcomes in order to develop (reach) some settings of classifier with minimum cost (losses).

4. RESULTS AND DISCUSSION

4.1 FIRST EXPERIMENT

The feed forward neural network is used in the settings mentioned in the previous section. The settings can be rewritten as in Table 4.1. the maximum number of epochs allotted for this predictor is 100. The predictor is seen reaching a good performance at single epoch only.

Table 4.1: Parameters of first predictor based FFNN.

Experiment	**1**
Layers	2
Nodes	200,1
Iterations	100
Best MSE	0.09659583
Epochs	1
Weight and bias	Pseudo random

Table 4.1 shows that Feed Forward Neural Network with two layers, the first layer is contained 200 nodes while the output layer is contained of single node. The experiment is repeated for one hundred times since the weight randomly generated, the results keen fluctuation in every run of the program. To tackle this issue, hundred experiments (runs) are made in order to find the best cost. The results shown that minimum mean square error is equal to (0.09659583) which resulted from single epoch. Figure 4.1 depicts the mean square error in all iterations.

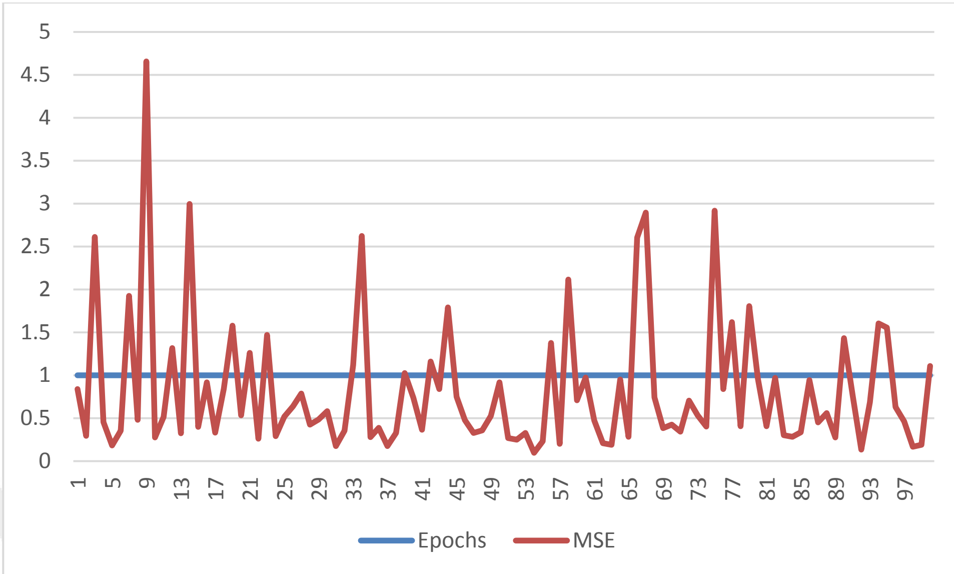

Figure 4.1: Mean square error versus iteration number while FFNN is used.

4.2 SECOND EXPERIMENT

With effort to deploy deep learning technology for predicting the occurrence of thyroid disease, a Long Short Time Memory neural network is implemented with various forms in order to optimize the results obtained from the Experiment I. the first settings of LSTM predictor is listed in the below Tables. It is important to note that the maximum epochs of this experiment are equal to ten.

4.2.1 First Modification

Experiment conducted with maximum number of epochs equal to ten epochs and hence four iterations (runs) are made in order to tackle the random impact of weight/bias generation. Table shows the best mean square error is equal to 0.06271883 at nine epochs. Experiment is made with LSTM neural network at two layers with 100 nodes at the input and single node at the output.

Table 4.2: Parameters of predictor-based LSTM in second experiment (1st modification).

Experiment	2A			
layers	2			
Nodes	100,1			
Iterations	4			
MES	0.07279769	0.10102842	0.06271883	0.09833466
Epochs	8	7	9	3

The result of this experiment can be demonstrated in Figures 4.2 and 4.3.

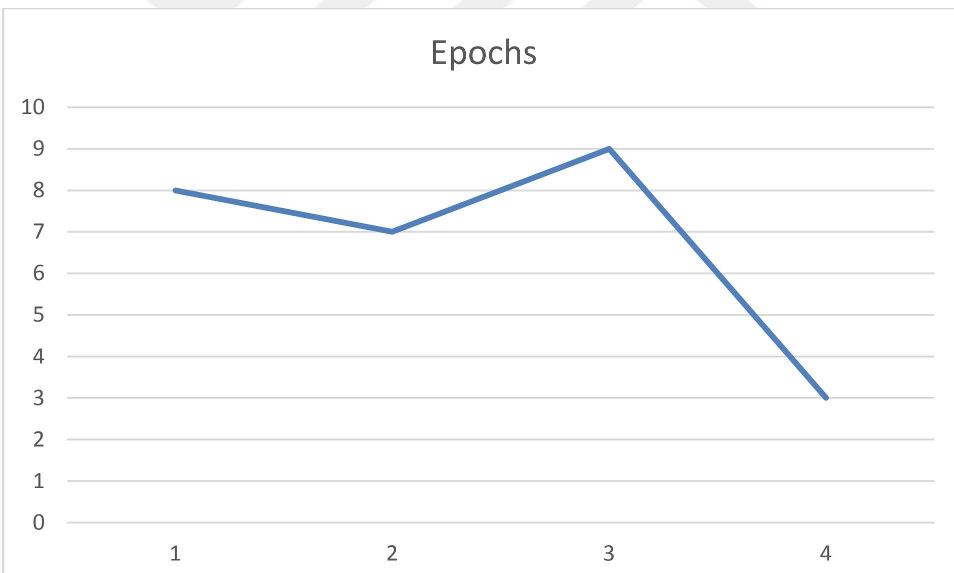

Figure 4.2: Number of epochs versus the iteration number for fist modification in LSTM predictor.

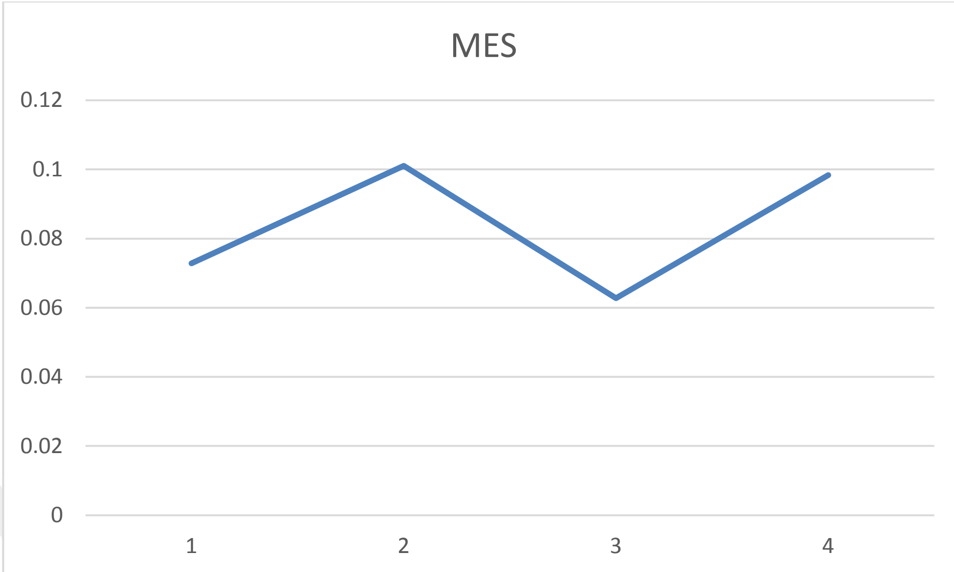

Figure 4.3: MSE versus the iteration number for fist modification in LSTM predictor.

4.2.2 Second Modification

LSTM neural network is used again with some modifications of the parameters with efforts to optimize the performance. The second settings of LSTM predictor is tabulated in Table 4.3. experiment conducted with maximum number of epochs equal to ten epochs and hence four iterations (runs) are made in order to tackle the random impact of weight/bias generation. Table shows the best mean square error is equal to 0.05540274 at seven epochs. Experiment is made with LSTM neural network at two layers with 200 nodes at the input and single node at the output.

Table 4.3: Parameters of predictor-based LSTM in second experiment (2nd modification).

Experiment	2B			
Layers	2			
Nodes	200,1			
Iterations	4			
MES	0.06856	0.05540274	0.09338043	0.05881513
Epochs	8	7	9	6

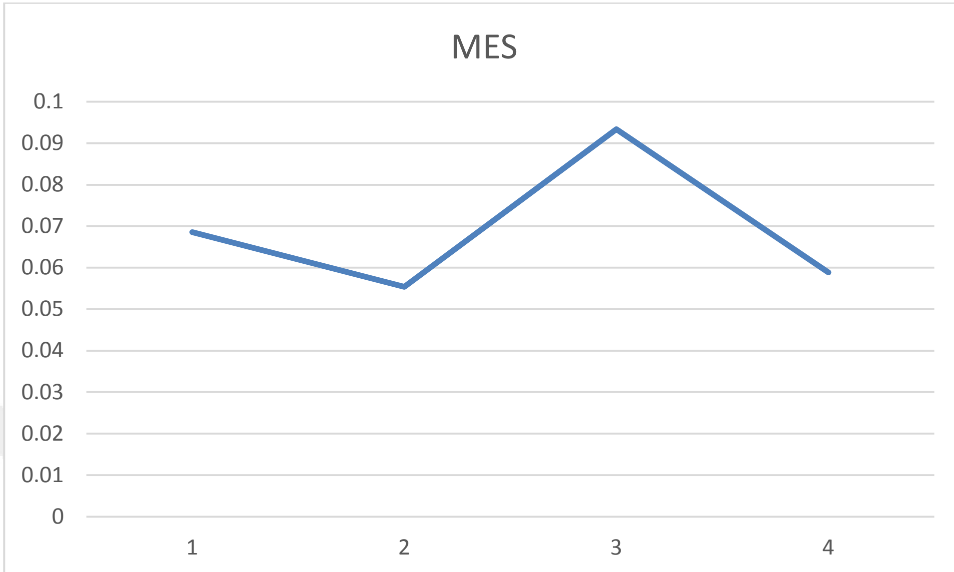

Figure 4.4: MSE versus the iteration number for 2nd modification in LSTM predictor.

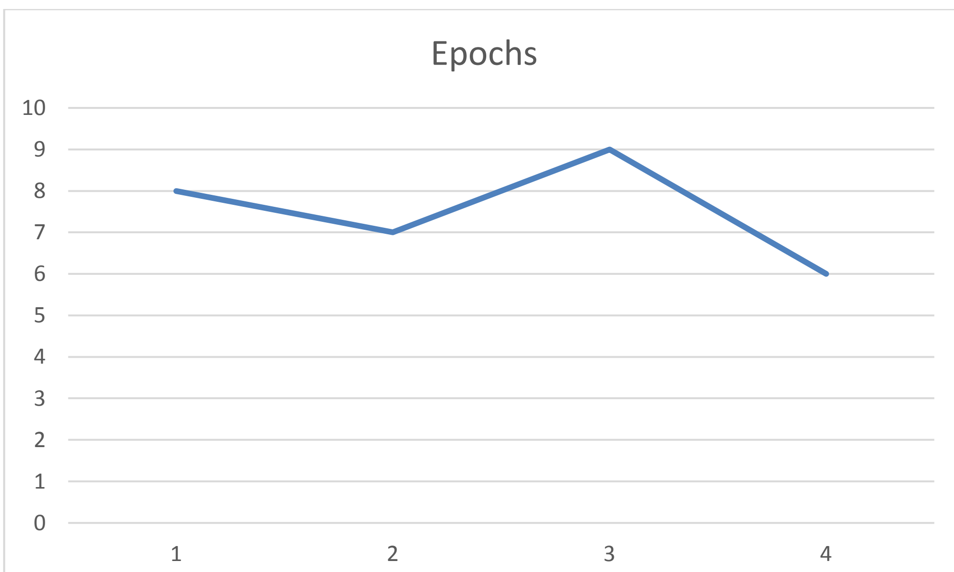

Figure 4.5: Number of epochs versus the iteration number for 2nd modification in LSTM predictor.

4.2.3 Third Modification

In this modification we increased the number of layers in the LSTM predictor wherein three layers are made with 100 nodes at the input layer, 100 nodes at the hidden layer and single node at the output layer. Table 4.4 represent the experiment details

Table 4.4: Parameters of predictor-based LSTM in second experiment (3rd modification).

Experiment	2C			
Layers	3			
Nodes	100,100,1			
Iterations	4			
MES	0.06255	0.05466761	0.06079289	0.05641229
Epochs	4	8	8	7

The results revealed a best MSE is 0.05466761at eight epochs, furthermore, results can be demonstrated at following Figures.

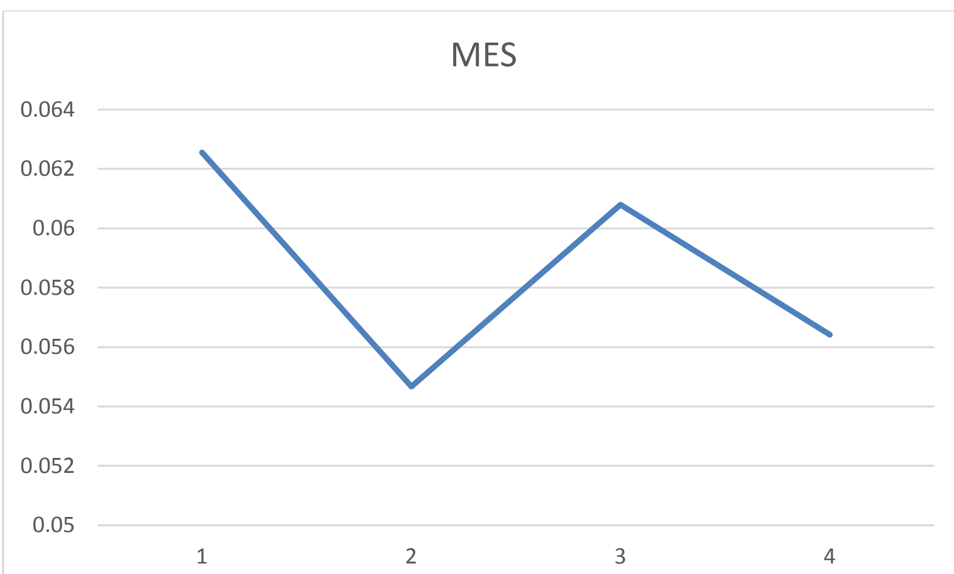

Figure 4.6: MSE versus the iteration number for 3rd modification in LSTM predictor.

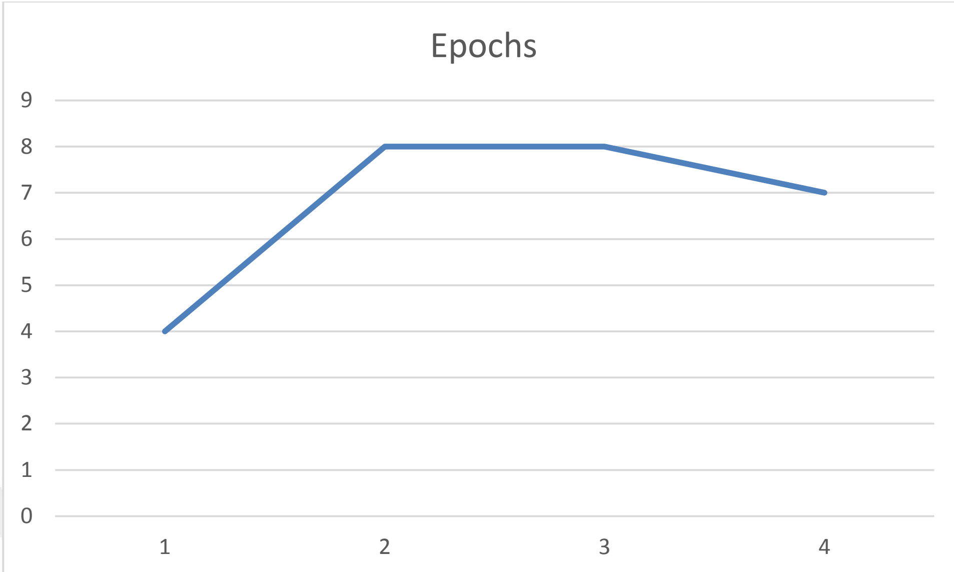

Figure 4.7: Number of epochs versus the iteration number for 3rd modification in LSTM predictor.

4.2.4 Fourth Modification

Four layers LSTM neural network is made according to the details of Table 4.5. two hidden layers are made instead of one hidden layer all layers with 100 nodes except the output layer which hold a single node.

Table 4.5: Parameters of predictor-based LSTM in second experiment (4th modification).

Experiment	2D			
Layers	4			
Nodes	100,100,100,1			
iterations	4			
MES	0.05582	0.05532829	0.05525305	0.05914223
Epochs	8	7	7	8

Experiment revealed that best cost is 0.05525305 at seven epochs, the results are graphically depicted in Figures below.

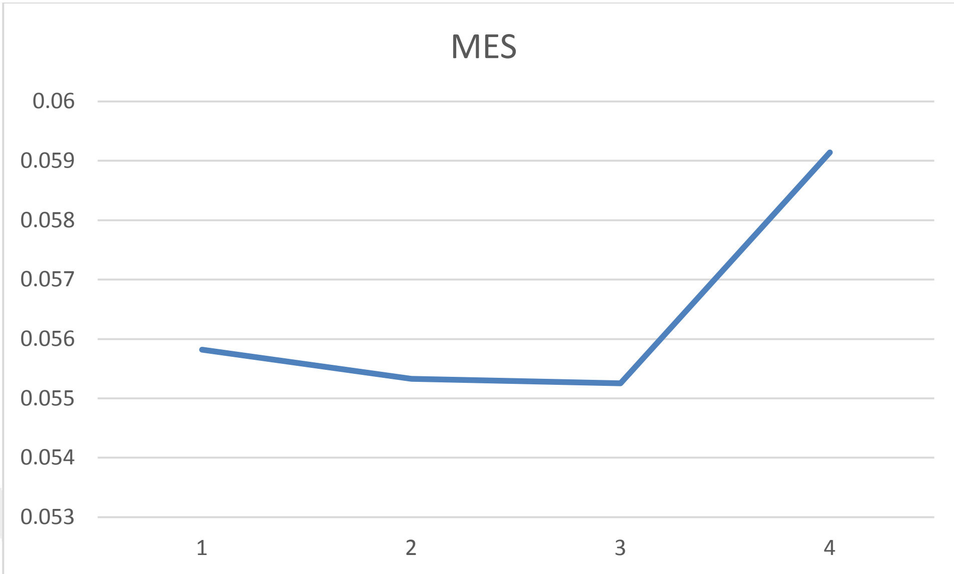

Figure 4.8: MSE versus the iteration number for 4th modification in LSTM predictor.

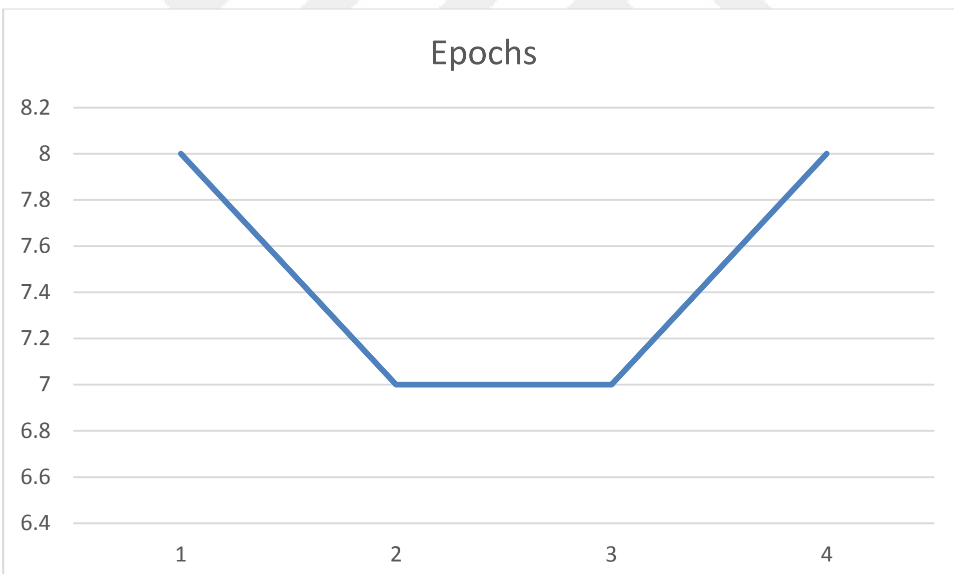

Figure 4.9: Number of epochs versus the iteration number for 4th modification in LSTM predictor.

4.2.5 Fifth Modification

Experiment is repeated by increasing the number of hidden layers so three hidden layers are made with further settings as listed in Table below:

Table 4.6: Parameters of predictor-based LSTM in second experiment (5th modification).

Experiment	2E			
Layers	5			
Nodes	100,100,100,100,1			
iterations	4			
MES	0.05493	0.05580519	0.055073	0.05590123
Epochs	7	7	3	7

The best cost obtained from this experiment is 0.05493001 at seven epochs. Further results are depicted in below Figures.

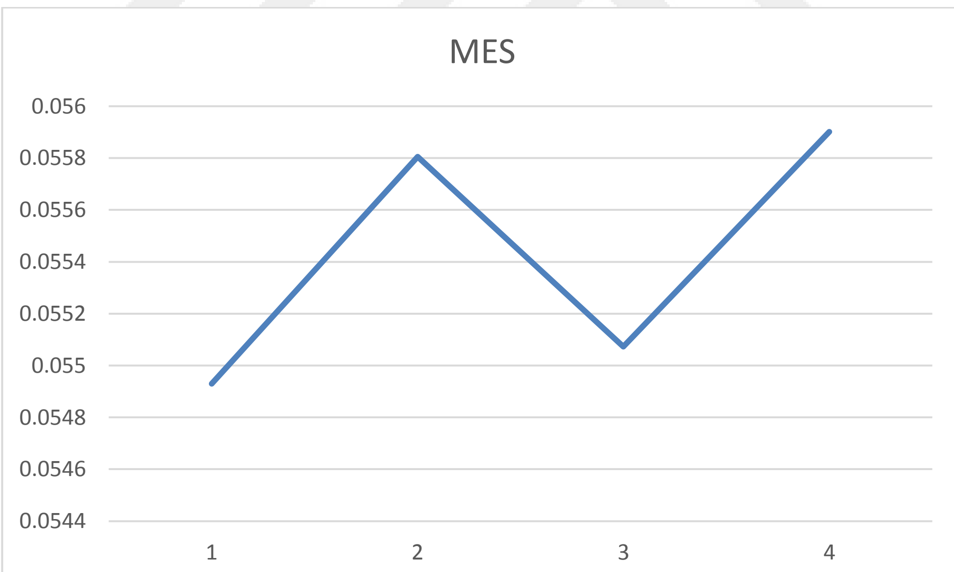

Figure 4.10: MSE versus the iteration number for 5th modification in LSTM predictor.

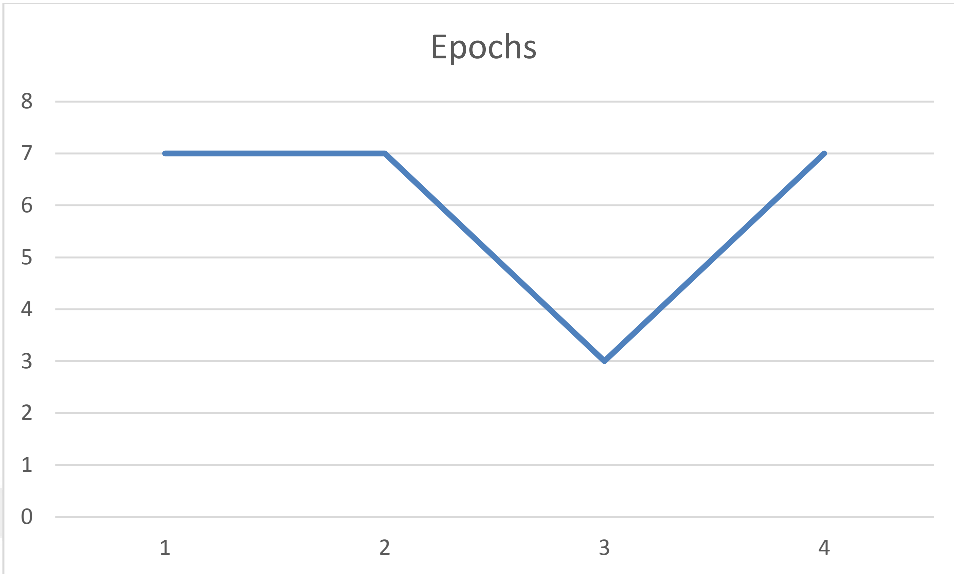

Figure 4.11: Number of epochs versus the iteration number for 5th modification in LSTM predictor.

4.3 THIRD EXPERIMENT

in this experiment we increase the number of epochs to twenty epochs (maximum epoch) and same modifications of the first experiment are repeated. So, tables of first experiment can be referred and results are listed hereinafter.

4.3.1 First Modification Results

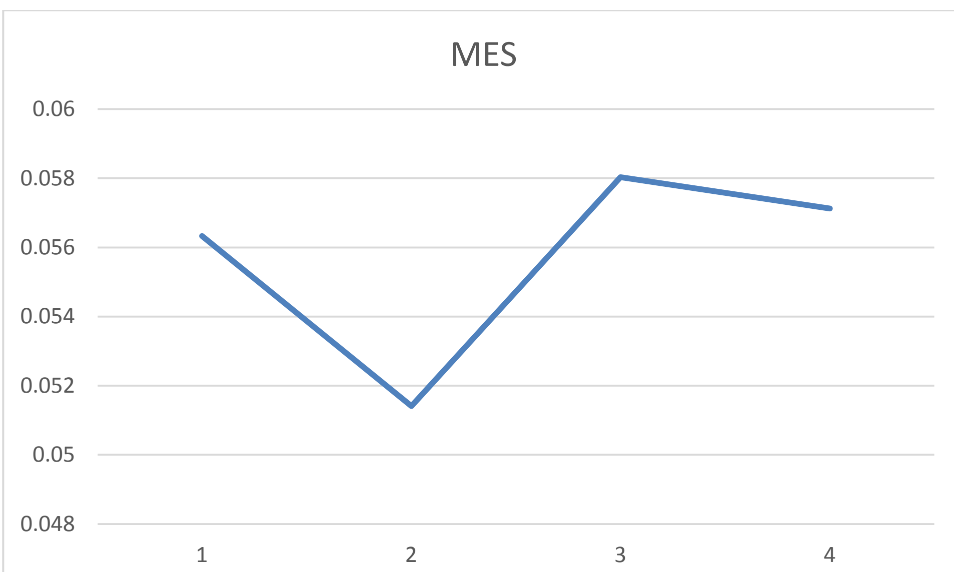

Figure 4.12: MSE versus the iteration number for 1st modification (third experiment) in LSTM predictor.

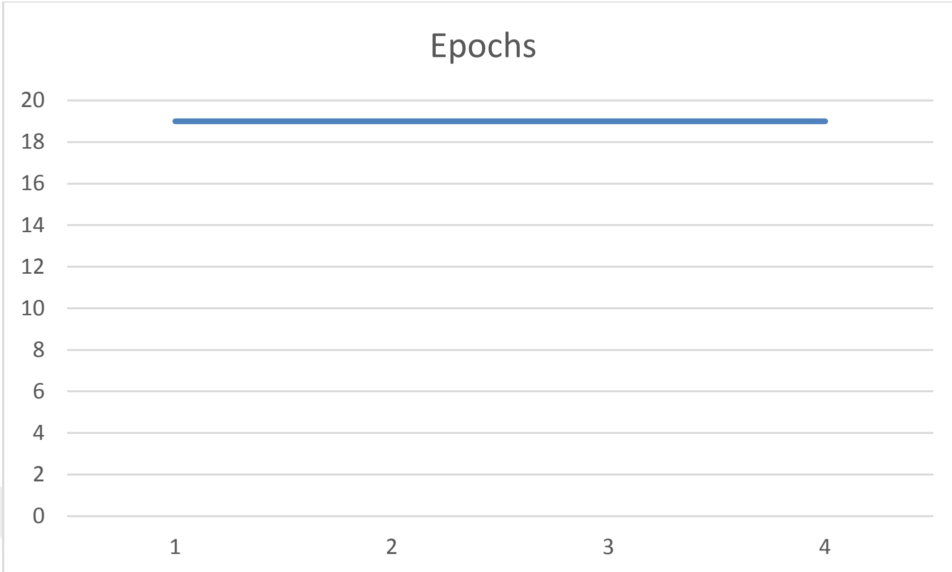

Figure 4.13: Number of epochs versus the iteration number for 1st modification (third experiment) in LSTM predictor.

Its cleared from the Figures above that best cost is 0.05140991 and the epoch number has remained same in all iterations (nineteen epochs).

4.3.2 Second Modification

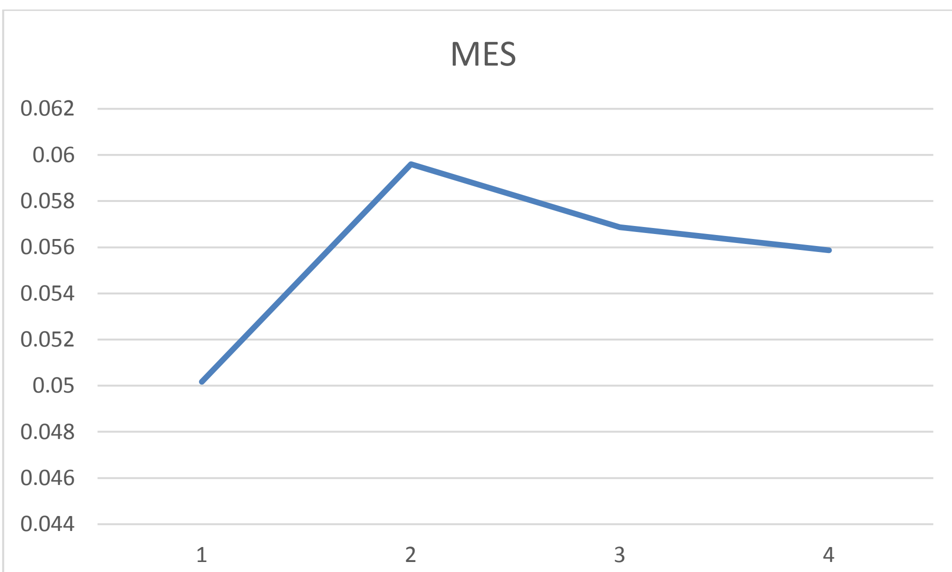

Figure 4.14: MSE versus the iteration number for 2nd modification (third experiment) in LSTM predictor.

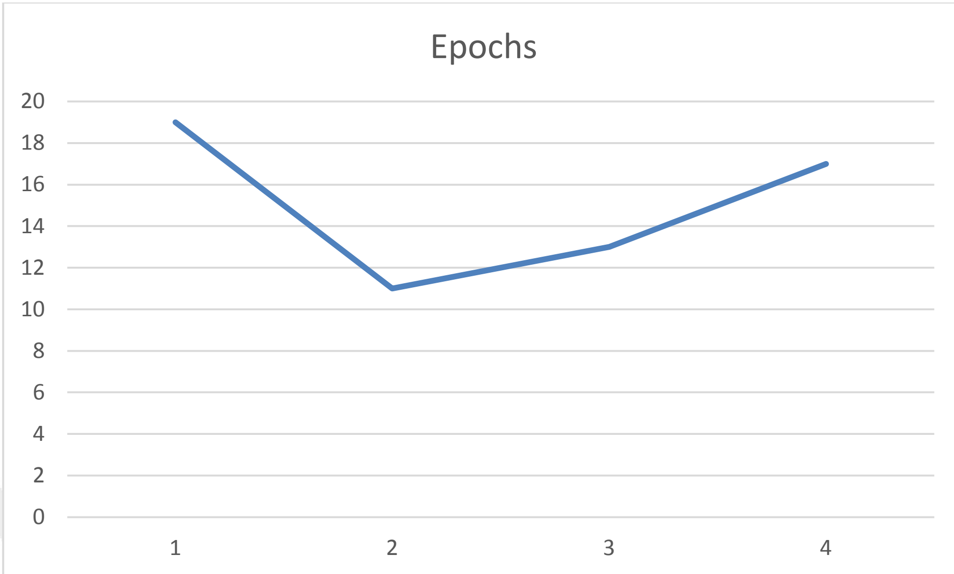

Figure 4.15: Number of epochs versus the iteration number for 2nd modification (third experiment) in LSTM predictor.

This experiment shows that best cost is 0.05017238 at nineteen epochs

4.3.3 Third Modification

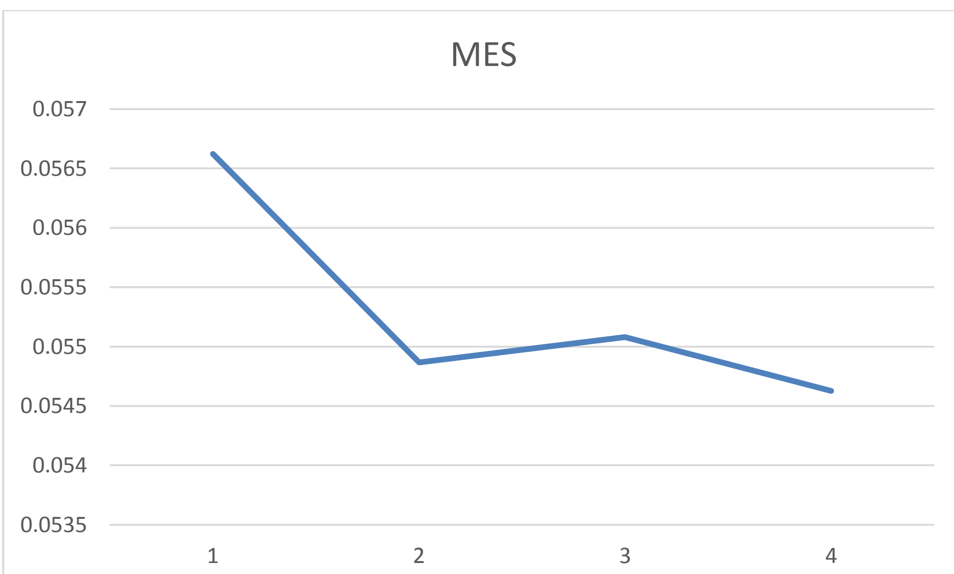

Figure 4.16: MSE versus the iteration number for 3rd modification (third experiment) in LSTM predictor.

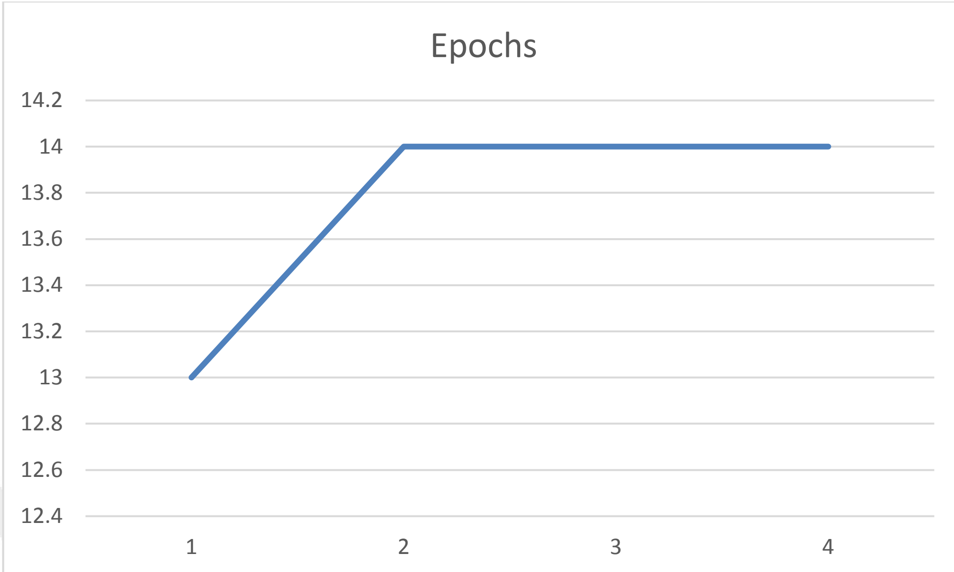

Figure 4.17: Number of epochs versus the iteration number for 3rd modification (third experiment) in LSTM predictor.

The best cost of this experiment is found 0.0546267, at fourteen epochs.

4.3.4 Fourth Modification

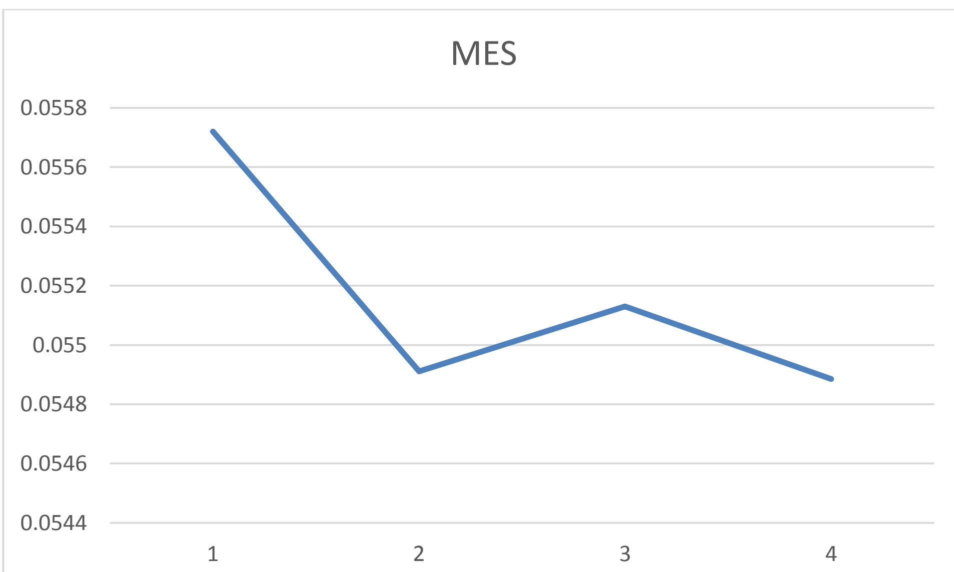

Figure 4.18: MSE versus the iteration number for 4th modification (third experiment) in LSTM predictor.

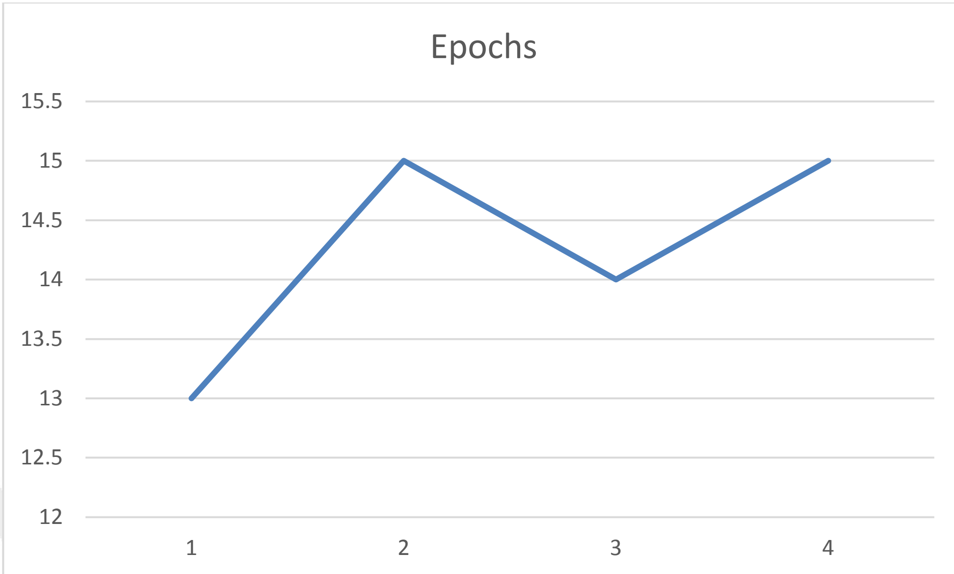

Figure 4.19: Number of epochs versus the iteration number for 4th modification (third experiment) in LSTM predictor.

The best cost yielded from this experiment is 0.05488537 at fifteen epochs.

4.3.5 Fifth Modification

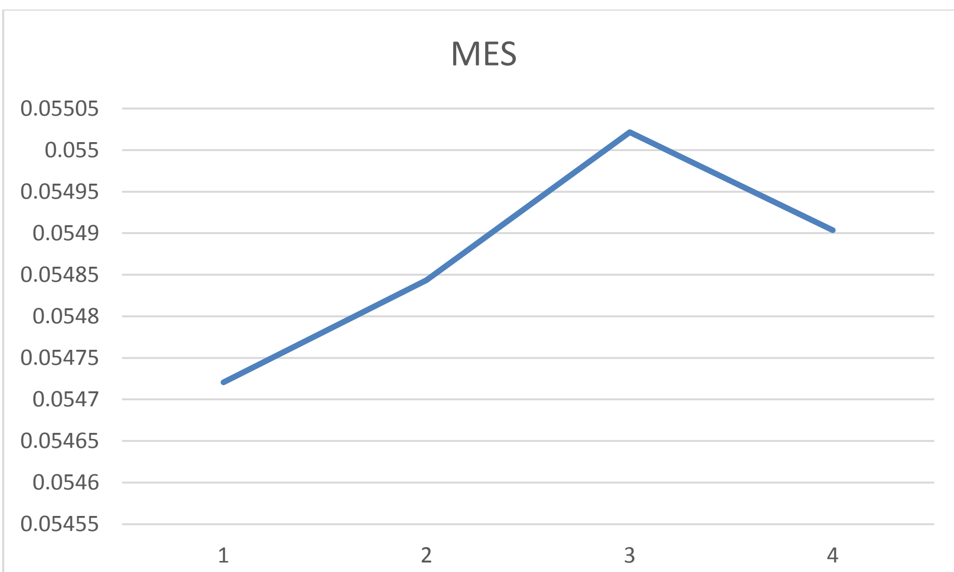

Figure 4.20: MSE versus the iteration number for 5th modification (third experiment) in LSTM predictor.

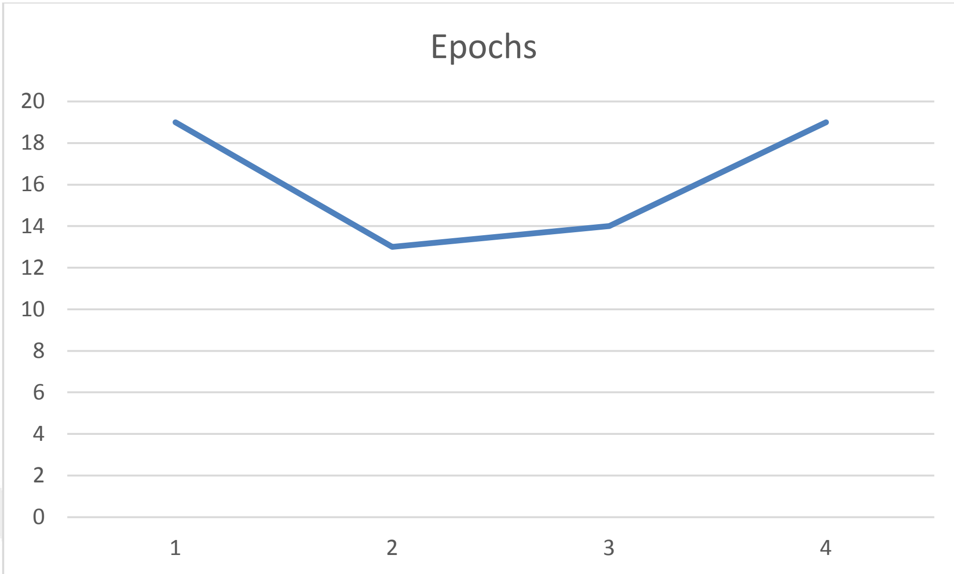

Figure 4.21: Number of epochs versus the iteration number for 5th modification (third experiment) in LSTM predictor.

The best cost yielded from this experiment is 0.05472066 at nineteen epochs.

4.4 FOURTH EXPERIMENT

At this time, one modification is done by increasing the number rod epochs to hundred epochs (maximum epochs). The previous experiment shown that two-layer LSTM neural network has given the best performance so that we modify the settings of two layers LSTM neural network of 200 nodes at the input layer and single node at the output layer with hundred epochs as maximum limit so that we got the result as shown in Table 4.7.

Table 4.7: Parameters of predictor-based LSTM in fourth experiment.

Experiment	4			
Layers	2			
Nodes	200,1			
iterations	4			
MES	0.03917	0.04492739	0.03976635	0.04478583
Epochs	98	99	98	99

The best cost obtained from this experiment is 0.03916952 at ninety-eight epochs. Further, the outcomes of this experiment can be graphically demonstrated as below:

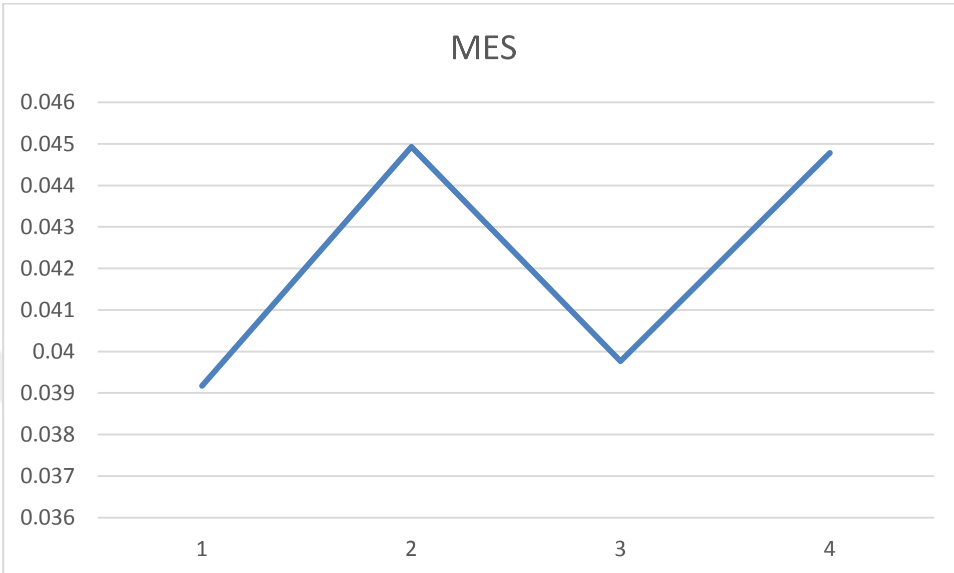

Figure 4.22: MSE versus the iteration number for (fourth experiment) in LSTM predictor.

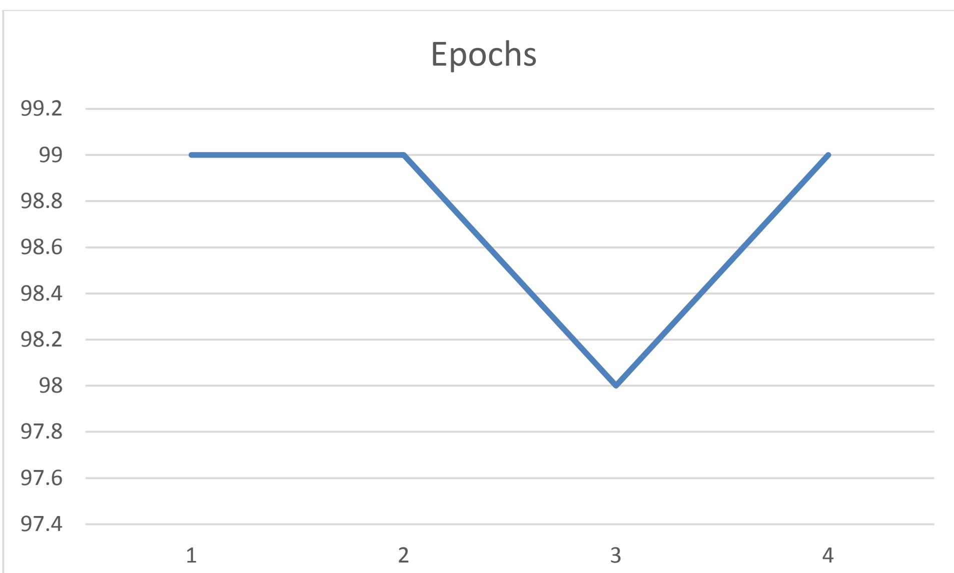

Figure 4.23: Number of epochs versus the iteration number for (fourth experiment) in LSTM predictor.

Using of crowding distance which extracted from NSGA-II optimization technique showed more improvement in system prediction with these parameters T3, TT4 and TSH. For more accurate it has been compared to five attributes evaluator techniques which used ranker searching method and then applying LSTM model as a better performances model for MSE calculations. Table 4.8. shows the different used attributer evaluator with its MSE performances after applying LSTM for specifying the risk factors attributes according to the sequences of importance by deleting last two evaluated column from data set.

Table 4.8: Risk Factor Specification for different Attributes Evaluator with LSTM .

Methods	Attributes Sequences					MSE
Crowding Distance	T3	TT4	TSH	FTI	T4U	**0.040785**
Information Gain	FTI	TT4	T3	TSH	T4U	0.047038
Gain Ratio	FTI	T3	TT4	TSH	T4U	0.0444952
Correlation	FTI	T3	TT4	TSH	T4U	0.04650
Relief	FTI	TT4	T3	T4U	TSH	0.04759
Symmetrical Uncertainty	FTI	T3	TT4	TSH	T4U	0.046353

5. CONCLUSION

Thyroid dataset is prepared for prediction of thyroid disease, data is about 2800 subjects' thyroid medical investigations such as hormones test and some other queries such as the history of other diseases and medication. However, the prediction begins with Feed Forward Neural Network (FFNN) with two layers (200 nodes at the input and single node at the output). Results made by observing the learning performance of the predictor more likely the epochs number and the mean square error (MSE). The result of FFNN is shown that best MSE that can be delivered from this predictor is (0.09659583) at one epoch only. Experiments are proceeded to use deep earning technology to predict of thyroid disease. The Long Short Time Memory Neural Network is used with different modifications. The first modification made with ten maximum epochs and various number of layers, the layers number is beginning with two through five layers and nodes were allotted to each layer were 100 and 200 nodes except the output layer which were given a single node. The results of first modification was 0.05540274 as minimum MSE at seven epochs. The second modification was made by increasing the maximum number of epochs to be twenty epochs with similar settings at the previous modification to get 0.05017238 MSE at 19 epochs. The last modification was made by maximizing the epochs to become 100 epochs with two layers consisting of two hundred nodes at the input layer and single node at the output layer so to say, the results obtained from last modification were 0.03916952 MSE at 98 epochs. From that we can conclude, LSTM neural network was drawn the optimum results with minimum MSE (0.03916952) at 98 epochs. The LSTM is seen utilizing the almost full number of epochs to reach the best approximation of results whereas the FFNN is seen utilizing a lesser epochs to reach some approximation, the reason behind this is the effect of memory (feedback) in LSTM predictor.

Moreover, risk attributes ought to be determined according to popular attributes evaluator techniques and search method which compared to proposed crowding distance technique. Omitting the last two column from data set and evaluates LSTM performances showed that T3, TT4 and TSH are the main effective parameters which has to be available for accurate prediction. In addition, T3 and TT4 are agreed for all searching method to be obligate for this type of data set.

REFERENCES

[1] Y. Wang, J. Zhou, K. Chen, Y. Wang and L. Liu, "Water quality prediction method based on LSTM neural network," *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, Nanjing, 2017, pp. 1-5.

[2] A. H. Mirza and S. Cosan, "Computer network intrusion detection using sequential LSTM Neural Networks autoencoders," *2018 26th Signal Processing and Communications Applications Conference (SIU)*, Izmir, 2018, pp. 1-4.

[3] Siyuan Chen, “A Deep Neural Network Model for Target-based Sentiment Analysis”, 978-1-5090-6014-6/18/$31.00 ©2018 IEEE.

[4] V Jithesh, “LSTM Recurrent Neural Networks for High Resolution Range Profile Based Radar Target Classification”, 978-1-5090-6218-8/17/$31.00 ©2017 IEEE.

[5] Gu Yuhai1, Liu Shuo 2, He Linfeng 3,Wang liyong4, “Research on Failure Prediction Using DBN and LSTM Neural Network”, 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). September 11-14, 2018, Nara, Japan.

[6] Martin Sundermeyer, Hermann Ney, Fellow, IEEE, and Ralf Schlüter, Member, IEEE, “From Feedforward to Recurrent LSTM Neural Networks for Language Modeling”, IEEE/Acm Transactions On Audio, Speech, And Language Processing, vol. 23, no. 3, March 2015.

[7] Dires Negash Fente, “Weather Forecasting Using Artificial Neural Network”, Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2.

[8] Yaohua Liu1, Yimin Zhou, Xiang Li, “Attitude Estimation of Unmanned Aerial Vehicle Based on LSTM Neural Network”, 978-1-5090-6014-6/18/$31.00 ©2018 IEEE.

[9] N Maria Klara Je ̨drzejewska, “Generating musical expression of MIDI music with LSTM neural network”, 978-1-5386-5024-0/18/$31.00 ©2018 IEEE.

[10] Yuzhen Lu and Fathi M. Salem, “Simplified Gating in Long Short-term Memory (LSTM) Recurrent Neural Networks”, 978-1-5090-6389-5/17/$31.00 ©2017 IEEE.

[11] Zaifa Chen, Yancheng Liu, “Mechanical State Prediction Based on LSTM Neural Network”, Proceedings of the 36th Chinese Control Conference July 26-28, 2017, Dalian, China.

[12] N B. Chandra, Rajesh Kumar Sharma, “On improving Recurrent Neural Network for Image Classification”, 978-1-5090-6182-2/17/$31.00 ©2017 IEEE.

[13] Xu Xu, Hao Ge, “An implementation of recurrent neural network by combining a convolutional neural network and simple initialization of weights”, 978-1-4673-7755-3/16/$31.00 ©2016 IEEE.

[14] Kewei Chen, Leilei Huang, Minjiang Li, Xiaoyang Zeng, Yibo Fan, “A compact and configurable long short-term memory neural network hardware architecture”, 978-1-4799-7061-2/18/$31.00 ©2018 IEEE.

[15] J. Li, L. Ding, and B. Li, “Differential evolution-based parameters optimization and feature selection for support vector machine,” Int. J. Comput. Sci. Eng., vol. 13, no. 4, pp. 355–363, 2016.

[16] R. Devi Priya and R. Sivaraj, “Pre-processing of microarray gene expression data for classification using adaptive feature selection and imputation of non-ignorable missing values,” Int. J. Data Min. Bioinform., vol. 16, no. 3, pp. 183–204, 2016.

[17] K. S. Vairavel and S. Valarmathy, “Implementing feature selection for multimodal biometrics,” Int. J. Biomed. Eng. Technol., vol. 23, no. 2–4, pp. 242–260, 2017.

[18] F. Zheng, X. Shen, Z. Fu, S. Zheng, and G. Li, “Feature selection for genomic data sets through feature clustering,” Int. J. Data Min. Bioinform., vol. 4, no. 2, pp. 228–240, 2010.

[19] H. F. Eid, “Binary Whale Optimization: An Effective Swarm Algorithm for Feature Selection,” Int. J. Metaheuristics, vol. 7, no. 1, pp. 67–79, 2018.

[20] Y. Praharsi, S.-G. Miaou, and H.-M. Wee, “Supervised learning approaches and feature selection - a case study in diabetes,” Int. J. Data Anal. Tech. Strateg., vol. 5, no. 3, pp. 323–337, 2013.

[21] G.-Z. Li, J. Y. Yang, W.-C. Lu, D. Li, and M. Q. Yang, “Improving prediction accuracy of drug activities by utilising unlabelled instances with feature selection,” Int. J. Comput. Biol. Drug Des., vol. 1, no. 1, pp. 1–13, 2008.

[22] R. Dash and B. B. Misra, "A multi-objective feature selection and classifier ensemble technique for microarray data analysis," Int. J. Data Min. Bioinform., vol. 20, no. 2, pp. 123–160, 2018.

[23] S. Hari, M. N. Murty, and S. Ramakrishnan, "Effective feature selection technique for text classification," Int. J. Data Min. Model. Manag., vol. 7, no. 3, pp. 165–184, 2015.

[24] O. A. Alomari, A. T. Khader, M. A. Al-Betar, and L. M. Abualigah, "Gene selection for cancer classification by combining minimum redundancy maximum relevancy and bat-inspired algorithm," Int. J. Data Min. Bioinform., vol. 19, no. 1, pp. 32–51, 2017.

[25] R. Anand, A. Vaid, and P. K. Singh, "Association Rule Mining Using Multi-objective Evolutionary Algorithms : Strengths and Challenges," 2009 World Congr. Nat. Biol. Inspired Comput., pp. 385–390, 2009.

[26] R. Acharya, S. Vipsita, and S. K. Baliarsingh, "Biclustering of Microarray Data Employing Multiobjective GA," 2017 14th IEEE India Counc. Int. Conf., vol. 15–17 Dec., pp. 1–6, 2017.

[27] E. D. Durmaz and R. Şahin, "NSGA-II and goal programming approach for the multi-objective single row facility layout problem," J. Fac. Eng. Archit. Gazi Univ., vol. 3, pp. 941–955, 2017.

[28] M. S. Mohammed and R. A. Vural, "NSGA-II+FEM Based Loss Optimization of Three Phase Transformer," IEEE Trans. Ind. Electron., pp. 1–1, 2018.

[29] N. E. Golea, K. E. Melkemi, and M. Melkemi, "A Novel Multi-objective Genetic Algorithm Optimization for Blind RGB Color Image Watermarking," Seventh Int. Conf. Signal Image Technol. Internet-Based Syst. Fr., p. 306–313., 2011.

[30] M. H. Zangooei, J. Habibi, and R. Alizadehsani, "Disease Diagnosis with a hybrid method SVR using NSGA-II," Neurocomputing, vol. 136, pp. 14–29, 2014.

[31] Cleveland database:, "No Title." [Online]. Available: https://archive.ics.uci.edu/ml/machine-learning-databases/thyroid-disease.

[32] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm :," 182 IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182–197, 2002.

[33] T. Fawcett and F. Provost, "Activity monitoring: Noticing interesting changes in behavior," in Proceedings of the fifth ACM SIGKDD interna- tional conference on Knowledge discovery and data mining, pp. 53–62, ACM, 1999.

[34] M. Markou and S. Singh, “Novelty detection: a reviewpart 1: statistical approaches,” Signal processing, vol. 83, no. 12, pp. 2481–2497, 2003.

[35] J. P. Anderson, “Computer security threat monitoring and surveillance,” Technical Report, James P. Anderson Company, 1980.

[36] D. E. Denning, “An intrusion-detection model,” IEEE Transactions on software engineering, no. 2, pp. 222–232, 1987.

[37] S. J. Stolfo, S. M. Bellovin, S. Hershkop, A. D. Keromytis, S. Sinclair, and S. W. Smith, Insider attack and cyber security: beyond the hacker, vol. 39. Springer Science & Business Media, 2008.

[38] C. Shields, “Machine learning and data mining for computer security, chapter an introduction to information assurance,” Springer, 2005.
K. Leung and C. Leckie, “Unsupervised anomaly detection in network

[39] intrusion detection using clusters,” in Proceedings of the Twenty-eighth Australasian conference on Computer Science-Volume 38, pp. 333–342, Australian Computer Society, Inc., 2005.

[40] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,” Neurocomputing, vol. 234, pp. 11–26, 2017.

[41] K. Greff, R. K. Srivastava, J. Koutn ́ık, B. R. Steunebrink, and J. Schmid- huber, “Lstm: A search space odyssey,” IEEE transactions on neural networks and learning systems, vol. 28, no. 10, pp. 2222–2232, 2017.

[42] L. Medsker and L. Jain, “Recurrent neural networks,” Design and Applications, vol. 5, 2001.

[43] M. Sakurada and T. Yairi, “Anomaly detection using auto encoders with nonlinear dimensionality reduction,” in Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis, p. 4, ACM, 2014.

[44] V. Kustikova and P. Druzhkov, “A survey of deep learning methods and software for image classification and object detection,” OGRW2014, vol. 5, 2014.