A Generalized Load Forecasting Methodology for Electric Power Systems

Submitted by
Shahid Mahmood Awan
2008-PhD-CS-04

Supervised by Dr. Muhammad Aslam

Department of Computer Science and Engineering
University of Engineering and Technology, Lahore
بسم الله الرحمن الرحيم
A Generalized Load Forecasting Methodology for Electric Power Systems

Dissertation submitted to the faculty of the Electrical Engineering,
Department of Computer Science and Engineering
of the University of Engineering and Technology, Lahore
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy
in
Computer Science.
by
Shahid Mahmood Awan (2008-PhD-CS-04)

Internal Examiner
Dr. Muhammad Aslam
Associate Professor
Department of Computer Science
and Engineering

External Examiner
Dr. Zia ul Qayyum
Professor / Vice Chancellor
University of Gujrat

Chairman
Dean
Department of Computer Science
Faculty of Electrical Engineering
and Engineering

University of Engineering and Technology, Lahore
Declaration

I declare that the work contained in this thesis is my own, except where explicitly stated otherwise. In addition this work has not been submitted to obtain another degree or professional qualification.

Signed: ___________________
Date: ___________________
Acknowledgments

Countless thanks to ALLAH (S.W.T), The Beneficent, The merciful for it is under His grace that we live, learn, prosper and flourish. Blessings and salutation on PROPHET MUHAMMAD (S.A.W) who is forever a center of knowledge and tower of guidance to humanity. The completion of this thesis would have been impossible without the extensive encouragement and guidance I received from my mentors, colleagues, and my family. I would like to specifically express my sincere gratitude to Dr. Muhammad Aslam and Prof. Dr. Zubair A. Khan. Dr. Khan offered profound technical insights from the energy forecasting domain. Dr. Aslam patiently provided me with countless hours of advice, insights, guidance, and encouragement throughout the course of this work. Their extensive support and mentoring made this work possible. I am thankful to Dr. Ana Maria Martinez Enriquez, Research Investigator, CINVESTA-IPN, D.F. Mexico., Prof. Dr. Wei-Chiang Hong, Management School, Hangzhou Dianzi University, P. R. China., and Dr. Muzammil Baig, System Analyst, InterSect Alliance, Australia, for their valuable feedback and suggestions to improve the quality of this thesis and model quality assessment and verification. I would like to express my appreciation to Prof. Dr. Waqar Mahmood and Prof. Dr. Haroon A. Babari for volunteering to offer their time, effort, and knowledge. Their valuable insights helped in model development and data analysis. I convey my genuine gratitude to Prof. Dr. Muhammad Ali Maud for his support and help towards completion of my thesis and administrative support during my studies. I would also like to thank my colleagues for sharing their ideas and offering help and encouragement during my studies. It was a pleasure working with them as student and colleague in attaining common goals.
Dedicated to my parents, brothers, sisters, and my wife. They have always supported and inspired me to achieve my academic goals.
## Contents

Acknowledgments iv

List of Figures x

List of Tables xii

Abbreviations xiii

Basic Definitions xiv

Abstract xv

1 Introduction 1

1.1 Electric Power Systems 1

1.2 Power System of Pakistan 3

1.3 Electric Load Forecasting 5

1.4 Classification of Electric Load Forecasts 7

1.4.1 Temporal Forecasts 7

- Short Term Electric Load Forecasting 7
- Medium Term Electric Load Forecasting 7
- Long Term Electric Load Forecasting 7

1.4.2 Spatial Electric Load Forecasting 8

1.4.3 Consumers based Forecasts 8

1.4.4 Multi-Scenario Load Forecasting 9

1.5 Business Needs of Load Forecasts 9

1.5.1 Energy Purchasing 9

1.5.2 Transmission and Distribution (T&D) Planning 9

1.5.3 Operations and Maintenance 9

1.5.4 Demand Side Management (DSM) 9

1.5.5 Financial Planning 10

1.6 Motivation 10

1.7 Problem Statement 11

1.8 Data Description 13

1.8.1 Electricity Consumption Data 13

1.8.2 Driving Factors of Electricity Consumption 14

1.8.3 Salient Features of Electric Load Series 17

1.9 Research Methodology 19

1.10 Thesis Organization and Conventions 22
# Contents

1.11 Thesis Objectives ........................................ 24  
1.12 Thesis Contributions ..................................... 25  

2 Literature Review ............................................. 26  
2.1 Classification of Electric Load Forecasting Methods ..... 26  
2.2 Traditional Forecasting Techniques ......................... 27  
2.3 Artificial Neural Networks (ANN) ........................... 28  
2.4 Support Vector Machines (SVM) ............................ 28  
2.5 Optimization Techniques .................................. 29  
2.6 Generalization Problem of LF Models ....................... 30  
  2.6.1 Noisy Data ........................................... 30  
  2.6.2 Relationship between Dependent and Independent Variables ........................................ 31  
  2.6.3 Data-Model Coupling .................................. 31  
  2.6.4 Applicability of Existing LF Models .................. 31  
  2.6.5 Over-Fitting .......................................... 31  
  2.6.6 Cross Validation ...................................... 32  
2.7 Conclusion ................................................. 32  

3 Artificial Bee Colony Algorithm based Load Forecasting Model ............................................. 33  
3.1 Introduction ................................................ 33  
3.2 Related Work .............................................. 34  
3.3 ANN Training ............................................... 36  
  3.3.1 ANN model ............................................ 36  
  3.3.2 ABC overview ....................................... 37  
  3.3.3 PSO overview ....................................... 37  
  3.3.4 GA overview ....................................... 37  
3.4 STLF Model Formulation .................................... 37  
  3.4.1 Input Factors ........................................ 37  
  3.4.2 ANN Settings ....................................... 38  
  3.4.3 PSO Settings ....................................... 38  
  3.4.4 GA Settings ....................................... 39  
  3.4.5 ABC Settings ....................................... 39  
  3.4.6 ABC Training Process ................................ 39  
  3.4.7 Stopping Criteria .................................... 41  
3.5 Experimental Results ...................................... 41  
  3.5.1 Data set ............................................. 41  
  3.5.2 Performance Evaluation .............................. 43  
  3.5.3 Forecasting Results and Discussion .................. 44  
3.6 Conclusion ................................................. 48  

4 Artificial Cooperative Search Algorithm based Load Forecasting Model ............................................. 49  
4.1 Introduction .............................................. 49  
4.2 Experimental Setup ...................................... 50  
  4.2.1 IESO Hourly Demand Data ............................ 51  
  4.2.2 Training, Validation & Testing Set .................. 51  
  4.2.3 Algorithm Parameters ................................ 51  
  4.2.4 ANN Architecture .................................... 51
4.2.5 Stopping Criteria ........................................ 51
4.3 ACS based ANN Optimization for LF model Development .... 52
4.4 Results and Discussion ...................................... 55
4.5 Conclusion .................................................. 57

5 Improved Artificial Bee Colony Algorithm based Load Forecasting Model 58
5.1 Introduction ................................................. 58
5.2 Evolving Behavior of ABC Algorithm ....................... 60
5.3 Applied Modifications ...................................... 62
  5.3.1 Modified Initialization ................................ 62
  5.3.2 Modified Search ..................................... 63
  5.3.3 Global Best Search .................................. 63
  5.3.4 Modification in Probability Function ................. 64
  5.3.5 Exploration by Genetic Operators .................... 64
  5.3.6 Exploitation by Genetic Operators ................... 65
  5.3.7 Evaluation on Statistical Benchmark Functions .... 65
5.4 Model Development for Peak Demand Estimation .......... 66
  5.4.1 ANN Training ....................................... 66
  5.4.2 Input Factors ...................................... 67
  5.4.3 ANN Settings ..................................... 67
  5.4.4 ABC Settings ...................................... 68
  5.4.5 GA Settings ...................................... 68
  5.4.6 PSO Settings ...................................... 68
  5.4.7 SA Settings ...................................... 68
5.5 Results .................................................... 69
  5.5.1 Case Study (NTDCL Data) .......................... 69
  5.5.2 Case Study (LESCO Data) .......................... 73
5.6 Conclusion ................................................. 77

6 Performance Evaluation of Proposed ELF Models 78
6.1 Evaluation of ELF models .................................. 78
6.2 Modeling Techniques ...................................... 78
  6.2.1 ANN Model Structure ................................ 79
  6.2.2 SVM Model ........................................ 79
  6.2.3 Optimization Techniques ............................ 79
  6.2.4 Performance Evaluation ............................. 80
6.3 Evaluation on NTDCL Data .................................. 80
  6.3.1 NTDCL Data Set .................................... 80
  6.3.2 Results ........................................... 82
6.4 IESO: Ontario Data Set .................................... 83
  6.4.1 Experiment Setup ................................... 84
  6.4.2 Results ........................................... 84
6.5 Medium Term Load Forecasting: Weekly Peak profile .... 86
  6.5.1 Results ........................................... 86
6.6 Long Term Load Forecasting: Industrial Profile .......... 90
  6.6.1 Data Set .......................................... 90
  6.6.2 Results ........................................... 90
6.7 Evaluation of Proposed Algorithms on Classification Problems 94
   6.7.1 Data Description 94
   6.7.2 Algorithm Parameters 95
   6.7.3 ANN Architecture 96
   6.7.4 Stopping Criteria 96
   6.7.5 Results and Discussion 96
6.8 Conclusion 98

7 A Generalized Load Forecasting Methodology 99
   7.1 Introduction 99
   7.2 Components of LF Model 100
      7.2.1 Data Pre-processing and Analysis 100
      7.2.2 Outlier Detection 101
      7.2.3 Treating Missing Values 101
      7.2.4 Duplicate Entries 101
      7.2.5 Correlation Matrix 102
      7.2.6 Over-Fitting 102
      7.2.7 Cross Validation 103
      7.2.8 Performance Evaluation 103
   7.3 Application on Noisy Data Set 103
      7.3.1 Load Data Characteristics 104
      7.3.2 Model Parameters 104
   7.4 Experimentation 106
      7.4.1 Hourly Demand Estimation 106
      7.4.2 Daily Peak Demand Estimation 106
   7.5 Results and Discussion 107
   7.6 Conclusion 114

8 Conclusion and Future Work 115

References 118
### List of Figures

1.1 Components of Electric Power System ........................................ 2  
1.2 Requirements of Each Component in Electric Power System .......... 2  
1.3 Break up of Pakistan Power System [1]................................. 4  
1.4 Electric Demand Forecasting .................................................. 6  
1.5 Load Forecasting Types .......................................................... 7  
1.6 Sector wise Energy Consumption [1] ......................................... 8  
1.7 Forecasting Function ............................................................... 11  
1.8 Domestic Consumer Growth ...................................................... 15  
1.9 Industrial Consumer Growth ..................................................... 15  
1.10 Year Wise Gross Domestic Product ............................................ 15  
1.11 Average Tariff ........................................................................... 16  
1.12 Effect of Temperature on Load ................................................ 17  
1.13 Relationship between Current and Previous Demand ................. 17  
1.14 Comparison of Load Profiles of two Different Week Days .......... 18  
1.15 Demand Shift after One Year ................................................... 18  
1.16 Modeling and Optimization Techniques ..................................... 19  
1.17 Components of ELF Process ..................................................... 20  
1.18 Thesis Layout ............................................................................ 22  
3.1 FFNN Model Training by ABC, PSO, and GA Algorithms ............. 38  
3.2 FFNN Model Training by ABC Algorithm ................................... 43  
3.3 Comparison of MAPE improvement over 500 iterations ................ 45  
3.4 Comparison of MSE obtained by GA, PSO and ABC assisted Networks . 45  
3.5 Actual vs Predicted demand by GA, PSO, and ABC based FFNN Models 46  
4.1 Comparison of MAPE Obtained by all Techniques ....................... 56  
5.1 Modifications in Base ABC Algorithm ....................................... 65  
5.2 ANN Model Training by ABC, PSO, and SA Algorithms .............. 66  
5.3 Actual demand vs Predicted by ABC, PSO, SA and Modified Versions of ABC .............................................................. 70  
5.4 Actual demand vs Predicted by ABC and ABC-GA assisted ANN .... 70  
5.5 Comparison of MAPE obtained by all Models ............................. 71  
5.6 Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC .............................................................. 72  
5.7 Actual demand vs Predicted by ABC and ABC-GA assisted ANN .... 72  
5.8 Comparison of MAPE Obtained by all Models ............................. 73  
5.9 Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC .............................................................. 74
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.10</td>
<td>Actual demand vs Predicted by ABC and ABC-GA assisted ANN</td>
<td>74</td>
</tr>
<tr>
<td>5.11</td>
<td>Comparison of MAPE obtained by all Models</td>
<td>75</td>
</tr>
<tr>
<td>5.12</td>
<td>Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC</td>
<td>75</td>
</tr>
<tr>
<td>5.13</td>
<td>Actual demand vs Predicted by ABC and ABC-GA assisted ANN</td>
<td>76</td>
</tr>
<tr>
<td>5.14</td>
<td>Comparison of MAPE Obtained by all Models</td>
<td>76</td>
</tr>
<tr>
<td>6.1</td>
<td>Components of LF Process</td>
<td>79</td>
</tr>
<tr>
<td>6.2</td>
<td>Electricity Demand profile of 3 consecutive days</td>
<td>81</td>
</tr>
<tr>
<td>6.3</td>
<td>MAPE Comparison of all Models</td>
<td>83</td>
</tr>
<tr>
<td>6.4</td>
<td>Four Day Profile - Actual v/s Forecasted</td>
<td>83</td>
</tr>
<tr>
<td>6.5</td>
<td>Actual v/s Forecasted by ANN-FFB-PSO and SVM-PSO</td>
<td>85</td>
</tr>
<tr>
<td>6.6</td>
<td>Actual v/s Forecasted by ANN-FFB-SA and SVM-SA</td>
<td>85</td>
</tr>
<tr>
<td>6.7</td>
<td>Actual v/s Forecasted by ANN-LMA</td>
<td>87</td>
</tr>
<tr>
<td>6.8</td>
<td>Actual v/s Forecasted by ANN-RPROP</td>
<td>87</td>
</tr>
<tr>
<td>6.9</td>
<td>Actual v/s Forecasted by ANN-SA</td>
<td>88</td>
</tr>
<tr>
<td>6.10</td>
<td>Actual v/s Forecasted by ANN-FFB-SCG</td>
<td>88</td>
</tr>
<tr>
<td>6.11</td>
<td>Actual v/s Forecasted by SVM-GreedySearch</td>
<td>88</td>
</tr>
<tr>
<td>6.12</td>
<td>Actual v/s Forecasted by all techniques</td>
<td>89</td>
</tr>
<tr>
<td>6.13</td>
<td>Actual v/s Forecasted by ANN-RPROP</td>
<td>92</td>
</tr>
<tr>
<td>6.14</td>
<td>Actual v/s Forecasted by ANN-PSO</td>
<td>92</td>
</tr>
<tr>
<td>6.15</td>
<td>Actual v/s Forecasted by ANN-PSO and ANN-RPROP</td>
<td>93</td>
</tr>
<tr>
<td>7.1</td>
<td>Components of LF model</td>
<td>100</td>
</tr>
<tr>
<td>7.2</td>
<td>Data Pre-Processing and Analysis Components of LF model</td>
<td>101</td>
</tr>
<tr>
<td>7.3</td>
<td>Five Day Profile of Load Data with Noise and after Pre-Processing</td>
<td>102</td>
</tr>
<tr>
<td>7.4</td>
<td>Relationship between Temperature and Demand</td>
<td>104</td>
</tr>
<tr>
<td>7.5</td>
<td>Relationship between Current and Previous Demand Profiles</td>
<td>105</td>
</tr>
<tr>
<td>7.6</td>
<td>Actual v/s Forecasted demand by ANN-RPROP</td>
<td>108</td>
</tr>
<tr>
<td>7.7</td>
<td>Actual v/s Forecasted demand by ANN-PSO</td>
<td>108</td>
</tr>
<tr>
<td>7.8</td>
<td>Actual v/s Forecasted demand by ANN-LMA</td>
<td>108</td>
</tr>
<tr>
<td>7.9</td>
<td>Actual v/s Forecasted demand by SVM-SA</td>
<td>109</td>
</tr>
<tr>
<td>7.10</td>
<td>Comparison of Actual Load with Forecasted values of four Techniques</td>
<td>109</td>
</tr>
<tr>
<td>7.11</td>
<td>Accuracy Comparison between Different Techniques on Hourly Data</td>
<td>110</td>
</tr>
<tr>
<td>7.12</td>
<td>Actual v/s Forecasted Demand by ANN-RPROP</td>
<td>110</td>
</tr>
<tr>
<td>7.13</td>
<td>Actual v/s Forecasted Demand by ANN-PSO</td>
<td>111</td>
</tr>
<tr>
<td>7.14</td>
<td>Actual v/s Forecasted Demand by ANN-LMA</td>
<td>111</td>
</tr>
<tr>
<td>7.15</td>
<td>Actual v/s Forecasted Demand by SVM-SA</td>
<td>112</td>
</tr>
<tr>
<td>7.16</td>
<td>Comparison of Actual Load with Forecasted values of Four Techniques</td>
<td>112</td>
</tr>
<tr>
<td>7.17</td>
<td>Accuracy Comparison between Different Techniques on Daily Data</td>
<td>113</td>
</tr>
</tbody>
</table>
List of Tables

1.1 Sectoral Electricity Consumption by Province 2009-10 (GWh) [1] .... 4
1.2 Supply and Demand Estimates [1] ............................................ 4
1.3 Needs of forecasts in utilities ................................................. 10

3.1 Input data set with Forecasts(Sample) .............................. 46
3.2 MAPE and standard deviation of error over 30 runs for three techniques . 47
3.3 Average MSE Comparison of Three Techniques .................. 47
3.4 Asymptotic Test ................................................................. 48

4.1 Comparison on Benchmark Functions ................................. 55
4.2 Results on Electric Demand data ........................................... 56

5.1 Evaluation of Modified Versions by Bench-Marking Functions .... 65
5.2 Sample LESCO Daily Peak Data ........................................... 68
5.3 Average MAPE over 10 runs of all Techniques on PEPCO Daily Peak Data 71
5.4 Average MAPE over 10 runs of all Techniques on PEPCO Weekly Peak Data ......................................................... 71
5.5 Average MAPE over 10 runs of all Techniques on LESCO Daily Peak Data 73
5.6 Average MAPE over 10 runs of all Techniques on LESCO Weekly Peak Data ......................................................... 75

6.1 Electricity Demand on January 14, 2010 ............................... 81
6.2 Performance Comparison of Models .................................... 82
6.3 Performance Comparison of Models on IESO data .................. 84
6.4 Weekly Peak Data of 4 Months ............................................ 86
6.5 Accuracy Comparison on Weekly Data .................................. 87
6.6 Industrial consumption data with input factors ..................... 91
6.7 Performance Comparison on Industrial Data ....................... 92
6.8 Results on Cancer data ....................................................... 97
6.9 Results on Heart data .......................................................... 97
6.10 Results on Diabetes data .................................................... 98
6.11 Results on Horse data ....................................................... 98

7.1 Correlation of Input Factors with Actual Load ..................... 102
7.2 Hourly Peak Demand Data with Weather Conditions ............. 106
7.3 Daily Peak Demand Data with Weather Conditions ............... 107
7.4 Accuracy Comparison between Different Techniques on Hourly Data 109
7.5 Accuracy Comparison between Different Techniques on Daily Data 111
Abbreviations

LF  Load Forecasting
ELF Electric Load Forecasting
STLF Short Term Load Forecasting
MTLF Medium Term Load Forecasting
LTLF Long Term Load Forecasting
PPS Pakistan Power Systems
LESCO Lahore Electric Supply Company
NTDCL National Transmission and Despatch Company Limited
NPCC National Power Control Center
IPP Independent Power Producer
IESO Independent Electrical Supply for Ontario
MSE Mean Squared Error
MAPE Mean Absolute Percentage Error
ANN Artificial Neural Networks
SVM Support Vector Machines
AI Artificial Intelligence
FFNN Feed Forward Neural Networks
ABC Artificial Bee Colony
PSO Particle Swarm Optimization
GA Genetic Algorithm
SA Simulated Annealing
ACS Artificial Cooperative Search
Definitions

1. **Electric Power System**: A network of electrical components used to supply, transmit and use electric power.

2. **Electric Power System**: A network of electrical components used to supply, transmit and use electric power.

3. **Power Purchasing Agency**: Responsible to purchase agreed capacity from power producers.

4. **Power Producers**: Power Generation Companies.

5. **Distribution Companies**: Power Distribution Companies connected to power consumers.

6. **Electric Power Distribution System**: It carries and delivers electricity from the transmission system to individual consumers.

7. **Consumers**: Electric Energy Consumers which can be further classified into industrial, domestic, commercial, and agricultural.

8. **Electric Utility**: Electricity Supply Company.

9. **Electric Load forecasting**: A process to predict the future demand of electric power consumers.

10. **Power System Planning**: A process to decide new installation or upgrading existing power system elements to satisfy the future load.

11. **The Sampling Interval**: The frequency of data collection.

12. **Optimization**: A process to find the most suitable value for a function within a given domain.

13. **Overfitting**: It generally occurs when a model is excessively complex.

14. **Global Minimum**: The absolute minimum, is the smallest overall value of a set, function, etc., over its entire range.

15. **Machine learning**: Scientific discipline that deals with the construction and study of algorithms that can learn from data.
Abstract

Electric load forecasting (ELF) involves the projection of peak demand levels and overall energy consumption patterns to support an electricity company’s future system and business operations. The ELF helps to anticipate these problems of power delivery: How much load needs to be delivered and when? One step ahead of this magnitude and temporal predication is the localization (where) of the load. Hence, it involves the accurate prediction of both the load magnitude and geographical locations of electric load over the different time periods. Short and mid-range predictions of power load allow electricity companies to retain high energy efficiency, reliable operation, and making optimized plan for power generation and distribution. As, power generation resources are limited and there is a crisis like situation in energy sector of Pakistan. Therefore, the foremost objective of this research is to empower Pakistan power systems with accurate load forecasting tools. Also, different type of planning and decision making requires different levels of forecasting such as: short term, medium term and long term forecasts. The ELF is a complex task due to the nature of load affecting variables. Dealing with abrupt changes in weather conditions and modeling consumer’s usage behaviors is a challenging task. Machine learning based statistical and artificial intelligence techniques are widely used for ELF. Among these, artificial neural networks (ANN) and support vector machines (SVM) emerge as competitive modeling approaches for ELF. Still, appropriate selection of SVM and ANN parameters requires proper optimization techniques to avoid slow convergence, local minima, and over-fitting of models. Another challenge lies in the generalization of these models; none of the reported models so far, has attained a sturdy stature as a generally applicable technique. The modern optimization techniques are robust, less time-consuming, dependable, and provide high quality solutions for parameter selection and model development. The accuracy of modeling techniques is extremely dependent on quality of historical data. Since, the recording of data in Pakistan power systems was previously manual. This data contained abnormalities like missing values, outliers, and duplication of records. Observing all the aforementioned problems, we got motivation to devise such an ELF model that can perform well on noisy data of Pakistan power systems and can handle load affecting parameters of this region. Hence, we proposed an integrated and customized model of short term ELF model. We included various components to the proposed system, such as: pre-treatment of historical data, analysis, transformation, cross validation, and over-fitting modules. The pre-processing
module deals with outliers, anomalies, resolving the missing data problems in load series, curve smoothing, and data normalization. In this thesis, we applied state-of-the-art machine learning techniques for data pre-processing, analysis, modeling, and optimization to develop a comprehensive ELF model to achieve goal of better generalization and improved accuracy. To achieve these goals by this research, we developed techniques for data pre-processing, data analysis, and load forecasting. We combined optimization techniques from statistical and artificial intelligence domains with ANN’s and SVMs. We tested our proposed models on electric demand and load affecting parameters data collected from national and local distribution companies of Pakistan and Ontario state electric supply company. We evaluated our models on the basis of forecasting accuracy, benchmark function, and by statistical significance tests.
List of Publications


Chapter 1

Introduction

Prediction is very difficult, especially if it’s about the future.

Nils Bohr, Nobel laureate in Physics

1.1 Electric Power Systems

Power industry provides backbone for a country’s development and is one of the most important utilities that underpin the survival of a nation’s critical industrial, telecommunication, and agricultural infrastructures. Electric power systems have four components, which are power generators, transmitters, distribution company, and consumers. These components are presented in Fig. 1.1. Power purchasing agency is responsible to purchase agreed capacity from power producers and to distribute to distribution companies. Power producers may directly be connected to distribution companies. Distribution companies are responsible to meet energy requirement of their consumers. On the other hand, consumers are classified into industrial, domestic, commercial, and agricultural. Which all have different energy requirements over different time spans.

All the components of an electric power system have their own requirements. The requirements of these stakeholders are shown in Fig. 1.2. Generation companies want to fully utilize their installed capacity. They need to go for yearly planned maintenance of their unit. They also have to declare their generation capacity in advance over the year. The purchasing authority has to make plans to meet energy requirements of its consumers. In a deregulated economy, purchasing authority makes unit commitment with independent power producers (IPP) to purchase the agreed capacity. The purchasing authority also maintains a merit order of IPP’s so that energy requirements can be met in an economical order. Some units are cheaper than others. Even if the energy requirements can be satisfied by cheaper units generated from hydro power generation sources, why to go for expensive fuel based thermal generation. In this way, purchasing agency tries to keep a balance between demand and supply. It has to plan infrastructure
development to meet future needs of consumers, as it takes years to establish new power plants. Similarly, distribution companies want full utilization of the available energy to maximize profits, by avoiding the over-spinning or over generation. They require area wise energy requirements to keep a balance between demand and supply and hence, avoid load shedding or system overloading. Consumers are essential component of the electric power system, as they generate power demand and require un-interrupted power supply. All other components are interested to know the current and future demand of the consumers. As, the power generation units cannot be started or stopped instantaneously; neither the generated energy can be stored efficiently for future use. It is, therefore, very important to plan generation, distribution, and transmission of electric power for current and more importantly future needs.

**Figure 1.1: Components of Electric Power System**

**Figure 1.2: Requirements of Each Component in Electric Power System**
1.2 Power System of Pakistan

Pakistan is a vibrant country with ever increasing demand for electric power in each sector of community. The gross electricity demand has increased from 12,400 MW in 2003 to 20,400 MW in 2010, roughly at the rate of 7.4% per annum. However, during the same period the installed capacity grew from 18,800 MW to only 22,300 MW which is at a rate of 3.2% per year [1]. The inability to meet the demand is exacerbated by the fact that not all the installed capacity is available when most needed due to capacity de-ratings, fuel constraints, forced outages and other operational difficulties such as with transmission line and transformer overloading. Due to the seasonal nature of hydroelectric generation, some of the 6,555 MW (30% of total installed system capacity) of the hydro capacity varies on a seasonal basis. The Pakistan Power Sector consists of two major Power Utilities, i.e. Pakistan Electric Power Company (PEPCO) and Karachi Electric (K-Electric) formerly known as Karachi Electric Supply Company (KESC). K-Electric has been privatized; while PEPCO consist of five generation companies, nine Distribution Companies (DISCO’s), National Power Control Center (NPCC), National Transmission and Despatch Company Limited (NTDCL), and IPP’s [1]. This break up is presented in Fig. 1.3. The electricity generation sector in Pakistan is a mix of hydel, thermal, and nuclear energy. The hydel generation system contributes around 31% of power; 66.8% of overall power is generated from thermal sources. Whereas, only 2.2 percent is obtained through nuclear sources. NTDCL is solely responsible for electricity transmission and dispatching in whole country (Except area covered by K-Electric) via national grid. Further, there are nine distribution companies, who channel electricity from the transmission grid stations (220kV) to the consumers. The customers can be categorized as domestic, industrial, commercial, agricultural, street lights, etc. Sector wise electricity consumption categorized by provinces is shown in Table. 1.1. The gross total of whole country is 74,348.2 (GWh), out of which domestic sector is withdrawing the major part of 34,272 (GWh). Domestic consumption is further divided into province wise load. Industrial, agriculture, and commercial sectors are major contributors of energy demand, respectively. Whereas, supply and demand estimates upto 2030 are presented in Table. 1.2. It is expected that net generation would be 90,000 (MW) by year 2030, with an average growth rate of 10%. On the other side, peak demand (MW) could rise upto 101,478 (MW) by year 2030, which is currently ranging between 25,000 (MW) to 32,000 (MW), with an average growth rate of 8%.

Due to limitation of energy resources, poor planning, environmental factors, ever increasing power demand, and limitation of transmission network; power sector in Pakistan is unable to fulfill the need of its consumers. Still, determining the need and behavior of consumers can help to provide available energy efficiently. Absence of this, results in a current handicapped situation of Pakistan power sector; in which, load is either under-estimated or overestimated; both situations are not acceptable for a reliable operation of electric utility.
Chapter 1. Introduction

Figure 1.3: Break up of Pakistan Power System [1]

Table 1.1: Sectoral Electricity Consumption by Province 2009-10 (GWh) [1]

<table>
<thead>
<tr>
<th>Sector</th>
<th>Punjab</th>
<th>Sindh</th>
<th>KP K</th>
<th>Balochistan</th>
<th>AJK</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>20,824.6</td>
<td>6,906.3</td>
<td>5,453.4</td>
<td>462.6</td>
<td>625.1</td>
<td>34,272.0</td>
</tr>
<tr>
<td>Commercial</td>
<td>3,528.7</td>
<td>1,452.0</td>
<td>483.4</td>
<td>92.6</td>
<td>48.8</td>
<td>5,605.4</td>
</tr>
<tr>
<td>Industrial</td>
<td>13,640.6</td>
<td>4,561.7</td>
<td>1,443.0</td>
<td>113.4</td>
<td>64.7</td>
<td>19,823.4</td>
</tr>
<tr>
<td>Agriculture</td>
<td>5,087.9</td>
<td>890.0</td>
<td>363.1</td>
<td>3,348.3</td>
<td>-</td>
<td>9,689.3</td>
</tr>
<tr>
<td>Street light</td>
<td>234.9</td>
<td>199.4</td>
<td>19.9</td>
<td>3.7</td>
<td>-</td>
<td>457.9</td>
</tr>
<tr>
<td>Traction</td>
<td>2.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.3</td>
</tr>
<tr>
<td>Bulk supplies</td>
<td>2,566.9</td>
<td>1,278.5</td>
<td>493.7</td>
<td>78.4</td>
<td>-</td>
<td>4,417.4</td>
</tr>
<tr>
<td>Other</td>
<td>20.3</td>
<td>5.6</td>
<td>2.9</td>
<td>0.1</td>
<td>51.7</td>
<td>80.6</td>
</tr>
<tr>
<td>Total (GWh)</td>
<td>45,906.2</td>
<td>15,293.3</td>
<td>8,259.4</td>
<td>4,099.1</td>
<td>790.3</td>
<td>74,348.2</td>
</tr>
</tbody>
</table>

Table 1.2: Supply and Demand Estimates [1]

<table>
<thead>
<tr>
<th>Fiscal Years</th>
<th>2010</th>
<th>2015</th>
<th>2020</th>
<th>2025</th>
<th>2030</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Dependable Capacity (MW)</td>
<td>19,477</td>
<td>27,000</td>
<td>40,000</td>
<td>70,000</td>
<td>90,000</td>
</tr>
<tr>
<td>Growth Rate</td>
<td>15%</td>
<td>9%</td>
<td>10%</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>Peak Demand (MW)</td>
<td>22,353</td>
<td>32,704</td>
<td>48,843</td>
<td>72,169</td>
<td>10,1478</td>
</tr>
<tr>
<td>Growth Rate</td>
<td>9%</td>
<td>8%</td>
<td>9%</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>Surplus/Deficit</td>
<td>-2,876</td>
<td>324</td>
<td>4,066</td>
<td>4,031</td>
<td>5,087</td>
</tr>
</tbody>
</table>
1.3 Electric Load Forecasting

Energy generation and distribution planning to meet the current and future needs of consumers is a challenging task. In developing countries, especially in Pakistan, generation is not up to the mark. Lack of energy resources, poor planning, and limitation of the transmission network are the prominent problems of the power sector. Still, to distribute the available energy efficiently, accurate forecasting of energy requirement of customers can contribute a lot. Absence of proper planning and forecasting result in power shortfall and system instability [2].

Electric Load Forecasting (ELF) is an essential component of electric power systems, which helps determine futuristic consumer side demand [2]. Accurate estimates of future demand help in planning generation and distribution of electricity [3]. It helps in keeping a balance between generation and distribution; increasing efficiency and reliability of the system in an economic manner [4]. Short and mid-range predictions of electricity load allow electricity companies to retain high energy efficiency and reliable operation. Absence of such prior planning results in a current crisis like situation in Pakistan, where power generation is not up-to the mark, its fallout is forced load shedding and voltage instability. The ELF helps an electric utility to make crucial and necessary decisions including decisions on purchasing, generating electric power, load switching, and infrastructural development. Load forecasts are extremely important for energy suppliers, financial institutions, and other participants in electric energy generation, transmission, distribution, and markets geographical. Accurate load forecast can be helpful in developing a power supply strategy and development plan; especially for developing countries where the demand is dynamic in nature due to high growth rate and high rural-urban drift. In today's world, ELF is an important process in most utilities with the applications spread across several departments, such as planning department, operations department, trading department, etc.

The nonexistence of precise ELF results in load under-estimation or over-estimation, both situations are not acceptable because load underestimated results in shortfall of energy and calls for costly emergency power purchases. Load over-estimated results in spinning up unnecessarily additional or reserve units. The accurate ELF facilitates to maintain a balance between electricity supply and demand in economical way.

The ELF helps to anticipate these problems of power delivery: How much load needs to be delivered and when? One step ahead of this magnitude and temporal prediction is the localization (where) of the load. It involves the accurate prediction of both the load magnitude and geographical locations of electric load over the different time periods.
Chapter 1. Introduction

The ELF can be described as extrapolation of trend line. In the graph presented in Fig. 1.4, the actual demand of daily peak load is presented. Whereas, the predicted line is following the actual demand curve line and is termed as forecasted line. The difference between both lines is termed as error of forecasting.

In today’s economy, there is a need to forecast and predict future load demands so as to determine an estimate of the amount of load required and to properly prepare for generation purposes. The ELF is a method of estimating the electric load required by a certain geographical area using previous records of load usage in that geographical area and other important factors such as: population, weather forecasts, temperature, etc. The issue in ELF is to obtain future electric demand values by mainly extrapolating past load consumption and considering other factors which affect the amount of electrical load used in any place at any point in time. Due to the fact that electric energy cannot be stored for future use with present technology, the knowledge of how electric load behaves in the future is very precious information. The forecasted demand data is also used for hydro-thermal co-ordination: if the requirement can be fulfilled by cheaper energy sources why to purchase expensive energy, resulting in fuel saving. The power interchange helps in allocating the spare energy of one area to other stations or consumers. Therefore, ELF also helps in contingency planning and making proper load shedding plans.

The ELF is generally categorized as: short, medium, and long term ELF [4]. The short-term LF (STELF) predicts the load demand ranging from one hour to one day [2]. It helps in day to day operation of power systems and in balancing demand and supply curves in real time. Medium-term ELF (MTELF) span up-to several weeks, and it supports decisions related to energy transactions and dispatching, coordination of hydro-thermal generation units, fuel allocation, and scheduling the maintenance plan. Whereas, long-term ELF (LTELFLF) range from one year up-to 10 years. These forecasts are used for system expansion, generation & distribution planning, and infrastructure development.
1.4 Classification of Electric Load Forecasts

There is no single forecast that can satisfy all of the needs of utilities. A common practice is to use different forecasts for different purposes. The classification of various forecasts is not only depending upon the business needs of utilities, but also the availability of the crucial elements that affect the energy consumption: weather (or climate in the long periods) and human activities. Load forecasts can be divided into different categories.

1.4.1 Temporal Forecasts

Temporal forecasts are time dependent forecasts. These can be further categorized as short, medium, and long term forecasts on basis of time horizon of forecasts. Short-term forecasts which usually range 1 hour to 1 week, medium forecasts span from a week to a year, and long-term forecasts are usually longer than a year. Load forecasting types and influencing factors are listed in Fig. 1.5.

![Figure 1.5: Load Forecasting Types](image)

**Short Term Electric Load Forecasting** *One hour to a week.* The short term load variation depends on weather, local events, type of day (Weekday or Holiday or Weekend), time of day, etc. Weather is the most influencing factor on short term load fluctuations, abrupt changes in temperature or weather condition cause variation in load curve. Time is also very important factor. As demand is different at different time of the day. Similarly, day of the week and special events are contributing factors. The day to day operations of the utility demand for precise short term load forecasting.

**Medium Term Electric Load Forecasting** *One week to a year.* Medium Term forecast depends on seasonal variations, electricity pricing, economical trends, river flow conditions, and irrigation/ crop seasons. The planning of maintenance, scheduling of the fuel supply, etc. calls for medium term load forecast.

**Long Term Electric Load Forecasting** *Longer than a year.* The long term load increase depends on the population growth, local area development, industrial expansion, etc. The electric system planning needs the forecast of the load for several years for infrastructure development.
1.4.2 Spatial Electric Load Forecasting

Forecasting the cumulative load of geographical diverse sites to fulfill the energy requirements of an area. Spatial load forecasts determine the consumption requirements on spatial basis. It helps to better understand and fulfill the energy requirements of an area. This could help in transmission & distribution (T&D) planning. NTDCL is responsible for medium and long term planning and management of power generation. NPCC handles the overall distribution of electrical energy on real time basis. Whereas, DISCO’s are responsible for distribution of electric power and cover different parts of country. Medium and long term consumer demand forecasts are required at NTDCL for generation planning. Short term and medium term load forecasts are required at NPCC and at each DISCO for proper utilization of available energy.

1.4.3 Consumers based Forecasts

The energy consumption of Pakistan can also be categorized on consumer classes. This categorization is shown in Fig. 1.6. The domestic users consumes about 46% of overall produced energy. Then is the industrial sector which requires about 28%, Commercial sector takes 8%, and agriculture sector consumes about 12% by tubewells, Others consumers include bulk supply and street lighting which utilizes 6% of total.

![Figure 1.6: Sector wise Energy Consumption](image)
1.4.4 Multi-Scenario Load Forecasting
Apart from the regular short term or medium term forecasts, electric utilities are often interested in what-if load scenarios for effective planning. The effective multi-scenario load forecasting is required to handle the uncertainty in load conditions for system stability and to develop contingency plans. Electric load may vary if the load affecting conditions are changed abruptly or situation is altered. Large industries consumers, like steel mills, may cause sudden load changes. Moreover, special events and certain conditions results in transient load changes such as Timing of Popular TV show or any schedule timing of shutdown of industrial operations effect on the total load of the Power System. Therefore, multi-scenario ELF is also helpful in Transmission and Dispatch planning at micro levels.

1.5 Business Needs of Load Forecasts
The business needs of the utilities can be summarized, but not limited to, the following:

1.5.1 Energy Purchasing
Whether a utility purchases its own energy supplies from the market place, or outsources this function to other parties, load forecasts are essential for purchasing energy. The utilities can perform bi-lateral purchases and asset commitment in the long term, e.g., 10 years ahead. They can also do hedging and block purchases one month to 3 years ahead, and adjust (buy or sell) the energy purchase in the day-ahead market.

1.5.2 Transmission and Distribution (T&D) Planning
The utilities need to properly maintain and upgrade the system to satisfy the growth of demand in the service territory and improve the reliability. Sometimes the utilities need to hedge the real estate to place the substations in the future. The planning decisions heavily rely on the forecasts, known as spatial load forecasts, that contain when, where, and how much the load as well as the number of customers will grow.

1.5.3 Operations and Maintenance
In daily operations, load patterns obtained during the load forecasting process guide the system operators to make switching, loading decisions, and schedule maintenance outages.

1.5.4 Demand Side Management (DSM)
Although lots of DSM activities are belong to daily operations, it is worthwhile to separate DSM from the operations category due to its importance in this smart-grid world. A load forecast can support the decisions in load control and voltage reduction. On the other hand, through the studies performed during load forecasting, utilities can perform long term planning according to the characteristics of the end-use behavior of certain customers.
1.5.5 Financial Planning

The load forecasts can also help the executives of the utilities project medium and long term revenues, make decisions during acquisitions, approve or disapprove project budgets, plan human resources, and technologies, etc. According to the lead time range of each business need described above, the minimum updating cycle and maximum horizon of the forecasts are summarized in Table 1.3. Other advantages of load forecasting include: optimization of network planning, investment, better heuristic management of risk, and reduction in cost.

<table>
<thead>
<tr>
<th>Application</th>
<th>Minimum updating cycle</th>
<th>Max horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy purchasing</td>
<td>1 hour</td>
<td>10 years and above</td>
</tr>
<tr>
<td>T&amp;D planning</td>
<td>1 day</td>
<td>30 years</td>
</tr>
<tr>
<td>Operations</td>
<td>15 minutes</td>
<td>2 weeks</td>
</tr>
<tr>
<td>DSM</td>
<td>15 minutes</td>
<td>10 years and above</td>
</tr>
<tr>
<td>Financial Planning</td>
<td>1 month</td>
<td>10 years and above</td>
</tr>
</tbody>
</table>

1.6 Motivation

Since, the electric load forecast is fundamental and first step in the planning process of any electric utility (Power Supply Company). It also involves the accurate prediction of both the load magnitude and geographical locations of electric load over the different time periods. Therefore, the motivation behind this research is to empower electric power system of Pakistan with accurate and timely forecasting models. Precise ELF would help in making proper plan for optimized power generation and distribution. We are aimed at facilitating Pakistan power system (PPS) in making accurate unit commitments for power purchases, scheduling maintenance plan of generation units and to reduce spinning reserve capacity to avoid over and under generation. In this way, estimates of future demand can help in maintaining the reliability of power system in an economical manner. Similarly, our objective is to find the best suitable combination of techniques for different types of forecasting. Short term, medium term, and long term forecasting models may be proposed and evaluated on the basis of accuracy and efficiency.

Observing the business needs of forecasting in PPS and scrutinizing the issues with recorded load data; we got motivation to solve the problem of accurate ELF for noisy data of indigenous power sector. The hypothesis behind this study is that modeling techniques can perform well on our datasets, when the data is smooth and free of abnormalities. In order to achieve better accuracy on noisy data sets; we are enthused to investigate different statistical and heuristic techniques to suggest an accurate load forecasting model that can fulfill the above mentioned requirements of PPS. For this, we are intended to incorporate machine learning techniques for model development, optimization, and data pre-processing. We are envisioned to perform an analytical study on varying behavior of different techniques on predicating futuristic load utilizing different combination of
Chapter 1. Introduction

Section 1.7 Problem Statement

The ELF in developing countries is very difficult because of the high growth rate of load demand and the wide differences in the modes and levels of consumption from one geographical area to another. The forecasting problem can be expressed as a mathematical form with the objective of finding a function that best fits to load affecting variables and produces the forecasted values for next time instance. The basic formulation of forecasting function is given in Eq. 1.1, to map independent variables to find dependent variable value.

\[ L_f(t + 1) = f(L(t), L(t - 1), \ldots, L(t - k), c, w, e) \]  

(1.1)

Where, \( L_f(t + 1) \) is the forecasted load value for time instance \( t + 1 \), and \( L(t) \) is the actual load value for the time instance \( t \), it also corresponds to historical load values. The variables \( c, w, \) and \( e \) represent the calendar, weather, and random events respectively.

The influence of various parameters on ELF is shown in Fig. 1.7. The function \( f \) in this figure corresponds to potential function or model which can produce the forecasted load values given load history and other load affecting parameters. This research work is focused to propose the best forecasting model while satisfying our research objectives.

![Figure 1.7: Forecasting Function](image)

Predicting load series is a complex task due to nature of load affecting variables. Dealing with abrupt changes in weather conditions, modeling consumer’s usage behaviors is challenging task. Especially, when there is mix load of different type of consumers, such as domestic, industrial, commercial, and agriculture. A lot of models are reported in literature to solve the ELF problem. The focus of research is now shifted towards building models for improved accuracy of forecasts with the use of modern statistical and computational intelligence methods.
Several ELF systems have been developed world wide and in most cases different countries have adopted different ELF systems suitable to their environment conditions [5–8]. Most of developed models serve the needs of one specific area or country and are customized according to their priorities, custom features and varied input parameters [4]. Beside a number of methods and techniques have been developed and reported in literature but there is a lack of an ample solution that handles all factors, covers all forecast types, and is adaptive enough to apply on different geographical locations [2]. Also different models consider varying number of input parameter set based on data availability and impact of these particular inputs on power demand [3].

Still, most of work is focused on data from one site only. These techniques cannot be applied to other geographical diverse areas. So, we cannot adopt any of these developed models to our local environment. Also, there is no single model in literature that handles all the load parameters [2]. As, discussed in [3, 4], data availability determines the input variables and model selection depends on the situation under consideration. Therefore, no general recommendation can be given to propose a global forecasting model for each scenario. So, one of the foremost decision to be made is therefore, selection of appropriate model based on our priorities, data availability and problem under consideration. We have unique problem in this way, we have many factors effecting load demand such as water flow conditions, political situations effecting socio-economic outlooks, weather conditions, electricity rates, and other community oriented conditions require consideration while modeling a load forecasting system. On the other side, we have immense geographical diverse areas which have a different mix of factors.

Despite, the availability of so many procedures which have been tested and successfully adopted. Another challenge lies in the generalization of these models, none of the reported models so far, has attained a sturdy stature as a generally applicable technique. This is due to varying nature of load affecting parameters, custom requirements, and particular circumstances of an area, which play an important role to choose the appropriate ELF model. Even if a forecasting model has performed best on one situation cannot be assured to show the same accuracy of forecasts on other scenario. The reported models are usually built for one specific problem of a utility. Therefore, large error appear when the same models are applied to other data sets of different utility or situation. How to select a method or a combination of such methods suitable for varying circumstances/scenario is difficult research problem. Hence, the results presented in the literature are usually not comparable to each other. This research is focused on evaluation of different models and to develop a comprehensive model to achieve desired goals.

Due to manual recording and data measurement problems of power station in Pakistan, data is prone to have abnormalities. This existence of outliers or missing values in historical load curves affect the accuracy of ELF results. Therefore, before passing to ELF models, all the impurities should be removed from raw data, through proper use of pre-processing techniques.
1.8 Data Description

Electric demand data and other load affecting attributes data has been collected from different sources like NTDCL, LESCO, meteorological department, and other concerning authorities. Weather forecasts and previous year’s meteorological data is collected from meteorological department of Pakistan. For hourly forecasts, there should be 3 to 5 years of hourly historical load data for each electric supply utility. So far, we have collected following type of data sets from different authorities. Detail of each data type has been covered in following subsections. In following section, we have listed the major data sources of energy demand and consumption data, sources of weather data and other effecting parameters.

1.8.1 Electricity Consumption Data

In following we have listed the major data sources of energy demand.

LESCO Data

We have collected daily and hourly load profiles for Lahore Electric Supply Company (LESCO).

- Hourly Load Data (2008-2011)

This data set contains the cumulative hourly load consumption data for the 4 years from year 2008 to 2011. It contains 24 hourly data entries for a single day.

- Daily Peak (Max) Load (2005-2011)

This data set contains the daily peak demand data for 6 years from 2005 to 2011. Daily peak data contains only one entry for a single day.

PEPCO Data

PEPCO maintains country wide consumption data for hourly and daily basis. The data sets collected are as follows.

- Hourly Load Data (Country Wide) (2006-2010)

This data set contains the cumulative hourly load consumption data for the 5 years from year 2006 to 2010. It contains 24 hourly data entries for a single day.
Daily Peak Demand (2006-2010)

This data set contains the daily peak demand data for 5 years from 2006 to 2010. Daily peak data contains only one entry for a single day that depicts the peak load demand of the day.

1.8.2 Driving Factors of Electricity Consumption

Load consumption pattern are driven by many influencing factors like weather, seasonal changes, economical trends, population and industry growth rates, usage patterns, and each and every human activity. Here, we have identified some of the key parameters that have affect on load variation.

Meteorological Data

Weather refers to the present condition of the meteorological elements, such as temperature, humidity, wind, rainfall, etc., and their variations in a given region over periods up to two weeks. Climate encompasses these same elements in a given region and their variations over long periods of time. Weather conditions are responsible for significant variations in the load pattern on short term load fluctuations. It is true because most utilities have large components of weather-sensitive load due to different index value that can be find from the different index system of weather, such as those due to space heating, air conditioning, and agricultural irrigation. Sudden changes in weather effects variation in load demand. Observations of different Metrological factors the temperature, relative humidity, wind speed, and the human comfort index effecting on the load. These conditions cause large variation in the aggregated load. In addition to the temperature, wind speed, cloud cover, and humidity have a large influence. Weather and meteorological related information has been collected from Meteorological Department of Pakistan.

Population Growth Rate

Population growth is one of the major factor that affect the load demand, population growth rate is not a constant factor in long run. However, it also need prediction according to the demographic area to calculate actual growth rate. It shows the increase in the users of electricity. It needs long term planning to overcome this factor by installation of power stations in next few years. The graph presented in Fig. 1.8. shows the year wise increase in domestic consumers. This data set is collected from NTDCL Pakistan.

Industrial Growth Rate

The cumulative industrial load is determined by electricity units consumed by industry. This load is quite steady and does not effected by calendar or weather variables. The year wise increase in industrial consumers is presented in Fig. 1.9. Year wise Industrial consumers count, Industrial tariff, and industrial usage history and demand data is collected from NTDCL Pakistan.
Economical Changes

Supply and demand is the fundamental concept of economy. The economic environment in which the utility operates has a clear affect on the electricity demand. More generally, economic trends have significant impacts on the system load growth/decline trend. The economy affects the development of new power generation or plants in future. Economist estimates the indexes of the country which are used in long term load forecasting. Mostly economic conditions decided either new generation plants need to develop or extension of old generation plants would be suitable. The year wise commutative gross domestic product (GDP) is presented in Fig. 1.10. Year wise and sector wise GDP data is collected from NTDCL Pakistan.
Chapter 1. Introduction

Load Consumption Factor

Load factor is defined as the ratio of the average load in kilowatts supplied during a selected period to the maximum load in kilowatts occurring in that period. So, this factor directly defines the number of peaks in certain period of time. In long term load forecasting, prediction of number of peaks in power system are required. This yearly data set is collected from NTDCL Pakistan.

Electricity Price

Electricity price is affected in load forecasting when fuel prices have sudden impact in forecasting interval which results decrease in demand which is more prominent in medium time interval.

Average tariff in paisa/KWh of 41 years for consumers is obtained from NTDCL Pakistan. The variation in average tariff on yearly basis is shown in Fig. 1.11.

![Figure 1.11: Average Tariff](image)

Calendar data: Day and Time Factors

Calendar events are major contributor towards the load fluctuation. The demand curve varies with time of the day so there is hourly cycle for each day. Load pattern of working days is different from the weekends so there exists a weekly daily cycle. National and religious holidays have different load profiles. There is also a seasonal effect as load curve have different pattern for different months of the year. Following graph shows the electric demand variation in different times of the day. It also indicates difference between actual load and supply. Short term load forecasting used for a single day to a week, so depends upon the day or week to be forecasted is either the particular day is holiday or any seasonal events is occur in the week, the information of the prediction day/week is very important which could vary from the forecasted normal day/week. This information is required to improve our prediction technique. For example, a normal day in which forecasted Load is average to the normal day, so load forecasting techniques have not required any additional inputs or forecasting factors, but if the technique is used with same input factors for a special day, e.g. holiday - industries and offices are
closed on that day, economical issues to power distribution companies increases to large extent. Following graph shows the effect of two consecutive but different types of days like working day and holiday on load demand. Calendar data, seasonal variations, special days, holidays, local and Hijri Calendar events have been processed and cross checked through internet and different newspapers.

1.8.3 Salient Features of Electric Load Series

In this section, complex relationship between demand and other factors have been shown using multiple graphs. As shown in Fig. 1.12, temperature and demand have peak values at different time interval of a typical summer day. Same is the case with day types; demand curve of one day does not match demand of next day or previous day. This is shown in Fig. 1.13. Load curves of previous day as well as same day of last week are plotted against current demand. We can see different demand profiles for different days of same week in Fig. 1.14.

![Figure 1.12: Effect of Temperature on Load](image1.png)

![Figure 1.13: Relationship between Current and Previous Demand](image2.png)
Seasonal and monthly changes also affect energy demand curves, as shown in Fig. 1.15. The demand of two consecutive days is plotted with one year shift. Although day types are different but there is a clear upward shift of demand.

Figure 1.14: Comparison of Load Profiles of two Different Week Days

Figure 1.15: Demand Shift after One Year
1.9 Research Methodology

The motive of this research is to propose the best ELF models which are quite suitable for geographical and demographical conditions of Pakistan. To achieve the objective of more generalized ELF model, we investigated many classical time series and regression based techniques, then we incorporated a number of modern approaches of forecasting. As shown in Fig. 1.16, different time series and regression methods are listed, which we used to analyze and predict the demand series. In the next level of this research, we opted to work with more advanced and modern approaches being adopted worldwide. A number of ANN and SVM based modeling techniques are utilized. Whereas, the objective function of ANN and SVM based models is optimized by using appropriate optimization methods listed. From this pool, we constructed numerous models by the combination of modeling and optimization techniques. These hybrid models are tested on various data sets in subsequent chapters. The process of obtaining optimized connection weights / parameters is shown in Fig. 1.17.

<table>
<thead>
<tr>
<th>Modeling Techniques</th>
<th>Architecture</th>
<th>Optimization Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feed Forward Neural Networks</td>
<td>1. Feed Forward With Bias</td>
<td>1. Fire Fly Algorithm</td>
</tr>
<tr>
<td></td>
<td>2. Feed Forward Without Bias</td>
<td>2. Particle Swarm Optimization</td>
</tr>
<tr>
<td>Recurrent Neural Networks</td>
<td>1. Elman</td>
<td>3. Back Propagation</td>
</tr>
<tr>
<td></td>
<td>2. Jordan</td>
<td>4. Levenberg Marquardt Algorithm</td>
</tr>
<tr>
<td>Support Vector Machine</td>
<td>1. NU-SVR</td>
<td>5. Scaled Conjugate Gradient</td>
</tr>
<tr>
<td></td>
<td>2. Epsilon-SVR</td>
<td>6. Resilient Back Propagation</td>
</tr>
<tr>
<td></td>
<td>3. Radial Basis Function</td>
<td>7. Simulated Annealing</td>
</tr>
<tr>
<td>Regression</td>
<td>1. Linear Regression</td>
<td>8. Quasi Newton Method</td>
</tr>
<tr>
<td></td>
<td>2. Polynomial Regression</td>
<td>9. Artificial Bee Colony</td>
</tr>
<tr>
<td></td>
<td>3. Multiple Polynomial Regression</td>
<td>10. Genetic Algorithm</td>
</tr>
<tr>
<td></td>
<td>2. Moving Average</td>
<td>12. Cuckoo Search</td>
</tr>
<tr>
<td></td>
<td>3. Weighted Moving Average</td>
<td>13. Artificial Cooperative Search</td>
</tr>
</tbody>
</table>

Figure 1.16: Modeling and Optimization Techniques

Initially, a survey on existing techniques and models is conducted. We further classify these techniques into different groups depending upon the nature of models. Since, a number of techniques are used and reported in literature to build an accurate forecasting model. Out of them ANN’s and SVM’s are proven most promising technique for ELF model building. Many learning schemes are being used to boost the ANN and SVM
performance with improved results. We explore better optimization approaches to devise a more suitable prediction technique. From the pool of optimization techniques given in Fig. 1.16, we propose a novel hybrid model for STLF by combining greater optimization ability of Artificial Bee Colony (ABC) algorithm with ANN. The ABC is a meta-heuristic optimization algorithm for nonlinear functions in multidimensional space. It is used as an alternative learning scheme to get optimized set of neuron connection weights of ANN. This formulation showed improved convergence rate without trapping into local minimum. Forecasting results obtained by this new approach have been presented and compared with other mature and competitive approaches, which confirms its applicability in forecasting domain. This formulation is tested on hourly data set obtained from IESO.

Artificial Cooperative Search (ACS) is a recently introduced algorithm for numerical optimization problems. Author of ACS algorithm examined its performance by 91 benchmark functions and compared with other mature computational intelligent algorithms. In order to verify the successes and analyse the optimization power of this algorithm for ELF, we propose a novel model of ACS based ANN for ELF. First ACS is evaluated against some well-known benchmarking functions and its performance is compared with other well matured evolutionary algorithms. Observing the better performance of ACS, we got motivation to employ ACS for weight optimization of ANN. This is a unique formulation of ACS based ANN model being presented. Furthermore, this hybrid model is applied for some real world classification and regression problems. The obtained results are then compared with other population based optimization algorithms for ANN. This empirical analysis affirms that ACS algorithm has better optimization ability when compared with other computational intelligent techniques being evaluated in this study.

The ABC algorithm is based on the exploration and exploitation behavior of honey bees. Its search mechanism is good in exploration but poor at exploitation of solution space. To improve these traits, several modified versions of ABC are introduced in literature. We analyze the evolving behavior of ABC algorithm and apply six modifications to original algorithm to improve its generalization and convergence ability. The individual modifications are applied to base algorithm in parallel as well as in series. Moreover, in order to further enhance the exploitation performance, genetic operators are coupled with the modified version of ABC, which makes it a unique formulation. This resulted in improved exploration and exploitation aspects of base algorithm; which is justified by benchmarking tests. Original and each modified version are then employed to train
artificial neural networks (ANN) with an objective to achieve better accuracy of load forecasts, an open issue of developing countries. These hybrid models are compared with each other and also with other competitive approaches used for ANN based model development. Results obtained from proposed formulation portray the efficacy in load forecasting domain achieving better accuracy. This integrated and modified version is tested on medium term forecasts. The performance of each modified version is compared with original and other well-known and mature models of forecasting. This experimentation is performed on daily and weekly peak data sets of both NTDCL and LESCO. The accuracy of modeling techniques is extremely dependent on quality of historical data. Since, the recording of data in Pakistan power systems was previously manual. This data contained abnormalities like missing values, outliers, and duplication of records. Observing all the aforementioned problems, we got motivation to devise such a ELF model that can perform well on noisy data of Pakistan power systems and can handle load affecting parameters of this region. Hence, we propose an integrated and customized model of short term ELF model. We include various components to the proposed system, such as: pre-treatment of historical data, analysis, transformation, cross validation, and over-fitting modules. The pre-processing module deals with outliers, anomalies, resolving the missing data problems in load series, curve smoothing, and data normalization. Further, this formulation with a set of modeling techniques is applied on LESCO hourly demand data.

Machine learning based statistical and artificial intelligence techniques are widely used for ELF. Among these, ANN and SVM emerge as competitive modeling approaches for ELF. To further improve the performance of these models, optimization techniques are being used to formulate hybrid ELF models. We have utilized the modern approaches to solve the issues with power planning in Pakistan. Hence, we contribute towards proposing machine learning based accurate model of ELF on Pakistan power system data set. Several forecasting models are formed using hybrid of optimization and model development techniques, which are ranked against their forecasting accuracy and performance. The SVM based models performed well and achieved 98.91% accuracy of forecasts. On the other hand, ANN based models showed comparable performance achieving 98.34% accuracy with added ability to avoid over-fitting, and efficiency with improved results. The evaluation and modeling of existing and new models is performed to find the best suitable combination of techniques for different types of forecasting. Short term, medium term, and long term forecasting models are evaluated on the basis on accuracy and efficiency. Several optimization techniques are combined with ANN’s and SVM’s to overcome the limitations of individual methods for better generalization and convergence ability.

This experimentation is performed on the Intel based Core i5 machine with 4GB of random access memory (RAM). The custom implementation of all the algorithms is produced on Java platform.
1.10 Thesis Organization and Conventions

This thesis is principally divided into four parts: introduction of load forecasting, existing techniques, evaluation of techniques, proposed models and unique contributions. The thesis organization is presented in Fig. 1.18. The thesis is organized as follows. Chapter 1, provides the introduction of ELF problem, its types, load affecting parameters, importance of ELF in power industry, motivation, objectives, problem statement, research methodology, and load shapes. Chapter 2, contains survey of existing literature. The
techniques are classified into different categories and shortcomings are highlighted. Literature review related to each of proposed models is also included in subsequent chapters. In chapter 3, a unique model for short term load forecasting is proposed. This model utilizes the ABC algorithm for ANN optimization. This formulation is tested on hourly data set obtained from IESO. The ACS is a recently introduced swarm based optimization algorithm. We have formulated its unique hybrid model with ANN. The generalization ability of this algorithm is evaluated on classification and forecasting problems in chapter 4. Further, The evolving behavior of ABC algorithm is discussed in chapter 5, six modifications are utilized in original algorithm to improve its generalization and convergence ability. These modifications are applied in parallel and as well as in series. This integrated and modified version is tested on medium term forecasts. The performance of each modified version is compared with original and other well-known and mature models of forecasting. The evaluation and modeling of existing and new models in presented in chapter 6. The motivation is to find the best suitable combination of techniques for different types of forecasting. Short term, medium term and long term forecasting models are evaluated on the basis on accuracy and efficiency. Several optimization techniques are combined with SVM’s and ANN’s to overcome the limitations of individual methods for better generalization and convergence ability. In chapter 7, proposed integrated and customized model of short term ELF model is presented. Various components of the system are discussed, such as: pre-treatment of historical data. This includes pre-processing of outliers or removing anomalies, resolving the missing data problems in load series, curve smoothing and data normalization. Moreover, correlation between input factors is presented in this chapter. Further, a set of techniques is applied on LESCO hourly demand data. Chapter 8, is the final chapter of thesis, it summaries the research contribution, future research directions and findings, and hence, closes the thesis.

In this thesis, the following conventions will be employed unless otherwise stated.

• The hourly data contains 24 sample load points per day. The sampling interval is 1 hour. The daily peak data contains 1 sample per day. The sampling period in this case is 1 day.

• The data set are obtained from NTDCL and LESCO. Canadian data is also employed for cross validation and to test the generalization ability of the proposed models. Since, the Pakistani load data is more difficult to deal with, therefore, LESCO and NTDCL data sets are used in most of the experiments.

• Mean Absolute Percentage Error (MAPE) will be employed to measure the error of the methods.

• Load and demand are alternatively used throughout the thesis. Actual load or demand are used to represent the actual readings of consumer demand data. Whereas, forecasted or output load means the predicted series of data.
1.11 Thesis Objectives

As power generation resources are limited and there is a crisis like situation in energy sector of Pakistan. Therefore, the foremost objective of this research is to empower Pakistan power systems with accurate load forecasting tools. As, different type of planning and decision making requires different levels of forecasting. Therefore, we aim to facilitate with short term, medium term and long term forecasts. The principal goal of this research is to design and develop a generalized and comprehensive load forecasting model for power systems which considers key load parameters and can be applied to geographical diverse areas. Through this research, we intend to solve power planning problem of local power industry. Some underlying objectives to achieve this goal are:

- To design and develop a load forecasting system for short term, medium term, and long term loads using analytical and intelligent modeling.
- To support local power industry through this research in most challenging problem of power planning. The forecasted data is useful for any electric utility in scheduling functions (unit commitment, hydro-thermal coordination, short-term maintenance, fuel allocation), day to day planning like load flow studies (contingency planning, load shedding) and infrastructure development (future planning, new power plants), etc.
- To detect and filter the errors in data through proper use of pre-processing techniques.
- To propose latest and better models with the incorporation of state of the art algorithms.

Another objective of this research is to study the most influencing factors effecting load variation in different seasons and in varying geographical localities. The development of ELF models in this case is a challenging problem. On the other side, traditional statistical and AI based techniques have been practically used for a long time, showing the forecasting accuracy to certain extent depending on the system at hand. These traditional methods need to be combined using some weighted multi-model forecasting techniques, which are more appropriate for our needs to show acceptable results.


1.12 Thesis Contributions

This dissertation has contributed towards completion of following milestones:

1. Development of unique hybrid models for electric load forecasting.
2. Integration of Artificial Bee Colony Algorithm (ABC) with Feed-Forward Neural Network (FFNN) model.
3. Integration of Artificial Cooperative Search (ACS) with FFNN model.
4. Performance analysis of different modifications applied on base Artificial Bee Colony Algorithm.
5. Development of hybrid model of mABC-GA-FFNN for load forecasting.
6. Improved accuracy of forecasts with proposed models.
8. Development of ELF models for short, medium, and long term forecasting.
10. Development of a generalized framework for ELF, consisting of data pre-processing, analysis, modeling, and optimization components.
Chapter 2

Literature Review

He who would search the pearls must dive below
John Dryden

2.1 Classification of Electric Load Forecasting Methods

Electric Load Forecasting (ELF) is an essential tool for any electric power supply utility in the world [2]. Accurate estimates of future demand help in planning generation and distribution of electricity [3]. The ELF has been a major area of research in the last decade, where more emphasis is laid on recent and more advanced approaches from both statistical and artificial intelligence (AI) domains [2]. Statistical models are generally considered rigid in nature and non-responsive to sudden variation in underlying input factors [3, 9]. The time series, regression based methods, radial basis functions (RBF), and support vector regression (SVR) [10, 11] fall into statistical category. The statistical methods developed for LF can be broadly categorized as, parametric and non-parametric [12]. Parametric models are developed by investigating qualitative correlation between the load and load affecting factors. ARMA models [13], multiple linear regression models [14], and exponential smoothing [15], are examples of parametric methods. Non-parametric methods directly build a load forecast model from historical data. As, in [16], a non-parametric regression based method is used for short term electric load forecasting.

On the other hand, artificial intelligence (AI) based approaches include expert systems, artificial neural network (ANN), and fuzzy logic [12, 17-19]. AI techniques use ANN for model development [20]. ANN models follow a heuristic approach to figure out the dependencies from historical data and have capability to model nonlinear behavior of underlying factors affecting variation in load curves [21]. Better predicting potential make them first choice for building load forecasting models. A wide range of ANN
models have been applied to load forecasting with a variety of network structures and training functions. Such models have been discussed and evaluated in [3, 12].

In contrast to the neural network (NN) theory, support vector machine (SVM) [22], is a statistical tool for classification and regression. It has greater ability of generalization and to avoid over-fit to data [23, 24]. Hence, computational power and accuracy make it an alternate choice for constructing load forecasting models. The ANN is known to suffer the slow convergence and trap into local minimum problems due to its gradient descent (GD) based learning process [21]. Similarly, finding the best parameters for SVM is another issue. There has been a tremendous development to overcome these issues, and a number of optimization techniques are used to train ANN’s and to find the optimum parameters of SVM [25]. Such techniques include levenberg-marquardt algorithm (LMA), simulated annealing (SA), particle swarm optimization (PSO), and resilient back-propagation (RPROP) method [2, 25]. A detailed comparison of these LF models has been presented in [26]. Recently much emphasis is being laid on the use of hybrid models [27]. The most recent survey of such models is reported by [2], covering state of the art methods of LF. ANN’s and SVM’s emerge as two competitive and successful techniques from AI and statistical domains respectively. AI models are discussed and evaluated in [28]. However, several LF models based on ANN and SVM are developed and their performance is compared in [29, 30].

In the gamut of hybrid approaches in ELF; genetic algorithm (GA) [31], PSO [32], ant colony optimization (ACO) [33], and artificial bee colony (ABC) [34], is being used to tune training parameters and weights of ANN [35], and SVM [22]. Applications of such optimization techniques to LF are discussed in [32, 36, 37].

### 2.2 Traditional Forecasting Techniques

The demand forecasting have been carried out by means of statistical or mathematical techniques. Such techniques include, time series methods, regression techniques, exponential smoothing, state space models, kalman filtering, and iterative re-weighted least-squares methods [38]. Time series methods are considered as oldest approaches applied to load forecasting. Time series methods also assume the linear relationship among load and influencing factors. These include: Box and Jenkins [39], moving average, auto-regressive moving average (ARMA), and auto-regressive integrated moving average (ARIMA) methods [40].

To understand inherent difficulty and limitation of these techniques, a comparative review of five widely used for load forecasting is presented in [38]. Regression techniques are used to model the relationship between consumer load and affecting parameters such as weather conditions and electricity prices. The mathematical model of regression based method is obtained by estimating the coefficients and assigning a polynomial degree to influencing factors. A regression based daily peak load forecasting method is proposed for Tokyo Electric Power Company [41]. To deal with seasonal, annual, and daily load
growth, and change a transformation function is estimated with the previous data points. Regression methods are easy to implement and their performance can easily be assessed [3]. The limitations and inherent problems with classical time series and regression methods are highlighted in [38, 42]. The drawbacks include problems in modeling the complex nonlinear relationship between load and the affecting parameters and inability to respond abrupt changes in these parameters [43].

2.3 Artificial Neural Networks (ANN)

Since, ANN’s are quite successful at modeling non-linearity and have characteristics such as being capable of generalizing, adaptability, self-organizing, real time operation, and fault tolerance. Still finding a suitable network structure and optimal weight values; makes design of ANN’s, a difficult optimization problem. The traditional approach for training ANN through error back propagation and weight adjustment by gradient descent (GD) based algorithm. Due to dependence of error surface shape and initial values of connection weights and parameters [35], GD based algorithms are trapped in local minima. As, this approach involves taking differentiation of error function, which is another issue due to which this algorithm cannot reach global minima. Recently different approaches are used; not only to overcome such problems, but also to train ANN for optimized set of connection weights and to obtain appropriate network topology. Xin Yao [20], reviewed different techniques to evolve ANN connection weights, learning schemes, network structures, and input feature selection.

2.4 Support Vector Machines (SVM)

The SVM is a statistical model, evolved from sound theory to experimentation. SVM’s have greater generalization ability and are less prone to over-fitting. However, its generalization ability highly dependent on the optimal selection of parameters. Therefore, to obtain better performance of forecasting with SVM; selection of appropriate settings to its parameters is critically important [44]. While, ANN follow a heuristic path of experimentation to theory. They have known problems of slow convergence and to stuck in local minima [21]. A lot of effort is reported to overcome the weaknesses of ANN model and to choose appropriate parameters for SVM by use of different optimization methods [25]. These optimization techniques are being used to develop hybrid LF models developed [2]. Since, the individual techniques have their own strengths and weaknesses, it becomes difficult to find suitable technique for forecasting. The ensemble models are formulated by combining two or more individual techniques. In this way, overall performance can be enhanced. The hybrid models for LF are widely discussed in literature [3, 9]. Whereas, the most recent survey has revealed the potential of ANN and SVM based hybrid models of LF [30]. Several other comparisons of state of the art techniques have emphasized on hybridization of forecasting methods to get more accurate results [26, 28, 29].
2.5 Optimization Techniques

Optimization techniques [45], play an important role in optimizing ANN structure, network layer weights, network size, number of neurons, and training functions [33]. The more recently developed evolutionary and nature inspired techniques for optimization have become popular methods for optimizing model structures and parameters. Out of these, ABC is used to optimize multi-variable and multi-model continuous functions [46].

Both mathematical and AI based optimization techniques are widely used and discussed in literature. A survey of optimization techniques is presented in [47], in which authors discussed a wide variety of techniques in relation to power systems. A bunch of load forecasting models developed using such optimization techniques are described in [3]. Wei-Chiang Hong [37], worked out different load forecasting models by combining swarm based optimization techniques with SVM; mainly chaotic ant swarm optimization, chaotic GA, chaotic PSO, and chaotic bee colony optimization (BCO) algorithms [48]. Author compared results with ANN and regression based models for forecasting performance. Based on a series of experiments, the author recommended SVR with hybrid chaotic evolutionary algorithms as a superior alternative to improve the load forecasting accuracy. In the gamut of hybrid approaches in load forecasting; for instance differential evolutionary algorithms are used in [49], to choose the appropriate parameters of SVM while developing hybrid LF model. Combination of GA and ACO is used with multi-layer perceptron (MLP) for hourly load prediction [33]. PSO is also a popular technique for ANN optimization, its applications in LF are reported in [32, 36].

The optimization techniques used in this study are obtained from different domains, for instance, LMA is a statistical learning scheme [50]. LMA is used to overcome the slow convergence of ANN for the estimation of daily peak demand [51]. An average accuracy of 97% was obtained by different models. On the basis of accuracy against different criteria’s, LMA based back propagation ANN model is confirmed as best technique for forecasting of daily peak loads [51]. The SA is an optimization algorithm derived from metallurgy. The SA finds the global optima by simulating the cooling process and energy states of metal under annealing process [52]. The SA is combined with SVM to solve the electric LF problem [53]. The SVM-SA based model achieved 98.25% of accuracy with better generalization performance compared to ARIMA and general regression NN (GRNN) models [53]. Adaptive PSO a computation intelligence algorithm, which is used to find optimal parameters of SVM for LF [36, 54]. To keep a balance among exploration and exploitation aspects of swarm; a new scheme for inertia weight adjustment is introduced [36]. This hybrid mode produced an accuracy of 98.32%; whereas, the Back-propagation neural networks (BPNN) and regression models produced 94.93% and 96.42% accurate results, respectively [36]. The resilient back-propagation (RPROP) method performs a direct adaptation of the weight step depending upon the local gradient information [55]. The RPROP trained ANN model shown an average accuracy of 97.40% for 7 days ahead electrical peak LF [55]. Similarly, firefly algorithm (FFA) and
artificial bee colony (ABC) algorithm are recently introduced optimization techniques [56, 57]. To overcome premature convergence, local optimum, and to get better performance in function optimization ABC is employed as learning scheme of SVM for accurate results of LF [34]. This formulation has achieved 97.62% accuracy; better than ARIMA and SA based models. Whereas, FFA is used for parameter optimization of SVM based forecasting model, achieving higher accuracy of forecasts with 98.42% accurate results [58]. Likewise, a recent study has evaluated different optimization algorithms for power demand forecasting in a smart grid environment [59].

2.6 Generalization Problem of LF Models

What emerges from our study is that number of methods and techniques have been developed and reported in literature but there is a lack of an ample solution that handles all factors, covers forecast types and is adaptive enough to apply on different geographical locations. Until now, there has not been one generalized load forecasting technique which applies to geographically diverse sites. Methods cited in the literature are tested basically using data from one site only. Large errors start to appear when these algorithms are tested over data from other sites. It is highly desirable, therefore, to develop a technique for load forecasting which applies to multiple sites in different seasons.

Different techniques have been applied to load forecasting. After surveying LF approaches, we can observe a clear trend toward new, stochastic, and dynamic forecasting techniques to obtain more generalized models while avoiding over-fitting. It seems that a lot of current research effort is focused on fuzzy logic, expert systems, support vector machines and particularly neural networks. There is also a clear move towards hybrid methods, which combine two or more of these techniques for better convergence and generalization. Majorly combining statistical, nature inspired or evolutionary techniques with ANN’s and SVM’s.

There are many challenges with individual techniques, existing models, and power demand data sets. The problems in data sets call for appropriate usage of pre-processing methods. This section is providing brief about these.

2.6.1 Noisy Data

Performance of LF models is directly proportional to quality of data [3]. Noise in the load series or affecting parameters data can disturb the whole demand patterns and hence, results in lower accuracy. Noise in the data may be of different forms, such as: missing values, outliers, and duplicated values. Pre-processing techniques help smoothing the load curves. Data treatment is discussed in Chapter 7.
2.6.2 Relationship between Dependent and Independent Variables

LF models estimate a regression function/equation to represent the relationship between dependent (Load) and independent variables (e.g., weather conditions). It requires detailed understanding of the problem and features must be selected carefully. The poor selection of independent features and improper regression function will reduce the forecasting accuracy. For this, proper use of statistical methods is recommended to analyse the co-relation and co-variance among input and output variables. Such methods to find the relationship among dependent and independent factors are discussed in Chapter 7.

2.6.3 Data-Model Coupling

What emerges from our study is that number of methods and techniques have been developed and reported in literature but there is a lack of an ample solution that handles all factors, covers forecast types and is adaptive enough to apply on different geographical locations. Until now, there has not been one generalized load forecasting technique which applies to geographically diverse sites. Methods cited in the literature are tested basically using data from one site only. Large errors start to appear when these algorithms are tested over data from other sites. It is highly desirable, therefore, to develop a technique for load forecasting which applies to multiple sites in different seasons. There are no rules as to when a particular technique is more or less suitable for forecasting; even a best suitable technique which has shown excellent results cannot be assured to have the same results for different meteorological or geographical localities.

2.6.4 Applicability of Existing LF Models

In continuation of the Sec.2.6.3, the techniques presented in the literature, propose forecasting solutions for a specific problem with customized requirements and different data sets. The features in each data set are not always the same. Hence, to compare the models, it requires the same data set with same features and setting of the models. Therefore, a direct comparison of the the existing techniques is not possible and is quite meaningless. Several existing and new approaches are compared in this study, but on the same criteria and same data sets. Performance comparison of various models is presented in Chapter 6.

2.6.5 Over-Fitting

In regression problems, whole data is classified into two groups: train set and test set. The training process aim at learning the input-output relationships and hence, modify the model parameters/structure. The trained model minimizes the forecasting error by adopting the trends in train data. Further, the test data is presented to model and predictions are made. If the model error on train data is low and error on test data is high, this is called over-fitting of the model. A regression function or a model is said to lesser generalized, if it could not show the same accuracy of forecasts on the unseen (test) data. Over-fitting, however, can be controlled by many factors; such as: size of training data, model structure, and cross-validation. The larger the train set, the less chances
of over-fitting. The model parameters should be carefully selected to achieve better mapping of inputs to outputs. In order to achieve better generalization of regression models, they should be properly validated. For this, K-fold cross-validation method is used.

2.6.6 Cross Validation

The cross-validation also called rotation estimation is a statistical method to estimate the performance of a predictive model on independent data. This is a model validation technique to assess its generalization ability of models. To perform the cross-validation, data is split into three groups: train, test, and validate. The data may be linearly divided or randomly selected for three subsets. We have used the random and sequential cross-validation techniques for data division.

2.7 Conclusion

After surveying LF approaches, we can observe a clear trend toward new, stochastic, and dynamic forecasting techniques. It seems that, a lot of current research effort is focused on fuzzy logic, expert systems, support vector machines, and particularly neural networks. There is also a clear move towards hybrid methods; which combine two or more of these techniques. Majorly combining optimization or evolutionary techniques with ANN and SVM. What emerges from our study is that number of methods and techniques have been developed and reported in literature but there is a lack of an ample solution that handles all factors, covers forecast types and is adaptive enough to apply on different geographical locations. Until now, there has not been one generalized load forecasting technique which applies to geographically diverse sites. Methods cited in the literature are tested basically using data from one site only. Large errors start to appear when these algorithms are tested over data from other sites. It is highly desirable, therefore, to develop a technique for load forecasting which applies to multiple sites in different seasons. Concluding, we can say that AI techniques, like all other approximation techniques, have relative advantages and disadvantages. There are no rules as to when a particular technique is more or less suitable for forecasting; even a best suitable technique which has shown excellent results cannot be assured to have the same results for different meteorological or geographical localities. It depends on our priorities, effectiveness of different input factors and geographical conditions. Though, load forecasting bench-marking and best practices of the industry should be incorporated, the tools and methods being used and their relative strengths and weaknesses should be considered while designing a new load forecasting system for our customized requirements. Finding appropriate solutions to above mentioned problems will be prime focus of our proposed work presented in subsequent chapters.
Chapter 3

Artificial Bee Colony Algorithm based Load Forecasting Model

Trying to predict the future is like trying to drive down a country road at night with no lights while looking out the back window.

Peter F. Drucker

3.1 Introduction

Accurate future electric demand estimation plays an important role in decision making for generation and transmission planning. It is essential for economic and reliable operations of electric utility, ensuring better synchronization between generating and distributing agencies [9, 60]. Especially in the developed countries, where electricity generation is deregulated into organized markets. In developing countries for instance, India and Pakistan, where generation capacity is not up-to the mark, voltage instability is often the case. Adopting better forecasting practices can avoid such over and under generation situations. Accurate load forecasting (LF) helps estimating amount of energy that must be delivered to a variety of customers located at geographical diverse sites on different time horizons [27].

Electric LF depends on several load affecting parameters, for instance, weather conditions, time of demand, number of customers. These influencing parameters may differ for different geographical regions. Varying number of parameters, nature of underlying load affecting features, priorities, and custom requirements call for a customized demand forecasting model development [3]. This results in high coupling between LF models being developed and data sets under consideration. Thus, one technique that performs well on one data set doesn’t confirm to show the same results on other data sets [12].
Now advanced and more effective forecasting models are being used in area of LF. The invention of new, powerful, and intelligent modeling techniques prompted researchers of forecasting domain to incorporate these models. Such variety of methods include both mathematical and artificial intelligence techniques. In the recent past emphasis is being laid on evolving hybrid methodologies [3, 9]. Combination of different techniques is being realized to enhance the performance and to overcome the weakness of existing models. From the pool of forecasting models, Artificial Neural Networks (ANN) emerged as most successful modeling techniques for LF. However, it is still known to suffer the slow convergence problem and falling into local minima. There has been a significant progress to overcome these limitations and different optimization techniques are being used to train ANN. Optimization techniques [45], play an important role in optimizing ANN structure, network layer weights, network size, number of neurons, and training functions [33]. The more recently developed evolutionary and nature inspired techniques for optimization have become popular methods for optimizing model structures and parameters. Out of these, Artificial Bee Colony (ABC) is used to optimize multi-variable and multi-model continuous functions [46].

Our hypothesis is to present ABC as a better approach for ANN optimization. In order to prove our hypothesis, in this experimentation, we propose an ABC based hybrid approach for model development of short term load forecasting (STLF). A customized ANN is used as modeling technique and ABC is employed as learning scheme. Optimized weights are obtained in a recursive manner by utilizing artificial bees to search best training parameters. Objective of this work is to improve the accuracy of forecasts by utilizing the inbuilt optimization and convergence capability of ABC. The results are also compared to other modern techniques of load forecasting like particle swarm optimization (PSO) [61] and genetic algorithm (GA) [62].

3.2 Related Work

The methods developed for LF can be broadly categorized as parametric, non-parametric, and artificial intelligence (AI) based approaches [12]. Parametric models are developed by investigating qualitative correlation between the load and load affecting factors. ARMA models [13], multiple linear regression models [14], and exponential smoothing [15], are examples of parametric methods. Non-parametric methods directly build a load forecast model from historical data. As in [16], a nonparametric regression based method is used for STLF. AI techniques use ANN for model development [20]. ANN models follow a heuristic approach to figure out the dependencies from historical data and have capability to model nonlinear behavior of underlying factors affecting variation in load curves. Better predicting potential make them first choice for building load forecasting models.

A wide range of ANN models have been applied to load forecasting with a variety of network structures and training functions. Such models have been discussed and evaluated in [3, 12]. In contrast to Neural Network (NN) theory, Support Vector Machine (SVM) [22], is a statistical tool for classification and regression. It has greater ability
of generalization and to avoid over-fit to data [23]. Hence, computational power and accuracy make it an alternate choice for constructing load forecasting models.

Since, ANN’s are quite successful at modeling non-linearity and have characteristics such as: being capable of generalizing, adaptability, self-organizing, real time operation, and fault tolerance. Still, finding a suitable network structure and optimal weight values; makes design of ANN’s, a difficult optimization problem. The traditional approach for training ANN through error back propagation and weight adjustment by gradient descent (GD) based algorithm. Due to dependence of error surface shape and initial values of connection weights and parameters [35], GD based algorithms are trapped in local minima. As this approach involves taking differentiation of error function, which is another issue due to which this algorithm cannot reach global minima. Recently different approaches are used; not only to overcome such problems, but also to train ANN for optimized set of connection weights and to obtain appropriate network topology.

Xin Yao [20], reviewed different techniques to evolve ANN connection weights, learning schemes, network structures, and input feature selection.

Both mathematical and AI based optimization techniques are widely used and discussed in literature. A survey of optimization techniques is presented in [47], where authors discussed a wide variety of techniques in relation to power systems. A bunch of load forecasting models developed using such optimization techniques are described in [3]. Wei-Chiang Hong [37], worked out different load forecasting models by combining swarm based optimization techniques with SVM; mainly chaotic ant swarm optimization, chaotic GA, chaotic PSO, and chaotic bee colony optimization (BCO) algorithms [48]. Author compared results with ANN and regression based models for forecasting performance. Based on a series of experiments, the author recommended SVR with hybrid chaotic evolutionary algorithms as a superior alternative to improve the load forecasting accuracy. In the gamut of hybrid approaches in load forecasting; for instance differential evolutionary algorithms are used in [49], to choose the appropriate parameters of SVM while developing hybrid LF model. Combination of GA and ant colony optimization (ACO) is used with multi-layer perceptron (MLP) for hourly load prediction [33]. PSO is also a popular techniques for ANN optimization, its application in LF are reported in [32, 36].

The ABC is relatively simple and is proved very effective to produce good results at a low computational cost. Therefore, this algorithm is successfully applied in different application areas. Originally, ABC optimization was proposed for solving numerical problems. Therefore, many studies are focused on evaluating performance of ABC on the widely used set of numerical benchmark test functions and to compare it with that of well-known evolutionary algorithms such as: GA, PSO, Differential Evolution (DE), and ACO. In [63], performance of ABC is compared against Bees and DE algorithms on eight well-known benchmark problems. For optimizing multi-variable functions, authors in [64], presented a comparative study of ABC with GA, PSO, and particle swarm
inspired evolutionary algorithm. In [65], authors compared the results of ABC performance assessment study of foraging, including ABC and evolutionary algorithms. A more recent and comprehensive survey is reported in [66], about ABC algorithm, its applications, modified versions, comparisons, and hybrid models.

The ANN’s are is proven successful in modeling LF problems. Still, aforementioned problems with GD based training functions and availability of alternative optimization algorithms motivated us to explore modern approaches to solve the accurate load forecasting problem. Part of this work is reported in [67], where nonlinear auto-regressive exogenous model (NARX) based feed-forward neural network (FFNN), SVR, and GD based ANN models are investigated to solve long term industrial load forecasting reaching 98% accuracy. As ABC is quite successfully applied to different application areas, its simplicity, efficiency, and accuracy of results also motivated authors to extend the use of this algorithm to energy demand forecasting domain. Due to the fact that different techniques proposed in literature, report varying degree of accuracy on different data sets [12]. Hence, a bunch of techniques should be evaluated on the same data sets with similar performance criteria. Therefore, our proposed approach is compared with other technique on same data sets with similar settings.

3.3 ANN Training

The proposed approach is based on the combination of FFNN with ABC. The FFNN is trained by ABC optimization algorithm to find optimal neuron connection weights. The following sub-sections briefly discuss the ANN model and ABC algorithm, followed by PSO and GA.

3.3.1 ANN model

The ANN is an information processing network inspired by human nervous system, i.e. to learn and build rules through experience [21, 35]. It is an adjustable parameterized system and network behavior can be tuned by network structure, interconnection schemes, learning rules, and output transfer functions. During training; neuron weights are optimized until network obtains a satisfactory level of accuracy. Once trained; network can accept new inputs to predict the output. Based upon their predictive performance and capability to handle nonlinear and dynamic systems ANN is considered as a powerful tool for time series prediction and modeling tasks.
3.3.2 ABC overview

The ABC algorithm [46], is inspired from intelligent foraging behavior of honey bee swarm searching for food. In bee colony algorithm position of each food source corresponds to a possible solution for the optimization problem, the amount of nectar of a food source depicts the fitness value (quality) of the associated solution. The total bees in colony can be categorized into three groups: employed, onlooker and scouts. The employed and onlookers are equal in number in a hive and correspond to equal number of solutions in population [57].

3.3.3 PSO overview

The PSO is an adaptive algorithm based on social-psychological behavior of swarm particles. These particles fly around in n-dimensional space to discover better solutions. They have two characteristics associated, velocity update and position update. At each iteration, new velocity and position for each particle is calculated [61].

3.3.4 GA overview

The GA is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetic. It is a multi agent search technique. The real power of GA comes from its selection mechanism, parallel processing and genetic operators. The genetic operators that form the basis of simulated genetic systems are crossover, fitness evaluation, reproduction and mutation [62]. Mutation operator can affect each solution with a given probability and is applied to maintain genetic diversity. Whereas, crossover operator is also called recombination, and is a process of combining existing solutions to produce new chromosomes/solutions that inherit some or all of their attributes from one or both of their parents.

3.4 STLF Model Formulation

In this section, formulation of a new customized model for STLF is discussed; parameters and settings for ANN, GA, PSO, and ABC are provided. ANN training process is shown in Fig. 3.1. the aforementioned techniques are incorporated as learning schemes. Inputs are passed to ANN, whose connection weights are obtained by each of learning schemes and results are evaluated over a number of iterations. The settings for ANN and learning schemes are given in following sections. Further, the algorithm used to implement the proposed ABC based ANN approach is described step-by-step. This method incorporates employed, onlooker and scout bees of ABC. Here, the whole process is divided into steps and equations.

3.4.1 Input Factors

There is a complex and nonlinear relationship between the electricity load and its influencing factors. Electric demand on short-term horizon is affected by calendar inputs, weather conditions, and energy price. The calendar events are time of day, day of week, and holidays; weather parameters include temperature and humidity. The inputs
to the neural network should be carefully selected as they affect the forecasting ability of
the network. The inputs are chosen based on their correlation with the output load. The
inputs of the neural network for this study include; calendar events, weather conditions,
day type, and previous hour, day, and week demand profiles.

3.4.2 ANN Settings

A three layered feed forward neural network is used in the experimentation. The inputs
are the variables affecting the variations in the load and the output is next hour load.
The number of neurons in hidden layer affects its learning and generalization ability so
they must be carefully selected. Hidden layer neurons are selected by taking the average
of number of inputs and number of outputs and then gradually increasing the neurons
avoiding under-fitting and over-fitting at the same time until optimal results are found.

The ANN used in this study consists of 9 input ($I$), 8 hidden ($H$) and 1 output ($O$)
neurons. Moreover, two additional bias ($b_1, b_2$) neurons are included, one at hidden layer
and other at output layer, resulting in total of 91 neuron connections. As, neurons of
each layer in feed-forward neural network (FFNN) are connected to all of the neurons
in the next layer. The neurons connections are calculated by multiplying neurons of
both layers. In our ANN model, 10 (9 input + 1 bias) neurons are connected to 8
hidden neurons resulting in 80 neuron connections. Further, the 9 (8 hidden + 1 bias)
neurons are connected to single output neuron. This can be formulated in an expression
as $N = (I+b_1)*H+(H+b_2)*O+b_1+b_2$. Hence, it becomes $91 = (9+1)*8+(8+1)*1+2$.

The sigmoid transfer function is used to get the neuron output. Mean Squared Error
(MSE) is used in order to evaluate the NN performance. The weights are initialized from
a uniform distribution in range $[-1, 1]$.

3.4.3 PSO Settings

This study employs the standard version of PSO with recommended values of two con-
stants $C_1$ and $C_2$, limiting factor and $\phi$ to update particle velocity and location [68].
Here, both constants $C_1$ and $C_2$ have same value i.e. $C_1 = C_2 = 1.494$, limiting factor
= 0.734 and $\phi = 4.1$. These values provide better compromise in global and local search
convergence. The population count of the swarm is 30.
3.4.4 GA Settings

The standard GA with recommended settings is applied in this study. Here, population size for GA is 5000, dimensions are equivalent to number of neuron weights; which is 91, recombination rate is set to 0.7 and mutation rate is 0.05.

3.4.5 ABC Settings

The number of food sources is 28 which is also the population count of honey bees. The count for each of employed and onlooker bees is 14. The dimension of the search space is equal to the 91 weight connections. The upper and lower bounds are $[-1, 1]$.

3.4.6 ABC Training Process

The training process of ABC comprises several steps which are discussed here, this sequence is elaborated in Algorithm 1, and shown in Fig. 3.2.

Initialization

At the initialization stage of ABC algorithm, an initial population of $N$ food sources is generated randomly within the range of variable boundaries. The Eq. 3.1. is used to initialize the population. Here, total population size is 28.

\[ x_{ij} = x_{j}^{\min} + \text{rand}(0, 1)(x_{j}^{\max} - x_{j}^{\min}) \]  

(3.1)

Where $i$ represent the $i_{th}$ food source number and $j$ is the optimization variable associated with $i_{th}$ food source.

Fitness Evaluation

Now, nectar amount in each food source is calculated and fitness of food sources is evaluated according to Eq. 3.2.

\[ \text{fitness}_i = \begin{cases} \frac{1}{1+|f_i|} & \text{if } f_i \geq 0 \\ 1 + \text{abs}(f_i) & \text{if } f_i < 0 \end{cases} \]  

(3.2)

Where $f_i$ is the cost associated with solution $x_i$, this function can be directly used as cost function in maximization problems.
**Neighbor Selection**

Now, each employed bee searches a new candidate food source \( v_i \) in the neighborhood of food sources already in memory. New food source is compared based upon the visual information of bees and evaluated according to Eq. 3.3. Here greedy selection is applied to select a better food source as new candidate food source.

\[
v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj})
\]  

(3.3)

Where, \( j \) is a random optimization variable in the range \([1, D]\) and \( k \) is a randomly selected food source different from \( i \), which is a uniformly distributed real random number in the range \([-1, 1]\). Here \( D \) is non-negative number.

**Solution Weighting**

Probability of selecting a solution in population is proportional to nectar amount in the source or fitness value of source in population. By using roulette wheel selection method [69], each onlooker bee selects a food source with higher probability, using Eq. 3.4. Then, it generates a candidate solution according to Eq. 3.3.

\[
p_i = \frac{\text{fitness}_i}{\sum_{i=1}^{N} \text{fitness}_i}
\]  

(3.4)

where \( p_i \) is probability of \( i_{th} \) food source. Here, solutions are evaluated and food source of which nectar is abandoned by bees or a better candidate food source is found and is replaced with a better one as the new food source according to greedy selection process.

**Solution Evaluation**

Further, the better solution which is also the combination of 91 neuron weights is passed to FFNN for processing. In FFNN several neurons are combined in three layers, where input signals and connection weights are processed in each neuron and passed to other neurons in next layer. A single neuron computes the weighted inputs using the formula given in Eq. 3.5. [21].

\[
y_j = \sum_{i=1}^{n} x_i w_{ij}
\]  

(3.5)

Where \( x_i \) is the input vector and \( w_{ij} \) are corresponding weights for this input vector, \( y_j \) is the response of neuron based upon transfer function.

For continuous output, neuron uses sigmoid transfer function as given in Eq. 3.6.

\[
y_j = \frac{1}{1 + e^{-x_j}}
\]  

(3.6)

Network computes the objective function also called the error \( E \), by taking MSE which is expressed by Eq. 3.7.

\[
E = \frac{1}{n} \sum_{i=1}^{n} (y_i - a_i)^2
\]  

(3.7)
Where $y_i$ is output obtained at neuron $i$ and $a_i$ is the desired output. On this step network passes its objective function back to learning scheme to further evaluate its solutions.

**Best Solution**

After each candidate solution is produced by artificial bees and then evaluated by FFNN, its fitness is always compared with old best solution. If the new solution is better or has equal fitness value to old solution; old best solution is replaced by new solution in memory. However, current best solution is kept in memory during this process.

### 3.4.7 Stopping Criteria

ABC algorithm employs the above described steps in a repeated manner through a predetermined number of iterations or termination criterion is met. Upon termination of algorithm, fitness value and position of optimal food source act as decision parameters. Algorithm stop finding new solution when the number of iterations reaches 500, or if the MSE decreases below $1e^{-5}$. The aforementioned criteria is same for PSO and GA.

The process from initialization step to termination is repeated for employed bees until they continue to find better solution, then onlooker bees are called to regenerate solutions based upon probabilities obtained by using Eq. 3.4. If onlooker bees are also unable to find a global minimum, or current best solution is exhausted then scout bees come into play by generating a random solution by Eq. 3.1. This process guarantees to avoid local minimum. Each type of bees passes their candidate solution to FFNN to calculate objective value, which is then returned back for possible selection or rejection of solution.

As illustrated in Algorithm 1, FFNN training through ABC is demonstrated in Fig. 3.2. Where, each of “Employed Bees” and “Onlooker Bees” block performs the step 1 of the sequence and has three output lines. One is connected to FFNN block to perform step 2 to step 5, and hence returns to originating block to perform step 7. This process is repeated until all the bees of same block are exhausted. The self-loop iterates for the number of bees. If the all the bees of one block are exhausted and objective function of FFNN is still not satisfied, then it moves to subsequent block for further exploration of solution search space.

### 3.5 Experimental Results

#### 3.5.1 Data set

This experimentation is performed on 10 years (2002-2012) hourly electricity load demand data sets obtained from Independent Electricity System Operator (IESO) for Ontario State. This data is publicly available and is highly accurate. Based upon high correlation, different input factors like calendar events, demand of last hour, and previous day, weather conditions, and hourly energy price are considered as input parameters in this study. In order to achieve the desired objective of accurate STLF model, it is necessary to compare all three models on same criteria, for that reason data and features
Chapter 3. *ABC-FFNN model*

1. **Initialization phase;**
   Load Sample Values of the input data;
   Initialize the optimization parameter;
   Generate the initial population $X_i, i = 1...N$;
   Evaluate the fitness value ($f_i$) of the population;
   Set iteration count to 1;
2. **Employed bees phase;**
   \[ \text{forall the employed bees do} \]
   - Apply greedy selection process;
   - Produce new solution $V_i$ by using (Eq. 3.3.);
   - Pass new solution to ANN;
   - ANN calculates the fitness value $f_i$;
   - \[ \text{if } f_i \text{ is better than current fitness then} \]
     - Current fitness = $f_i$;
   - \[ \text{else} \]
     - Discard new solution;
   - \[ \text{end} \]
   - Increase trial count;
   - Calculate the probability values $p_i$ for the solutions ($x_i$) by (Eq. 3.4.);
3. **Onlooker bees phase;**
   \[ \text{forall the onlooker bees do} \]
   - Select a solution $x_i$ depending on $p_i$;
   - Apply greedy selection process;
   - Produce new solution $v_i$;
   - Pass new solution to ANN;
   - ANN calculates the fitness value $f_i$;
   - \[ \text{if } f_i \text{ is better than current fitness then} \]
     - Current fitness = $f_i$;
   - \[ \text{else} \]
     - Discard new solution;
   - \[ \text{end} \]
   - Increase trial count;
4. **Scout bees phase;**
   \[ \text{while iteration <= Maximum iteration count do} \]
   - Replace abandoned solution for the scout bees by a new solution produced randomly by equation Eq. 3.1.;
   - Memorize the best solution obtained so far;
   - iteration = iteration+1;
\[ \text{end} \]

Where $x_i$: a solution, $f_i$: the fitness value of $x_i$, indicates a neighbor solution of $x_i$, $p_i$ is the probability value of $x_i$.

**Algorithm 1**: Customized Algorithm for iterative FFNN Training by ABC Algorithm

sets are identical for all models. For training the neural network models, 70% of data is utilized and 30% is used for testing and validation by further dividing the data into two equal parts. Further the attribute data set is scaled to fit in a specific range, this is process of normalization and there are many techniques for it. Here data is normalized.
Figure 3.2: FFNN Model Training by ABC Algorithm

into range $[0, 1]$ before training using min-max normalization method. This is defined by the formula given in Eq. 3.8.

$$x' = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$  \hspace{1cm} (3.8)

where $x'$ is the normalized value, $x_{\text{min}}$ is the minimum and $x_{\text{max}}$ is the maximum value for variable $x$.

3.5.2 Performance Evaluation

In this study, model accuracy is ranked on the basis of Mean Absolute Percentage Error (MAPE) which is the most common method to evaluate LF accuracy. It expresses accuracy as a percentage. This is defined by the formula as given in Eq. 3.9.

$$M = \frac{100}{N} \sum_{t=1}^{N} \left| \frac{A_t - F_t}{A_t} \right|$$  \hspace{1cm} (3.9)

Where $A_t$ is actual load at time period $t$ and $F_t$ is forecasted demand value for same time period, $N$ is number of samples.
3.5.3 Forecasting Results and Discussion

In this study, maximum iteration count for one run of each technique is 500. For each algorithm 30 independent runs are made to get mean value of MAPE and MSE. By this customized and extended formulation of ABC based NN, a greater accuracy of results is achieved with 2.3% of average MAPE. Average MAPE is obtained by the Eq. 3.10.

\[
\text{AvgM} = (\text{TrainM} \times 0.70) + (\text{TestM} \times 0.15) + (\text{ValM} \times 0.15)
\]  

(3.10)

As ANN’s follows heuristic way of learning, the network improves over number of iterations; consequently, results are improved. Improvement in accuracy of results by reducing MAPE over 500 iterations can be witnessed in Fig. 3.3. The GA, PSO, and ABC are compared with respect to convergence rate towards lowest error. During initial iterations, MAPE is reduced swiftly for PSO and ABC. For ABC it dropped from 7.54 to 2.93 then MAPE of ABC based model started to converge slowly and it reached around 2.40 after first 250 cycles and then a slow improvement can be seen in next 250 iterations and ultimately it reached 2.30 after 500th iterations. For PSO, convergence rate is initially high but it could not show any improvement after first 90 iterations and remained at 3.0 over the subsequent iterations. The PSO reached its global minima in a fast manner but it is unable to converge further. Similarly, the GA performed poorly with higher MAPE. First, it started to converge and achieved its lowest possible values in initial 200 iterations. Then it remained consistent for the subsequent iterations resulting in higher MAPE value. This is due to inherent problems with GA’s such as: premature convergence and slow finishing [70].

The ANN computes the objective function or network error E, by taking MSE that is expressed by Eq. 3.7. The comparison of all three optimization techniques for their ability to reach lowest error is presented in Fig. 3.4. The MSE of our proposed hybrid approach of ABC based neural network is dropped quickly in initial iterations then it started to converge and over the maximum number of iterations to reach lowest point. Whereas, MSE obtained by GA based ANN models remains higher over the complete loop of iterations. However, PSO based neural network is showing good results and its performance is comparable to ABC based model. The ABC model in this case outperformed others by faster rate of convergence and achieving lowest MSE in lesser iterations. The observed behavior of PSO and GA strengthen our hypothesis to present ABC as alternative and better approach to train ANN model and to obtain better accuracy of forecasts.

Three days hourly profile of actual and predicted demand values is shown in Fig. 3.5. It provides evidence that results obtained through ABC-FFNN hybrid model closely fits the actual load values. It is clear from the graphs that ABC algorithm outperformed other techniques while training ANN’s model for STLF by attaining lowest MSE, achieving lower MAPE and producing more accurate results. Table 3.1. shows input parameters
and comparison of actual and predicted demand values over a few data samples.

Accuracy comparison in terms of MSE and MAPE obtained from aforementioned techniques is shown in Table 3.2 and 3.3. In Table 3.2, average, best, and worst MAPE of
Table 3.1: Input data set with Forecasts (Sample)

<table>
<thead>
<tr>
<th>Date</th>
<th>Hour</th>
<th>LastHour</th>
<th>LastDay</th>
<th>Weekday</th>
<th>Price Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>5/16/2012</td>
<td>4</td>
<td>13614</td>
<td>13893</td>
<td>5</td>
<td>20.18</td>
<td>13711</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>5</td>
<td>13711</td>
<td>14272</td>
<td>5</td>
<td>19.37</td>
<td>14010</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>6</td>
<td>14010</td>
<td>15401</td>
<td>5</td>
<td>18.45</td>
<td>14838</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>7</td>
<td>14838</td>
<td>17218</td>
<td>5</td>
<td>23.38</td>
<td>16942</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>8</td>
<td>16942</td>
<td>17804</td>
<td>5</td>
<td>22.65</td>
<td>17856</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>9</td>
<td>17856</td>
<td>17990</td>
<td>5</td>
<td>23.67</td>
<td>18079</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>10</td>
<td>18079</td>
<td>18077</td>
<td>5</td>
<td>23.1</td>
<td>18309</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>11</td>
<td>18309</td>
<td>17991</td>
<td>5</td>
<td>21.56</td>
<td>18373</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>12</td>
<td>18373</td>
<td>17892</td>
<td>5</td>
<td>21.47</td>
<td>18229</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>13</td>
<td>18229</td>
<td>17738</td>
<td>5</td>
<td>21.98</td>
<td>18093</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>14</td>
<td>18093</td>
<td>17621</td>
<td>5</td>
<td>20.77</td>
<td>18024</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>15</td>
<td>18024</td>
<td>17545</td>
<td>5</td>
<td>20.83</td>
<td>18079</td>
</tr>
<tr>
<td>5/16/2012</td>
<td>16</td>
<td>18079</td>
<td>17689</td>
<td>5</td>
<td>20.17</td>
<td>18217</td>
</tr>
</tbody>
</table>

each technique is tabulated. As all techniques are executed 30 times with 500 iterations of each run, therefore, spread of MAPE is also presented by the technique of standard deviation. This is evident from the figures given in Table 3.2, that ABC based ANN is showing better accuracy. In Table 3.3, average MSE of ANN over 30 runs is shown, ABC is outperforming here again with lesser error.

![Figure 3.5: Actual vs Predicted demand by GA, PSO, and ABC based FFNN Models](image)

The comparison of forecasting accuracy is also obviously important for forecasters to evaluate their hypotheses/ models. As, all the learning schemes employed in this study for FFNN training are population based. Therefore, it becomes necessary to verify the contributions of proposed approach. There are numerous statistical procedures proposed in literature for forecast accuracy comparisons. To ensure the statistical significance of
Table 3.2: MAPE and standard deviation of error over 30 runs for three techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>ABC</th>
<th>PSO</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Best</td>
<td>Worst</td>
</tr>
<tr>
<td>Training</td>
<td>2.341</td>
<td>1.936</td>
<td>3.434</td>
</tr>
<tr>
<td>Validation</td>
<td>2.177</td>
<td>1.728</td>
<td>2.739</td>
</tr>
<tr>
<td>Testing</td>
<td>2.295</td>
<td>1.871</td>
<td>3.111</td>
</tr>
<tr>
<td>Average</td>
<td>2.309</td>
<td>1.896</td>
<td>3.281</td>
</tr>
</tbody>
</table>

Table 3.3: Average MSE Comparison of Three Techniques

<table>
<thead>
<tr>
<th>Model Accuracy</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO</td>
<td>8.79E-04</td>
</tr>
<tr>
<td>GA</td>
<td>3.95E-03</td>
</tr>
<tr>
<td>ABC</td>
<td>7.16E-04</td>
</tr>
</tbody>
</table>

our proposed ABC based FFNN hybrid model for forecasting accuracy improvement, we used a statistical test proposed by Diebold and Mariano [71]. This test is applied in [34, 72] to compare forecasting accuracy of models. Diebold-Mariano(D-M) test statistic retrieves the equality of forecast accuracy of forecasts obtained from two forecasting models. It calculates the D-M test statistic on the basis of the loss differential, which is the difference of squared forecast error.

The D-M test is performed at two values of $\alpha = 0.05$ and $\alpha = 0.10$, with a null hypothesis $H_0$: forecast accuracy is same for the ABC-FFNN model when compared with PSO-FFNN and GA-FFNN models. The test results are provided in Table 3.4. Here, $e_1$, $e_2$, and $e_3$ denote the vectors of the forecast errors from the ABC-FFNN, PSO-FFNN, and GA-FFNN models, respectively. The p-value is a significant test output that either accepts or rejects the $H_0$. If p-value is lesser than $-1.96$, or greater than $1.96$, it rejects the null hypothesis. Whereas, H-value determines the decision of $H_0$. The H-value = 1, indicates the rejection of null hypothesis and if H-value = 0, it refuses to indicate any significant difference between the two models.

From the obtained test results, it is evident that the $H_0$ is proven false for both values of alpha when accuracy of ABC-FFNN model is compared with both PSO-FFNN and GA-NN models. On the basis of observed facts, it is concluded that ABC-FFNN model achieved significant forecasting accuracy, which is not equivalent to accuracy of PSO-FFNN or GA-FFNN model and hence, the null hypothesis is rejected. However, it is interesting to note that $H_0$ is not rejected, when GA-FFNN and PSO-FFNN models are compared with each other for both values of alpha.
### Table 3.4: Asymptotic Test

<table>
<thead>
<tr>
<th>Electric load forecasting</th>
<th>Asymptotic ($S_1$) test</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO-ANN vs ABC-ANN</td>
<td>$H_0 : e_2 = e_1$</td>
<td>$S_1 = 33.8904; p\text{-}value = 4.4456\times 10^{-92}$</td>
<td>$S_1 = 33.8904; p\text{-}value = 4.4456\times 10^{-92}$</td>
</tr>
<tr>
<td></td>
<td>$H_0 : e_2 = e_1$</td>
<td>$H\text{-}value = 1$: Reject $H_0$</td>
<td>$H\text{-}value = 1$: Reject $H_0$</td>
</tr>
<tr>
<td>GA-ANN vs ABC-ANN</td>
<td>$H_0 : e_3 = e_1$</td>
<td>$S_1 = 62.0015; p\text{-}value = 1.3013\times 10^{-18}$</td>
<td>$S_1 = 62.0015; p\text{-}value = 1.3013\times 10^{-18}$</td>
</tr>
<tr>
<td></td>
<td>$H_0 : e_3 = e_1$</td>
<td>$H\text{-}value = 1$: Reject $H_0$</td>
<td>$H\text{-}value = 1$: Reject $H_0$</td>
</tr>
<tr>
<td>GA-ANN vs PSO-ANN</td>
<td>$H_0 : e_3 = e_2$</td>
<td>$S_1 = 51.9470; p\text{-}value = 0.1493$</td>
<td>$S_1 = 51.9470; p\text{-}value = 0.1493$</td>
</tr>
<tr>
<td></td>
<td>$H_0 : e_3 = e_2$</td>
<td>$H\text{-}value = 0$: Not Reject $H_0$</td>
<td>$H\text{-}value = 0$: Not Reject $H_0$</td>
</tr>
</tbody>
</table>

#### 3.6 Conclusion

In this work, a new hybrid model for short term load forecasting is formulated. It reveals optimization power of ABC algorithm for training multilayer ANN and finding global minimum in an efficient manner. This is a unique formulation to draw attention of researchers towards the innate optimization capability of ABC algorithm for LF and related problems. The FFNN weights are adjusted recursively by ABC algorithm to achieve better accuracy of results. Results are presented and compared in tables and graphs, which proved that ABC based LF model is better approach than PSO and GA. A greater accuracy with MAPE of 1.89% is achieved. Accuracy comparison test proves the significance of the forecasting accuracy improvement from the proposed model [73]. The proposed model in this chapter is further improved by introducing several modifications to base ABC algorithm and hence, an integrated LF model is presented in Chapter 5.
Chapter 4

Artificial Cooperative Search
Algorithm based Load Forecasting Model

4.1 Introduction

Artificial neural networks (ANN) are quite popular for solving complex non-linear problems proving their superiority over traditional methods because of their robustness, generalization, parallelism, fault tolerance and learning abilities [35]. They have been successfully applied to problems like function approximation, regression analysis, filtering, clustering, classification, pattern recognition, financial market prediction, and forecasting. These tasks are achieved by iteratively adjusting the weights of the neural network, in order to minimize the mean squared error (MSE) between the input and target values over the entire training pattern. Back propagation based on gradient descent algorithm is the most commonly used method to train multi-layer perceptron neural network model whose performance is highly dependent upon the shape of error surface, initial structure of the network and its parameters because it's known to stick in local minima [21]. In order to surmount the problems faced by gradient based algorithms, different global optimization techniques have been introduced to train neural networks such as: particle swarm optimization (PSO) [54], artificial bee colony algorithm (ABC) [46], differential evolution (DE) [74], genetic algorithms (GA) [62], and evolutionary programming algorithms [75].
In this context, artificial cooperative search (ACS) is a recently introduced, two-population based algorithm for global search [76]. It is inspired of mutualism and cooperation based biological interaction of two eusocial superorganisms living in the same habitat. Author of ACS reported its success to solve numerical optimization problems and examined its optimization ability against 91 benchmark functions. This motivated us to explore the optimization ability of ACS to train ANN. In this experimentation ACS is adopted as an alternative optimization algorithm for training ANN weights, which is a unique formulation of a hybrid model. In order to achieve the goal of better results from the newly introduced ACS algorithm, same set of data sets and default parameter settings are used as given in [76, 77].

The objective of this study is to analyze the performance of ACS on well-known classification and regression problems from different domains. Further, a comparative investigation of the successes stated in [76] is also carried out. For this, four classification problems, discussed in PROBEN1 [77] are used. The PROBEN1 is a set of ANN benchmark problems and rules to evaluate the performance of optimization algorithms.

In this experimentation ACS algorithm is used in order to train neural network and its performance is compared with the popular evolutionary algorithms like GA, PSO, cuckoo search (CS), differential search algorithm (DSA), and ABC on some known classification problems. Performance of ACS to train ANN is also investigated with aforementioned techniques on real world problem of accurate load forecasting (LF). Accurate electric demand estimation plays an important role in generation and transmission planning, scheduling, and unit commitment for economic and reliable operations of electric utility [2]. Electric LF is an active area of research; where, more advanced and effective forecasting models are proposed. In the gamut of hybrid approaches in LF; GA, PSO, ACO, and now the ABC is being used to tune training parameters and weights of ANN, and support vector machines (SVM) [3]. Problems with gradient descent based training functions [35] and availability of alternative optimization algorithms also motivated us to explore modern approaches to solve the accurate LF problem. Applications of such optimization techniques to LF are discussed in [32, 36, 37]. In order to reveal the optimization capability of ACS, ANN based hybrid model for LF is formulated. This is a unique formulation to draw attention of researchers towards the innate optimization capability of ACS algorithm for LF and related problems.

### 4.2 Experimental Setup

This section covers the description of data set used for this study, settings of each algorithm used for ANN training, and ANN architecture.
4.2.1 IESO Hourly Demand Data

To predict the hourly electric energy demand of Ontario State. This dataset is obtained from IESO data repository. It contains hourly demand of along with several load effecting parameters like hourly energy price and calendar information. Historical demand and price data of last 11 years (2003 – 2013) is used in this study.

4.2.2 Training, Validation & Testing Set

The whole dataset is divided into training set (50%), validation set (25%) and testing set (25%). The data is linearly transformed into 0 and 1. The missing values are encoded using binary inputs.

4.2.3 Algorithm Parameters

ACS Settings: For evaluating ACS against other algorithms following parameters are used in this study. Its population size is 30 and maximum iteration count is 3000.

CS Settings: CS has two control parameters for its learning process, in this experimentation $\beta = 1.50$ and $p0 = 0.25$ are used as recommended in [78].

PSO Settings: The control parameters of PSO include, population count $P$, inertia weight $W$, position limit $L1$, velocity limit $L2$, and two learning rate constants $C1$, $C2$. Here $P$ is 30, $W$ is 0.6, both $C1$ and $C2$ have the same value of 1.80. Here range of $L1$ and $L2$ is $[-1.0, 1.0]$. These parameter settings are recommended in [64].

ABC Settings: The number of food sources is 28 which is also the population count of honey bees. The count for each of employed and onlooker bees is 14. The dimension $D$ of the search space is 50 as recommended in [64]. The upper and lower bounds are $[-1.0, 1.0]$.

DSA Settings: DSA has two control parameters, $p1$ and $p2$. Both have the same value of $3 * \text{rand}$.

4.2.4 ANN Architecture

For comparison 30 runs of each algorithm are made on fixed neural network architecture for a same problem. MSE is used in order to evaluate the ANN performance.

4.2.5 Stopping Criteria

During the training of ANN the following stopping criteria is used

1. Stop when the number of epochs reaches 3000.
2. Stop if the MSE decreases below 0.01.
3. Stop if $\frac{GL(t)}{P_a(t)} > a$ where, $a$ is the training strip length and $GL(t)$ is the generalization loss on validation set.
4.3 ACS based ANN Optimization for LF model Development

The ACS algorithm is based on the migrating behavior of artificial super organisms as they biological interact in order to achieve the global minimum value pertaining to the problem. In ACS algorithm, a super organism consisting of random solutions of the related problem corresponds to an artificial super organism migrating to more productive feeding areas. ACS algorithm contains two super organisms; \( \alpha \) and \( \beta \) that contain artificial sub-super organisms equal to the dimension of the population \( N \). The number of individuals within each sub-super organism equal to the dimension of the problem \( D \). In ACS algorithm, \( \alpha \) and \( \beta \) super organisms are used for the detection of artificial Predator and Prey sub-super organisms. The Predator sub-super organisms in ACS algorithm can pursue the Prey sub-super organisms for a period of time while they migrate towards global minimum of the problem. When the iterative calculation process of ACS algorithm that is named as co-evolution process is considered, it can be seen that the two super organisms looking for the global minimum of the related problem, establish a co-operation based biological interaction between each other. In ACS algorithm the individuals of the \( i^{th} \) sub-super organism are initialized by Eq. 4.1.

\[
\alpha_{i,j,g=0} = \text{rand.(}up_j - \text{low}_j\text{)} + \text{low}_j \\
\beta_{i,j,g=0} = \text{rand.(}up_j - \text{low}_j\text{)} + \text{low}_j 
\]

(4.1)

Where, \( i = 1, 2, 3, \ldots, N \) and \( j = 1, 2, 3, \ldots, D \) and \( g = 0, 1, 2, 3, \ldots, \text{maxcycle} \). The \text{rand} is a random number chosen from a uniform distribution \([0, 1]\). The \text{up}_j and \text{low}_j are the upper and lower bounds for \( j^{th} \) dimension of the search space. The fitness level of each sub-superorganism is calculated by using Eq. 4.2

\[
y_{i;\alpha} = f(\alpha_i) \\
y_{i;\beta} = f(\beta_i)
\]

(4.2)

In ACS algorithm predator super organism is determined randomly in each cycle using the rule given in Algorithm 2.

Algorithm 2: Determination of Predator, \( y_i;\alpha \) and \( y_i;\beta \)

Data: \( \alpha, \beta, y_\alpha, y_\beta \)  
Result: Predator, key
1 if \( \text{rnd} < \text{rnd} \) then  
2 \hspace{0.5cm} Predator = \alpha, \ y_{\text{Predator}} = y_\alpha, \ key = 1  
3 else  
4 \hspace{0.5cm} Predator = \beta, \ y_{\text{Predator}} = y_\beta, \ key = 2  
5 end
Similarly, prey super organism is determined as given in Algorithm 3.

\begin{algorithm}
\textbf{Data:} $\alpha, \beta$
\textbf{Result:} Prey
\begin{itemize}
    \item [1] if $\text{rnd} < \text{rnd}$ then Prey = $\alpha$ else Prey = $\beta$
    \item [2] Prey := permuting(Prey)
\end{itemize}
\textbf{Algorithm 3:} Determination of Preyi
\end{algorithm}

Where, \text{permute}() is function which produces random permutations of a given set. The biological interaction location between predator and prey sub-super organism is determined using the Eq. 4.3.

\begin{equation}
\text{Location}x = \text{Predator} + R \ast (\text{Prey} - \text{Predator})
\end{equation}

Where, $R$ is the scaling factor which controls the biological interaction. $R$ is obtained using the rule given in Algorithm 4.

\begin{algorithm}
\textbf{Data:} $p$ : Probability of biological interaction in ACS.
\textbf{Result:} $M$ : Binary-valued integer-map of passive individuals.
\begin{itemize}
    \item [1] $M_{1:N,1:D} = 1$
    \item [2] for $q \leftarrow 1$ to $N \cdot D$
        \begin{itemize}
            \item [3] if $\text{rnd} < (p \cdot \text{rnd})$ then $M_{r nd int(N),r nd int(D)} = 0$
        \end{itemize}
    \item [4] end
    \item [5] if $\text{rnd} < (p \cdot \text{rnd})$ then
        \begin{itemize}
            \item [6] for $i \leftarrow 1$ to $N$
                \begin{itemize}
                    \item [7] for $j \leftarrow 1$ to $D$
                        \begin{itemize}
                            \item [8] if $\text{rnd} < (p \cdot \text{rnd})$ then
                                \begin{itemize}
                                    \item [9] $M_{i,j} = 1$
                                \end{itemize}
                            \item [10] else
                                \begin{itemize}
                                    \item [11] $M_{i,j} = 0$
                                \end{itemize}
                            \end{itemize}
                        \end{itemize}
                \end{itemize}
        \end{itemize}
    \item [13] end
    \item [14] end
    \item [15] end
    \item [16] for $i \leftarrow 1$ to $N$
        \begin{itemize}
            \item [17] if $\sum_{j=1}^{D} M_{i,j} = D$ then $M_{i,r nd int(D)} = 0$
        \end{itemize}
    \end{itemize}
\textbf{Algorithm 4:} Determination of $M$ of the Passive Individuals
\end{algorithm}

Where, $k$ returns a random number chosen from gamma distribution with shape parameter 4 and scale parameter 1.
In ACS algorithm the individuals which are migrating are called active individuals and the individuals which are not involved in migration are called passive individuals. The passive individuals of the predator are determined using the Algorithm 5.

**Algorithm 5:** Upgrading $x'_i$s by using Active Individuals

```
Data: M
Result: Updated $x$
1 for $i \leftarrow 1$ to $N$ do
   2 for $j \leftarrow 1$ to $D$ do
      3 if $M_{i,j} > 0$ then $x_{i,j} := \text{Predator}_{i,j}$ end
   4 end
5 end
```

Where, $p$ is the probability of the biological interaction restricting the number of passive individuals in each sub-super organism. The value of the $p$ varies from 0.1 to 0.9. In ACS algorithm only active individuals can discover new biological interaction locations. This situation is modeled by rule given in Algorithm 6.

**Algorithm 6:** Generation mechanism of Scale Factor; $R$

```
Result: $R$
1 if $\text{rnd} < \text{rnd}$ then
   2 \hspace{1em} $R = 4 \cdot \text{rnd} \cdot (\text{rnd} - \text{rnd})$
4 else \hspace{1em} $R = \kappa |\kappa \sim \Gamma (4 \cdot \text{rnd}, 1)$
5 end
```

In ACS algorithm the strategies for population generation, mutation, crossover, and boundary control mechanisms are different from the other population based algorithms. The ACS return the best solution whose dimension is equivalent to the dimension of the problem ($D$). Here, dimension of the problem corresponds to total number of neuron connection weights, which is 80 in this case. The population size is 30 and maximum iteration count is 3000. The ACS algorithm iterates over the specified number of executions and during each execution it produces random solution of superorganisms containing sub-superorganisms equivalent to $N$. The number of individuals within each sub-superorganisms is $D$. This solution is then passed to ANN block to re-assign neuron connection weights. The ANN produces the forecasts and calculates the MSE. Further, the objective function of ANN is passed back to ACS for a new generation of better superorganisms. This process is repeated for the specified iteration count.
4.4 Results and Discussion

This section discusses and compares the successes of ACS algorithm with three other computational intelligence based algorithms (i.e. PSO, GA, and ABC) for numerical optimization problem solving. These algorithms are compared by using benchmark problems. This comprises of four standard and widely used benchmark problems [76, 77]. The dimensions of the benchmark functions are 30 and 50. The detailed description of the benchmark problems used in this set are given in [77]. The comparison on benchmarking functions is provided in Table 4.1.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>ABC</th>
<th>ACS</th>
<th>PSO</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
<td>N/A</td>
<td>N/A</td>
<td>Weight = 0.9, C1=C2=1.494</td>
<td>Mutation = 0.05, Recombination=0.7</td>
</tr>
<tr>
<td>Population size</td>
<td>28</td>
<td>50</td>
<td>25</td>
<td>5000</td>
</tr>
<tr>
<td>Dimension = 30</td>
<td></td>
<td></td>
<td>Sphere [-100,100]</td>
<td>1.41E-09, 3.5E-38, 14.4888, 1.686</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Griewank [-600,600]</td>
<td>8.07E-12, 0, 51.677, 0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rastrigin [-5.12,5.12]</td>
<td>1.3, 0, 86.347, 0.319</td>
</tr>
<tr>
<td>Dimension = 50</td>
<td></td>
<td></td>
<td>Sphere [-100,100]</td>
<td>4.53E-05, 6.52E-37, 52.462, 6.292</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Griewank [-600,600]</td>
<td>2.34E-06, 0, 147.102, 0.1470</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rastrigin [-5.12,5.12]</td>
<td>10.256, 0, 174.915, 0.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rosenbrock [-30,30]</td>
<td>38.243, 48.979, 312.727, 151.464</td>
</tr>
</tbody>
</table>

Each algorithm is executed 30 times to train ANN. Results are presented as a mean of 30 runs. In order to reveal the optimization capability of each algorithm to reach global minimum, several performance measure of great importance are used i.e. mean epochs of training, mean training error, mean, max, and min classification rates. Mean of these performance measures is also provided.

This experimentation is performed on 11 years (2003 – 2013) hourly electricity load demand data sets obtained from Independent Electricity System Operator (IESO) for Ontario State. This data is publicly available and is highly accurate. Based upon high correlation, different input factors like calendar events, demand of last hour, and previous day, weather conditions, and hourly energy price are considered as input parameters in this study. In order to achieve the desired objective of accurate STLF model, it is necessary to compare all three models on same criteria, for that reason data and features sets are identical for all models. For training the neural network models 70% of data is utilized and 30% is used for testing and validation by further dividing the data into two equal parts. In this study, model accuracy is ranked on the basis of mean absolute percentage error (MAPE) which is the most common method to evaluate LF accuracy. It expresses accuracy as a percentage. This is defined by the formula as given in Eq. 4.4.
\[ M = \frac{1}{n} \sum_{t=1}^{n} |A_t - F_t| \]  

(4.4)

Where, \( A_t \) is actual load at time \( t \) and \( F_t \) is forecasted demand value at same time, \( n \) is number of samples.

The results are presented in Table 4.2, where, average MAPE of 10 runs is presented. Training, validation, and testing MAPE of each technique is presented. The standard deviation of all three MAPE is also presented. ACS showed the more generalization ability, also it has lesser standard deviation of MAPE and there is less difference between its training, validation, and testing MAPE. On the other hand, ABC has produced better standard deviation on MAPE and stands second to show better generalization. The PSO based ANN model has produced lower MAPE but the standard deviation is higher than ABC, GA, and ACS based models. The GA based could not perform well to produce lower MAPE; still, the standard deviation is lower than of PSO. From the results presented in Table 4.2, it can be concluded that ACS avoided the overfitting and underfitting scenarios on this data set. Comparison of MAPE is also presented by chart in Fig. 4.1.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>PSO stdev</th>
<th>ABC stdev</th>
<th>GA stdev</th>
<th>ACS stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training MAPE</td>
<td>2.865</td>
<td>1.273</td>
<td>3.609</td>
<td>0.382</td>
</tr>
<tr>
<td>Validation MAPE</td>
<td>2.514</td>
<td>1.097</td>
<td>3.557</td>
<td>0.588</td>
</tr>
<tr>
<td>Testing MAPE</td>
<td>2.767</td>
<td>1.532</td>
<td>3.555</td>
<td>0.525</td>
</tr>
</tbody>
</table>

Table 4.2: Results on Electric Demand data

![MAPE Comparison](image)

**Figure 4.1:** Comparison of MAPE Obtained by all Techniques
4.5 Conclusion

In this study, five different evolutionary optimization techniques are examined to train ANN’s. The formulation of hybrid model of ACS with ANN is a unique contribution being presented. All optimization techniques are evaluated on different benchmark functions and real-world problem of load forecasting from regression domain. Results are presented and compared in tables and graphs. It can be clearly observed that ACS has shown better optimization capability on benchmark functions as well as on demand data set. It has produced better accuracy while achieving lower MAPE on train, validation, and test data. Also, the standard deviation of MAPE on 10 iteration is low when compared to PSO, ABC, and GA. The ABC model as discussed in Chapter 3, stands second in this case. From the obtained results, it can be concluded that ACS has inherent potential of optimization.
Chapter 5

Improved Artificial Bee Colony Algorithm based Load Forecasting Model

The easiest way to predict the future is to invent it.

Anonymous

5.1 Introduction

Meta-heuristic optimization algorithms based upon swarm intelligence and foraging behavior of bees, are popular in solving complex and multidimensional optimization problems. Such soft computing techniques are an attractive alternative to traditional hard computing archetypes. Unlike to statistical techniques [79, 80], these methods are not dependent to data and gradient, so these hardly get trapped in local minima and divergence situations. These algorithms are easy to implement and can easily be coupled with other algorithms. Such meta-heuristic techniques are inspired from different natural phenomena of evolution; like genetic algorithm (GA), that simulates survival of the fittest law [62, 81], particle swarm optimization (PSO), which is inspired from the social behavior of swarms [54], ant colony optimization (ACO) that stimulates the foraging behavior of ant colonies [82, 83], and firefly algorithm (FA) which is inspired by fireflies [84].

Based upon the foraging behavior of honey bee swarm, Karaboga [85], introduced a new kind of optimization algorithm called artificial bee colony (ABC) algorithm for numerical function optimization [46]. Although ABC is relatively a new algorithm for function optimization. Still, it received much attention of researchers to solve complex optimization problems, due to its simplicity, fewer parameters, ease of implementation and competitive performance against other population based algorithms. Performance and quality
of results are comparable to other mature optimization techniques. A variety of applications and modified versions of ABC algorithm are discussed in [86–91]. Applications of this algorithm include electric power systems, parallel and grid computing, data clustering, image analysis, signal processing, and computer science applications. Comparative analysis of this algorithm are conducted and compared with other meta-heuristic and hybrid algorithms like, PSO, GA, differential evolution (DE), and Back Propagation using benchmarking optimization functions [46, 92–96]. A number of evolved versions of ABC algorithm are also proposed with modified techniques for initialization, searching, selection, and finding best solutions [87, 97–103]. As in [104] a modified version is presented where crossover and mutation operators were incorporated to improve the quantum evolutionary algorithm (QEA). To increase the diversity of population, another modified version is proposed in [105], in which original probability function of ABC algorithm is replaced.

However, the real power of ABC algorithm in solving a specific problem dwells in its mechanism of solution search, which consists of exploration and exploitation traits of bee colony. Here exploration refers to ability to reach global optimum by investigating the various unknown regions in the solution space i.e. each point of search space is evaluated for global optimum value. While exploitation is an ability to produce better solution from the knowledge gained from previous ones [106, 107]. However, the exploration and exploitation contradict each other. Still, success of ABC algorithm or other population based optimization algorithms resides in keeping a good balance in above mentioned traits. As mentioned in [106–108] that mechanism of solution search of ABC algorithm is good at exploration but poor at exploitation. To improve the exploitation ability of ABC, many improvements are suggested in the literature mentioned earlier. Observing the greater optimization power and inherent potential for improvement, we got motivation to further extend the optimization ability of base algorithm by adopting modifications in its processes. Objective of this work is to improve the accuracy of forecasts by utilizing the improved optimization and convergence capability of ABC to train artificial neural networks (ANN). As ANN is proven successful in modeling load forecasting (LF) problems. Still, problems with gradient descent based training functions [35] and availability of alternative optimization algorithms motivated us to explore modern approaches to solve the accurate LF problem. Part of this work is already reported in [67].

In this experimentation, evolving behavior of ABC algorithm is presented by applying modifications in various different ways to achieve a good combination of the global exploration and local search without being trapped into local optima. Individual modifications are introduced in parallel at each level of base algorithm like population initialization, search process, probability functions and best solution selection phases. Further all the modifications are also applied in series to develop a new extended and customized version of base ABC algorithm. To utilize the local search ability of genetic algorithm (GA); genetic operators are integrated with modified ABC to extend its searching in reduced
Chapter 5. Improved ABC-GA-FFNN model

search space. Adopting modifications of ABC in series and integration of GA is the main contribution of this research work.

Optimization capability of newly formulated algorithm is coalesced with ANN at every step to formulate a hybrid LF model and the results are evaluated for forecasting accuracy. This effort is carried out to explore different ways of improving exploration and exploitation capabilities of ABC algorithm, and to further improve optimization ability of this algorithm to achieve better performance on peak load estimation problems. The work presented, which evolves ABC algorithm for ANN parameter optimization is a unique formulation to draw attention of researchers towards the inherent optimization capability of ABC algorithm to find global minimum in an efficient manner. Mathematical modeling of base algorithm and evolving steps are presented. Results obtained by this new approach are presented and compared with other mature and competitive approaches like simulated annealing (SA) [109], and PSO techniques for forecasting which confirms the effectiveness and applicability of proposed approach in LF domain.

5.2 Evolving Behavior of ABC Algorithm

ABC algorithm is a swarm based meta-heuristic algorithm, inspired from intelligent foraging behavior of honey bee swarm searching for food. In bee colony algorithm position of each food source corresponds to a possible solution for the optimization problem, and the amount of nectar of a food source depicts the fitness value (quality) of the associated solution. The total bees in colony (population) can be categorized into three groups: employed, onlookers, and scout bees. Where employed bees and onlookers are equal in number in a hive and correspond to equal number of solutions in population [46]. Employed bees fly around randomly and search for the food sources (solutions) and keep the quality (nectar amount / fitness) of food source in their memory. On their return to hive, they perform waggle dance to share the information of food sources with onlooker bees waiting on the dance area [57]. Onlooker bees evaluate the fitness level of a food source from information provided by employed bee and assign a probability according to its nectar amount.

Onlooker bees now send employed bees to collect food from the particular selected source. The employed bee whose food source is exhausted becomes the scout bee and starts her search again to find new and/or better food sources from other candidate solutions. Scout bees first search food randomly and locate sources with rich amount of nectar. This process continues until whole colony finds a best solution from the solution space or maximum count of cycles is reached. This whole process can be translated into mathematical steps as given below.

An initial population of $N$ food sources is generated randomly within the search space boundaries. The Eq. 5.1. is used to initialize the population.

$$x_{ij} = x_{j}^{min} + rand(0,1)(x_{j}^{max} - x_{j}^{min}) \quad (5.1)$$
Where \( i \) represent the \( i_{th} \) food source number and \( j \) is the optimization variable associated with \( i_{th} \) food source. Here each food source is a d-dimension vector.

Now nectar amount in each food source is calculated and fitness of food sources is evaluated using Eq. 5.2:

\[
\text{fitness}_i = \begin{cases} 
\frac{1}{(1+f_i)} & \text{if } f_i \geq 0 \\
1 + \text{abs}(f_i) & \text{if } f_i < 0 
\end{cases}
\]  
(5.2)

Where \( f_i \) is the cost associated with solution \( x_i \), this function can be directly used as cost function in maximization problems.

Now each employed bee searches a new candidate food source \( v_i \) in the neighborhood of food source already in memory. New food source is compared based upon the visual information of bees and evaluated according to Eq. 5.3. Here greedy selection is applied to select a better food source as new candidate food source.

\[
v_{ij} = x_{ij} + \phi_{ij} (x_{ij} - x_{kj})
\]  
(5.3)

Where \( j \) is a random optimization variable in the range \([1, D]\) and \( k \) is a randomly selected food source different from \( i \), which is a uniformly distributed real random number in the range \([-1, 1]\).

Probability of selecting a solution in population is proportional to nectar amount in the source or fitness value of source in population. By using roulette wheel selection method [69], each onlooker bee selects a food source with higher probability \( p_i \), by using Eq. 5.4, and generates a candidate solution according to Eq. 5.3.

\[
p_i = \frac{\text{fitness}_i}{\sum_{i=1}^{N} \text{fitness}_i}
\]  
(5.4)

Here solutions are evaluated and poor food source are replaced by better ones through greedy selection process. These steps are repeated until a termination criterion is reached i.e. best solution found.

Success of ABC resides in exploration and exploitation of solution search space and keeping a good balance between these two features can ensure better optimization. Original ABC algorithm has great potential to be evolved in both terms. As per concluding remarks of Karaboga, mentioned in [64], ABC has great ability of optimization; still, some modifications to the original structure are necessary in order to significantly improve its performance in convergence. New strategies can be defined for searching, selection, new neighbor production and scout production. Following this, different techniques are adopted in this study, to evolve base algorithm. Modifications are introduced in initialization, solution weighting, evaluation, and neighbor selection. These are discussed in following section.
5.3 Applied Modifications

As discussed, there are four major steps of ABC algorithm, i.e. initialization, fitness evaluation of solutions, neighbor selection and solution weighting. There are two basic principles of ABC algorithm, exploration and exploitation of solution in search space. As mentioned in [106-108] ABC is good in exploration but it lacks the exploitation capability. Therefore, to enhance exploration and exploitation traits of the algorithm and to get a best combination of both, different modification are introduced at each aforementioned step of base algorithm, as follows.

5.3.1 Modified Initialization

The first modification introduced is in the population initialization phase of the ABC algorithm. The idea is that if a solution $x$ is too far from the optimal solution then its opposite $x'$ will be near to the optimal solution. This modification is based upon opposition based learning (OBL) [110], according to which, there are some scenarios while evolving the population, when best solution is in the opposite location of a randomly initialized solution, this could increase the computational time of search. So to overcome this scenario generalized opposition based learning (GOBL) by employing a space transformation technique is proposed in [111]. OBL approach is successfully integrated within Differential Evolution (DE) in [112], PSO in [113, 114], and ABC in [115]. While GOBL strategy is used with DE [116, 117], and PSO [118], to enhance the performance of both. In opposition based learning approach, the initial population is first initialized and then for every solution its opposite solution is found using Eqs. 5.5 and 5.6.

$$ox_{ij} = k \times (x_{max ij} + x_{min ij}) - x_{ij} \quad (5.5)$$

$$ox_{ij} = UB_{ij} + LB_{ij} - x_{ij} \quad (5.6)$$

This generalized opposition based initialization of solutions is limited by the same upper and lower bounds of the search domain[110]. This is to avoid lose of information gathered in the search process. However, the remaining algorithm remains the same, after generating the opposite solutions against the original ones, both food sources are evaluated and greedy selection is applied. For abandoned solutions, opposite food sources are generated again and then evaluated.

Let $x$ be a number in the range $[a, b]$ then its opposite is found by Eq. 5.7.

$$x' = a + b - x \quad (5.7)$$

This modification is termed as “Opposition based ABC” and referred in text as ABC-Opp.
5.3.2 Modified Search

Next modification applied is in search process of base algorithm; the original ABC algorithm applies greedy selection to find new solution from the existing parent solutions. In basic ABC algorithm the new solutions are produced as given in Eq. 5.3.

Using this equation new solution $v_i$ get only one change of parameter from parent solution $x_i$, this results in slow convergence. To improve the convergence rate of ABC algorithm, a modification is proposed in [86], by combining $x_i$ and its former neighbor $x_{i-1}$ to get the new solution $v_i$.

\[
v_{ij} = x_{ij} + \varphi_{ij}(x_{ij} - x_{kj}), \text{ for } i = 1 \tag{5.8}
\]

\[
v_{ij} = x_{i-1j} + \varphi_{ij}(x_{ij} - x_{kj}), \text{ for } i > 1 \tag{5.9}
\]

Where $x_{i-1j}$ is the previous neighbor of $x_{ij}$ and is the better than selected by greedy selection process. This result in the larger search range than in the basic ABC algorithm, and in improved convergence rate. This improved equation is only applied in the exploration of employed bees, and onlooker bees still apply Eq. 5.1, to keep the local searching. This results in a combination of global and local exploration to get a better balance in food searching and exploitation of bee colony by avoiding the optimization to be trapped into the local best value. This modification is termed as “ABC with Modified Search” and referred as ABC-Mod.

5.3.3 Global Best Search

In ABC algorithm if a particular solution could not be improved after a certain number of iterations, it is discarded and replaced by a new solution generated by scout bees, which is selected based on probabilities calculated according to Eq. 5.3 and Eq. 5.4, respectively.

To improve the exploitation feature of population based optimization algorithms [107], introduced a global best (gbest) solution to improve the search of candidate solutions by adding a gbest term in Eq. 5.1. Adding this term can accelerate in optimizing objective function. Now the modified equation becomes,

\[
v_{ij} = x_{ij} + \theta_{ij}(x_{ij} - x_{kj}) + \psi_{ij}(x_{ij} - x_{kj}) \tag{5.10}
\]

Where $\theta$ and $\psi$ are random numbers in the range $[0, C]$, $C$ is a non-negative number and is the $j_{th}$ term of global best solution. This change is termed as ABC-GBest.
5.3.4 Modification in Probability Function

To improve the exploitation capability of the standard ABC algorithm, a modified version is proposed in [106], which uses an inversely proportional mutation function. As in ABC algorithm employed and onlookers bees are responsible for exploitation process; each new solution is assigned a probability value proportional to fitness level. In [106] a different approach is used to calculate probabilities, inspired by affinity maturation process of the immune system [119, 120] as given in Eq. 5.11.

\[
P_i = 1 - (0.05 * (1 + f_i) * \exp(-f_i))
\]

where \(f_i\) is the normalized fitness value between 0 and 1. The probability value is inversely proportional to the fitness value where probability decreases as the fitness value increases. Another modification proposed is to mutate the best solution \(x_k\) by a randomly selected solution different from \(x_i\) [106].

\[
V_{i,j} = X_{best,j} + \theta(X_{best,j} - X_{k,j}) + \varphi(X_{i,j} - X_{k,j})
\]

Where \(X_k\) is the randomly selected solution, \(j\) is randomly selected parameter in \(D\) dimension, \(\theta\) and \(\varphi\) are random numbers in range \([-1, 1]\). This modification is applied to evolved version of ABC and named as ABC-Inv.

5.3.5 Exploration by Genetic Operators

GA based on Darwinian evolution theory, “survival of the fittest”, is a multi agent search technique. The real power of GA comes from its selection mechanism, parallel processing and genetic operators. The term genetic operators (GO’s) is applied to the methods used to simulate nature in computer-based evolutionary systems. The GO’s that form the basis of simulated genetic systems are crossover, fitness evaluation, reproduction, and mutation [62, 81, 121]. The GO’s used here are mutation and crossover. Mutation operator can effect each solution with a given probability and is applied to maintain genetic diversity. Whereas, crossover operator is also called recombination, and is a process of combining existing solutions to produce new chromosomes/ solutions that inherit some or all of their attributes from one or both of their parents. As mentioned in [64] ABC like all other evolutionary optimization approaches, does not use GO’s. This results in slow convergence performance of ABC for local minimum. To cope this factor, GO’s are applied at both phases of exploration and exploitation.

In this phase, the original employee bees search process is replaced, as given in Eq. 5.3, and Eq. 5.4. with the crossover and mutation operators to analyze their effect on exploration of base algorithm. First crossover operator is applied on initial population then mutation operator is applied and next generation of solutions is obtained. This modification is termed as GO’s based ABC or GO-ABC.
5.3.6 Exploitation by Genetic Operators

For the purpose of further exploiting the solutions obtained by modified version of ABC as given above. Here GO’s (mutation, crossover) are applied after the ABC selected a final solution. GO’s come into play at this stage and further evolve the quality of solutions. Each solution is evaluated against current best and finally best solution is selected. ABC results in reducing large search space into small search space, as GA can perform well on small search space, that’s why this formulation is applied to find local best from reduced search space of ABC. As this model integrates original ABC algorithm with GA, which results in improved exploitation power of hybrid approach; that’s why this modification is termed as ABC-GA.

The diagram in Fig. 5.1. shows the modifications and their sequence of application. Some modifications are applied in parallel, which can also be witnessed.

![Diagram](image.png)

**Figure 5.1**: Modifications in Base ABC Algorithm

5.3.7 Evaluation on Statistical Benchmark Functions

In order to justify the performance of proposed modified versions of ABC. Their efficiency is evaluated against four benchmark functions. The populations size and iteration count of each problem are 50 and 1000 respectively. The upper and lower limits are provided against each function. These functions are performed on two dimensions, i.e. 30 and 50. The results are provided in Table 5.1, as a mean of 10 runs.

<table>
<thead>
<tr>
<th>ABC Versions</th>
<th>ABC</th>
<th>ABC-Opp</th>
<th>ABC-Mod</th>
<th>ABC-Gbest</th>
<th>ABC-GA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension = 30</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphere [-100,100]</td>
<td>6.70E-10</td>
<td>3.83E-10</td>
<td>3.19E-14</td>
<td>1.35E-14</td>
<td>6.41E-16</td>
</tr>
<tr>
<td>Griewank [-600,600]</td>
<td>4.84E-12</td>
<td>2.60E-12</td>
<td>9.99E-16</td>
<td>7.88E-16</td>
<td>5.33E-16</td>
</tr>
<tr>
<td>Rastrigin [-5.12,5.12]</td>
<td>2.75E-13</td>
<td>6.42E-06</td>
<td>5.33E-16</td>
<td>3.55E-16</td>
<td>0</td>
</tr>
<tr>
<td><strong>Dimension = 50</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphere [-100,100]</td>
<td>3.45E-05</td>
<td>9.38E-06</td>
<td>6.19E-07</td>
<td>1.20E-06</td>
<td>9.05E-08</td>
</tr>
<tr>
<td>Griewank [-600,600]</td>
<td>1.49E-06</td>
<td>7.01E-07</td>
<td>1.12E-08</td>
<td>1.72E-08</td>
<td>1.46E-10</td>
</tr>
<tr>
<td>Rosenbrock [-30,30]</td>
<td>2.34E+00</td>
<td>6.06E-01</td>
<td>4.18E-01</td>
<td>1.06E+00</td>
<td>1.04E+00</td>
</tr>
<tr>
<td>Rastrigin [-5.12,5.12]</td>
<td>9.95E-02</td>
<td>2.50E-09</td>
<td>6.04E-15</td>
<td>2.13E-15</td>
<td>0</td>
</tr>
</tbody>
</table>
5.4 Model Development for Peak Demand Estimation

In this section, formulation of a new customized model for LF is discussed; parameters and settings for ANN, PSO, SA, and ABC are provided. ANN training process is shown in Fig. 5.2, where aforementioned techniques are incorporated as learning schemes. Inputs are passed to ANN, whose connection weights are obtained by each of learning schemes and results are evaluated over a number of iterations.

\[ y_j = \sum_{i=1}^{n} x_i w_{ij} \]  \hspace{1cm} (5.13)

Where \( x_i \) is the input vector and \( w_{ij} \) are corresponding weights for this input vector, \( y_j \) is the response of neuron based upon transfer function. For continuous output, neuron uses sigmoid transfer function as given in Eq. 5.14.

\[ y_j = \frac{1}{1 + e^{-x_j}} \]  \hspace{1cm} (5.14)

Network can compute the error \( E \), by taking Mean Squared Error (MSE) that is expressed by Eq. 5.15.

\[ E = \frac{1}{n} \sum_{i}^{n} (y_i - a_i)^2 \]  \hspace{1cm} (5.15)

Where \( y_i \) is output obtained at neuron \( i \) and \( a_i \) is the desired output.

\textbf{Figure 5.2:} ANN Model Training by ABC, PSO, and SA Algorithms
Here, ANN is trained by ABC algorithm to find optimal connection weights. The objective function is to minimize the MSE of the ANN model; and hence, minimize the Mean Absolute Percentage Error (MAPE) of the forecasting model. MAPE expresses accuracy as a percentage, and is defined by the formula as in Eq. 7.2.

\[ M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \]  

(5.16)

Where \( A_t \) is actual load at time \( t \) and \( F_t \) is forecasted demand value at same time, \( n \) is number of samples.

### 5.4.2 Input Factors

This work to predict daily and weekly peak demand, would help power supply companies in medium term demand estimation for generation and distribution planning. In this section aforementioned modified versions of the ABC algorithm are applied to estimate daily and weekly peak load. This experimentation is performed on 5 years daily peak demand data obtained from National Transmission and Dispatch Company (NTDC) of Pakistan and from Lahore Electric Supply Company (LESCO). The sample data used for this study is shown in Table 5.2.

There is a complex and nonlinear relationship between the electricity load and its influencing factors. Electric demand on short-term horizon is affected by calendar inputs, weather conditions, and energy price. The calendar events are time of day, day of week, and holidays; weather parameters include temperature and humidity. The inputs to the ANN should be carefully selected as they affect the forecasting ability of the network. The inputs are chosen based on their correlation with the output load. The inputs of the ANN for this study include; calendar events, weather conditions, day type, previous hour, day, and week demand profiles.

### 5.4.3 ANN Settings

A three layered FFNN is used in this experimentation. The inputs are the variables affecting the variations in the load and the output is next hour load. The number of neurons in hidden layer affects its learning and generalization ability so they must be carefully selected. Hidden layer neurons are selected by taking the average of number of inputs and number of outputs and then gradually increasing the neurons avoiding underfitting and overfitting at the same time until optimal results are found. The ANN used in this study consists of 9 input layer neuron, 8 hidden layer neurons, and one at output layer. This results in 91 neuron weight connections. The sigmoid transfer function is used in the hidden layer. Two biases are added; each on the hidden and output layer.

MSE is used in order to evaluate the ANN performance. The weights are initialized from a uniform distribution in range \([-1, 1]\).
5.4.4 **ABC Settings**

The number of food sources is 28 which is also the population count of honey bees. The count for each of employed and onlooker bees is 14. The dimension of the search space is equal to the 91 weight connections. The upper and lower bounds are $[-1, 1]$.

5.4.5 **GA Settings**

The standard GA with recommended settings is applied in this study. Here, population size for GA is 5000, dimensions are equivalent to number of neuron weights; which is 91, recombination rate is set to 0.7 and mutation rate is 0.05.

5.4.6 **PSO Settings**

This study employs the standard version of PSO with recommended values of two constants $C_1$ and $C_2$, limiting factor and $\phi$ to update particle velocity and location [68]. Here, both constants $C_1$ and $C_2$ have same value i.e. $C_1 = C_2 = 1.494$, limiting factor $= 0.734$ and $\phi = 4.1$. These values provide better compromise in global and local search convergence. The population count of the swarm is 30.

5.4.7 **SA Settings**

The name and inspiration of SA comes from annealing in metallurgy, a method involving heating and controlled cooling of a material. The algorithm starts with a higher value of temperature $T$, and then $T$ is decreased gradually at each step. In this experimentation start value of $T$ is 10.0, and stop temperature is set to 0.1; which should always be $T >= 0$.
In the following section, results obtained by model formulation of ANN with different modified variants of ABC is discussed along with tables and graphs. A total of 9 different models are applied, which include original ABC based ANN, hybrid of ANN with 6 modified versions of ABC, and for cross comparison with other modern approaches, PSO, and SA based ANN are also applied on the same data sets. As, each data set has different type of factors effecting load patterns. However, same data is used for all techniques for one case study. Here 70% of available data set is used for model training while remaining 30% is used for testing and validation by further dividing the data into two equal parts. Here maximum iteration count for each technique is 5000.

5.5 Results

Here aforementioned models are applied to solve two problems of LF, namely daily peak and weekly peak energy demand estimation, and results obtained are discussed. This experimentation is performed on 5 years daily and weekly peak demand data obtained from NTDC and from LESCO.

5.5.1 Case Study (NTDCL Data)

Daily and weekly demand data sets of NTDCL represents peak demand occurred at national grid of Pakistan over 5 years time frame. Daily peak demand is fluctuated by seasonal variation, weather conditions, working days, and weekends. Influencing factors considered for this study are: day of week, day type, week of year, month of year, etc. All above mentioned techniques are applied on this data and forecasting accuracy of different techniques is shown by graphs and tables.

Daily Peak Load Profile

This study is performed to predict daily peak energy demand of NTDCL system and to analyze behavior of different modifications on base algorithm. As ABC is modified in serial manner; step by step, after each new modification, forecasting results and accuracy is also improved, which can be witnessed in graph shown in Fig. 5.3. Where original demand line is compared with predicted values of different models. Maximum improvement is reached at fifth step where GO’s were applied after combined steps of ABC-Opp, ABC-Mod, ABC-GBest, and ABC-Inv. However, when GO’s are used for exploration, it produced poor results. This is closely shown in Fig. 5.4. The graph in Fig. 5.4, closely compares the original maximum demand with predictions obtained from base ABC and modified ABC and further optimized with GO’s (ABC-GA). It can clearly be seen that ABC-GA showed better results when compared to original ABC and it closely follows the trend of original line of daily peak energy demand. It is clear from the graphs that, LF model based on GA based ABC-ANN combination outperformed others in accuracy. The MAPE obtained from all techniques is shown in Table 5.3. Total 9 different techniques are applied, training, testing, and validation MAPE is provided in Table 5.3. Original ABC version showed good accuracy when compared to PSO and
Chapter 5. Improved ABC-GA-FFNN model

Figure 5.3: Actual demand vs Predicted by ABC, PSO, SA and Modified Versions of ABC

Figure 5.4: Actual demand vs Predicted by ABC and ABC-GA assisted ANN
SA based ANN. But it started to improve further with each modification applied. Moreover, ABC-GA outperformed all techniques under study. However, when GO’s are used in-place of employee bees, its accuracy is reduced. This comparison is also presented in bar graph in Fig. 5.5.

Table 5.3: Average MAPE over 10 runs of all Techniques on PEPCO Daily Peak Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>ABC</th>
<th>ABC-Opp</th>
<th>ABC-Mod</th>
<th>ABC-GBest</th>
<th>ABC-Inv</th>
<th>GO-ABC</th>
<th>ABC-GA</th>
<th>PSO</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>4.19</td>
<td>3.52</td>
<td>3.43</td>
<td>3.44</td>
<td>3.26</td>
<td>5.91</td>
<td>2.69</td>
<td>3.83</td>
<td>4.41</td>
</tr>
<tr>
<td>Test</td>
<td>3.51</td>
<td>3.48</td>
<td>3.4</td>
<td>3.14</td>
<td>3.23</td>
<td>7.63</td>
<td>2.49</td>
<td>3.55</td>
<td>4.28</td>
</tr>
<tr>
<td>Validation</td>
<td>3.98</td>
<td>3.58</td>
<td>3.42</td>
<td>3.35</td>
<td>3.25</td>
<td>6.42</td>
<td>2.63</td>
<td>3.74</td>
<td>4.37</td>
</tr>
</tbody>
</table>

Weekly Peak Load Profile

The same set of techniques is applied to NTDC weekly peak load data, the graph shown in Fig. 5.6, shows comparison results obtained from above mentioned techniques. The line graph in Fig. 5.7, reveals how accurately ABC-GA is following the actual trend line, and how it improved results of original ABC algorithm. It is clear from the graph given in Fig. 5.7, that model based on GA based ABC-ANN outperformed others in accuracy. The MAPE obtained from all techniques is shown in Table 5.4. The same set of 9 models is applied on weekly peak demand data. Original ABC showed better results than PSO, SA, and other modified versions but ABC-GA showed even better results. Train, test and validation MAPE are summarized in Table 5.4. This comparison is also presented in bar graph in Fig. 5.8.

Table 5.4: Average MAPE over 10 runs of all Techniques on PEPCO Weekly Peak Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>ABC</th>
<th>ABC-Opp</th>
<th>ABC-Mod</th>
<th>ABC-GBest</th>
<th>ABC-Inv</th>
<th>GO-ABC</th>
<th>ABC-GA</th>
<th>PSO</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>3.36</td>
<td>3.87</td>
<td>3.81</td>
<td>3.84</td>
<td>3.88</td>
<td>5.47</td>
<td>3.18</td>
<td>4.25</td>
<td>4.47</td>
</tr>
<tr>
<td>Test</td>
<td>3.21</td>
<td>2.85</td>
<td>2.74</td>
<td>2.52</td>
<td>2.89</td>
<td>4.81</td>
<td>3.13</td>
<td>3.75</td>
<td>3.89</td>
</tr>
<tr>
<td>Validation</td>
<td>3.31</td>
<td>3.56</td>
<td>3.48</td>
<td>3.44</td>
<td>3.58</td>
<td>5.27</td>
<td>3.16</td>
<td>4.1</td>
<td>4.29</td>
</tr>
</tbody>
</table>
Figure 5.6: Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC

Figure 5.7: Actual demand vs Predicted by ABC and ABC-GA assisted ANN
Also comparison of MAPE shows test error is lower than training error, which suggests the appropriateness of models as there is no case of over-fitting in any model.

5.5.2 Case Study (LESCO Data)

This accumulative peak demand for daily and weekly for last 5 years is obtained from LESCO. It represents the daily and weekly peak demand profiles observed at Lahore region of Pakistan. Influencing factors considered for this study include: day of week, day type, week of year, month of year and weather conditions like temperature, humidity, and dew point at the time of peak load. All above mentioned techniques are applied on this data and forecasting accuracy of different techniques is shown by graphs.

Daily Peak Load Profile

All of the above mentioned techniques are applied on LESCO daily peak data sets. The predicted demand profiles of all techniques are presented by graph in Fig. 5.9, whereas, the comparison of actual load versus predicted values of ABC and ABC-GA is presented by graph in Fig. 5.10. The Results are summarized in Table 5.5, whereas, bar graph for this is presented in Fig. 5.11.

<table>
<thead>
<tr>
<th>Technique</th>
<th>ABC</th>
<th>ABC-Opp</th>
<th>ABC-Mod</th>
<th>ABC-GBest</th>
<th>ABC-Inv</th>
<th>GO-ABC</th>
<th>ABC-GA</th>
<th>PSO</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>4.4</td>
<td>4.29</td>
<td>4.3</td>
<td>4.28</td>
<td>4.42</td>
<td>5.72</td>
<td>4.42</td>
<td>4.66</td>
<td>4.99</td>
</tr>
<tr>
<td>Test</td>
<td>3.63</td>
<td>3.35</td>
<td>2.99</td>
<td>2.77</td>
<td>2.37</td>
<td>6.84</td>
<td>3.2</td>
<td>4.06</td>
<td>4.31</td>
</tr>
<tr>
<td>Validation</td>
<td>4.16</td>
<td>4.08</td>
<td>3.97</td>
<td>3.82</td>
<td>3.85</td>
<td>5.95</td>
<td>3.05</td>
<td>4.48</td>
<td>4.78</td>
</tr>
</tbody>
</table>

Weekly Peak Load Profile

The same set of techniques are applied on LESCO weekly peak data sets, results obtained are presented by graph in Fig. 5.12, whereas, close comparison of actual vs ABC and ABC-GA is presented in Fig. 5.13. Results obtained are summarized in Table 5.6, whereas, bar graph for this is presented in Fig. 5.14.
Figure 5.9: Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC

Figure 5.10: Actual demand vs Predicted by ABC and ABC-GA assisted ANN
Chapter 5. Improved ABC-GA-FFNN model

Figure 5.11: Comparison of MAPE obtained by all Models

Figure 5.12: Actual demand vs Predicted by ABC, PSO, SA, and Modified Versions of ABC

Table 5.6: Average MAPE over 10 runs of all Techniques on LESCO Weekly Peak Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>ABC</th>
<th>ABC-Opp</th>
<th>ABC-Mod</th>
<th>ABC-GBest</th>
<th>ABC-Inv</th>
<th>GO-ABC</th>
<th>ABC-GA</th>
<th>PSO</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>4.54</td>
<td>4.21</td>
<td>4.46</td>
<td>4.46</td>
<td>4.43</td>
<td>6.41</td>
<td>4.26</td>
<td>5.33</td>
<td>5.66</td>
</tr>
<tr>
<td>Test</td>
<td>4.97</td>
<td>4.42</td>
<td>3.83</td>
<td>3.36</td>
<td>3.55</td>
<td>6.07</td>
<td>3.02</td>
<td>4.89</td>
<td>5.33</td>
</tr>
<tr>
<td>Validation</td>
<td>4.5</td>
<td>4.43</td>
<td>4.27</td>
<td>4.13</td>
<td>4.16</td>
<td>6.3</td>
<td>3.88</td>
<td>5.19</td>
<td>5.56</td>
</tr>
</tbody>
</table>
Figure 5.13: Actual demand vs Predicted by ABC and ABC-GA assisted ANN

Figure 5.14: Comparison of MAPE Obtained by all Models
5.6 Conclusion

In this paper, we incorporated several modifications in base ABC algorithm to improve its exploitation aspect. Optimization ability of these modified versions is evaluated by statistical benchmark functions. Modifications are applied in series to obtain an integrated and evolved version of ABC. Further, GO’s are applied to witness its effect on optimization and forecasting accuracy. Base algorithm and all the evolved versions of ABC are used to develop LF model. Results obtained from these models are compared with each other and other modern techniques of optimization for ANN; for instance; PSO and SA. Obtained results portray that base algorithm shown better accuracy when modified in search and probability functions. However, GO’s could not improve results and computation time. Authors achieved a greater accuracy with MAPE of 2.49% which is quite remarkable and is comparable to any other model development technique for LF even on noisy data of such utility companies. Computation time of this approach is quite stumpy when compared to other competitive approaches used with ANN like PSO and SA.
Chapter 6

Performance Evaluation of Proposed ELF Models

An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.

---

6.1 Evaluation of ELF models

This chapter covers the modeling and evaluation of a number of available machine learning based LF models on different electric demand data sets. The characteristics of each data set are discussed. Moreover, models are compared on the basis of MAPE and accuracy of forecasts. The models are ranked on their optimization ability, capability to reach global minima and avoiding over-fitting of models. The results and performance comparisons are presented by tables and graphs. The best performing models are then selected to explore strategies to improve further by introducing several modifications in their processes and parameters, which is covered in subsequent chapters.

6.2 Modeling Techniques

In this study, ANN and SVM based models are evaluated based on their accuracy of results. Here, 10 optimization techniques are used to get the optimized set of ANN connection weights and best parameters for SVM based regression model. This makes 20 different forecasting models with the combination of optimization and modeling techniques, these are listed in Table 6.2. The diagram in Fig. 6.1, shows the essential components of this process. Here inputs (load affecting parameters such as: calendar attributes and hourly demands) are passed to LF modeling techniques and objective functions are optimized by learning schemes. Further, the results are evaluated and learning process is terminated upon achieving desired value of objective function.
6.2.1 ANN Model Structure

ANN model used in this study consists of 3 layers of neurons connected to next layer, namely input layer, hidden layer and output layer. Input layer consists of 9 neurons, 6 neurons in hidden layer, and one neuron in output layer. We have constituted feed-forward ANN models with one additional bias neuron at each layer. Thus, total weight connections becomes $85 = 10 \times 7 + 7 \times 2 + 1$. These weights are optimized by utilizing different learning schemes.

6.2.2 SVM Model

In this experimentation two types of SVM based regression models are used, namely Epsilon-SVR and NU-SVR. The RBF kernel is used in this study, its control parameters are tuned by PSO, SA, and default search method of SVM.

6.2.3 Optimization Techniques

The learning schemes include Back-propagation method; which is the default learning scheme of ANN, scaled conjugate gradient method (SCG), RPROP, LMA, SA, PSO, ABC, FFA, and default search method for SVM’s. The standard learning parameters for FFA, SCG, and SVM-Search method are used in this experimentation. The settings of RPROP, LMA, SA, PSO, and ABC are provided here.

**RPROP**

This technique has two learning parameters, which are: initial update value and step size. The initial update value for RPROP algorithm is 0.1 and maximum step size is 50.

**LMA**

The LMA has two parameters named as lambda scale and lambda maximum value. The initial value of lambda scale is 10.0, and maximum value is $1e^{25}$.

**SA**

This algorithm has two control parameters, initial temperature $T_1$ and stop temperature $T_2$. In this study, value of $T_1$ is 10.0, and $T_2$ is set to 0.1; which should always be $T_2 >= 0$. 
Chapter 6. Evaluation of ELF Models

PSO

The control parameters of PSO are population count \( P \), inertia weight \( W \), position limit \( L_1 \), velocity limit \( L_2 \), and two learning rate constants \( C_1, C_2 \). Here \( P \) is 25, \( W \) is 0.7, both \( C_1 \) and \( C_2 \) have the same value of 1.49. Here, range of \( L_1 \) and \( L_2 \) is \([-1.0, 1.0]\).

ABC

The number of food sources is 28 which is also the population count of honey bees. The count for each of employed and onlooker bees is 14. The dimension of the search space is equal to the 85 weight connections. The upper and lower bounds are \([-1; 1]\).

6.2.4 Performance Evaluation

All of the ANN and SVM based models are ranked on the same criterion, i.e. mean absolute percentage error (MAPE). It is widely used method to evaluate the accuracy forecasting model [29]. It represents the accuracy as percentage of error, and it is expressed by the formula given in Eq. 6.1.

\[
M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right |
\]  

(6.1)

Where \( n \) is total number of instances, \( A_t \) is the actual load value at time \( t \), and \( F_t \) is the forecasted value for the same time instance.

6.3 Evaluation on NTDCL Data

6.3.1 NTDCL Data Set

This experimentation is carried out on 6 years (2005 – 2010) hourly electricity demand data set obtained from National Transmission and Despatch Company Limited (NTDCL) of Pakistan. This is the recording of cumulative load of national grid, it represent the consumer demand of whole country, excluding Karachi city of Pakistan. Karachi electric supply company (K-Electric) is responsible for the electric power supply in Karachi region. Hourly data contains 24 demand data entries for each day. Consumer demand fluctuates on hourly basis throughout a day. Sample data of one day (January 14, 2010) is shown in Table 6.1, where, load affecting input parameters and hourly demand is presented. For STLF, weather is considered as most influencing parameter responsible for demand curve fluctuations. As, weather conditions are not same for whole country at the same time instance, due to geographical diversity of Pakistan. Therefore, weather is not considered in this case study. Other input parameters include calendar inputs (year, month, day of month, time of day, day of week, day type, etc.). Usually, there are two peaks and two valleys in everyday demand data. On the other hand, demand profile of working days is different from the demand profile of off days. As shown in Fig. 6.2, demand profile of three consecutive days is provided i.e. Saturday, Sunday, and Monday of (14, 15, 16) January 2010. It can be witnessed that peak and off peak loads appear at different hours of the day and demand profile is different for different days. Whole data
set is divided into three groups, for training, 70% of data is used; while remaining 30% is divided into two equal halves for testing and validation.

Table 6.1: Electricity Demand on January 14, 2010

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Week Day</th>
<th>Day Type</th>
<th>Hour</th>
<th>Demand (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>9019</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>9211</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>9213</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>9002</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>5</td>
<td>9226</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>10065</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>10718</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>8</td>
<td>10598</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>11211</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>11095</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>11</td>
<td>11061</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>12</td>
<td>10544</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>13</td>
<td>10666</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>14</td>
<td>10697</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>15</td>
<td>10625</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>16</td>
<td>10727</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>17</td>
<td>10695</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>18</td>
<td>11768</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>19</td>
<td>12128</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>20</td>
<td>12003</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>21</td>
<td>11859</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>22</td>
<td>11133</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>23</td>
<td>10379</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>14</td>
<td>5</td>
<td>1</td>
<td>24</td>
<td>9902</td>
</tr>
</tbody>
</table>

Figure 6.2: Electricity Demand profile of 3 consecutive days
6.3.2 Results

This experimentation is performed on real electricity demand data collected from NT-DCL. The results are obtained as average of 10 runs for each hybrid model. Accuracy comparison is presented in Table 6.2, which is compared on the same criteria by taking MAPE, time of execution (in seconds), and number of iterations. Training, testing, and validation MAPE is presented here. Results are categorized into three groups, i.e. ANN based models, Epsilon-SVR models, and NU-SVR models. It is evident that Epsilon-SVR models have consumed lesser time, but MAPE is very high. On the other hand, NU-SVR based models taken more time to train; still, these have shown better accuracy with lower MAPE. The MAPE is inversely proportional to accuracy of results, higher MAPE corresponds to lower accuracy. Out of NU-SVR models, NU-SVR-Search method have shown greater accuracy with only 0.60% of train MAPE. But the test MAPE and validation MAPE are higher than train MAPE, resulting in over-fitting of model. From the pool of ANN based models; RPROP with 5000 iterations and PSO with 10000 iterations have shown impressive results. Especially RPROP based ANN-FFB model have produced above 98% accurate results, it is efficient and avoided over-fitting of model. The MAPE comparison is also presented via bar graph in Fig. 6.3.

<table>
<thead>
<tr>
<th>Techniques (10 run)</th>
<th>Train Error</th>
<th>Test Error</th>
<th>Validation Error</th>
<th>Time (Sec)</th>
<th>Iterations</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-FFB-FFA</td>
<td>5.63</td>
<td>6.63</td>
<td>5.93</td>
<td>1173</td>
<td>5000</td>
<td>94.17</td>
</tr>
<tr>
<td>ANN-FFB-PSO5</td>
<td>2.95</td>
<td>3.26</td>
<td>3.04</td>
<td>237</td>
<td>5000</td>
<td>96.99</td>
</tr>
<tr>
<td>ANN-FFB-PSO3</td>
<td>3.81</td>
<td>4.44</td>
<td>4.00</td>
<td>158</td>
<td>3000</td>
<td>96.06</td>
</tr>
<tr>
<td>ANN-FFB-PSO10</td>
<td>2.46</td>
<td>2.81</td>
<td>2.57</td>
<td>515</td>
<td>10000</td>
<td>97.46</td>
</tr>
<tr>
<td>ANN-FFB-LMA</td>
<td>4.21</td>
<td>2.34</td>
<td>3.64</td>
<td>2006</td>
<td>5000</td>
<td>96.15</td>
</tr>
<tr>
<td>ANN-FFB-BACKPROP</td>
<td>7.65</td>
<td>9.05</td>
<td>8.07</td>
<td>58</td>
<td>5000</td>
<td>92.07</td>
</tr>
<tr>
<td>ANN-FFB-RPROP</td>
<td>1.55</td>
<td>2.11</td>
<td>1.72</td>
<td>57</td>
<td>5000</td>
<td>98.34</td>
</tr>
<tr>
<td>ANN-FFB-SCG</td>
<td>6.39</td>
<td>7.73</td>
<td>6.79</td>
<td>156</td>
<td>3000</td>
<td>93.34</td>
</tr>
<tr>
<td>ANN-FFB-ABC</td>
<td>5.09</td>
<td>7.97</td>
<td>6.59</td>
<td>389</td>
<td>5000</td>
<td>93.62</td>
</tr>
<tr>
<td>ANN-FFB-SA</td>
<td>2.82</td>
<td>3.43</td>
<td>3.00</td>
<td>1546</td>
<td>5000</td>
<td>97.06</td>
</tr>
<tr>
<td>EPSILON-SVR-Search</td>
<td>7.09</td>
<td>7.93</td>
<td>7.34</td>
<td>2</td>
<td>5</td>
<td>92.74</td>
</tr>
<tr>
<td>EPSILON-SVR-PSO20</td>
<td>4.71</td>
<td>6.57</td>
<td>5.27</td>
<td>42</td>
<td>30</td>
<td>94.92</td>
</tr>
<tr>
<td>EPSILON-SVR-PSO10</td>
<td>4.68</td>
<td>6.58</td>
<td>5.25</td>
<td>24</td>
<td>10</td>
<td>94.95</td>
</tr>
<tr>
<td>EPSILON-SVR-PSO5</td>
<td>5.50</td>
<td>6.20</td>
<td>5.71</td>
<td>4</td>
<td>5</td>
<td>94.36</td>
</tr>
<tr>
<td>EPSILON-SVR-SA</td>
<td>6.50</td>
<td>7.48</td>
<td>6.79</td>
<td>1</td>
<td>5</td>
<td>93.30</td>
</tr>
<tr>
<td>NU-SVR-PSO</td>
<td>2.30</td>
<td>3.49</td>
<td>2.72</td>
<td>160</td>
<td>1</td>
<td>97.39</td>
</tr>
<tr>
<td>NU-SVR-PSO5</td>
<td>2.67</td>
<td>4.46</td>
<td>3.21</td>
<td>500</td>
<td>5</td>
<td>96.98</td>
</tr>
<tr>
<td>NU-SVR-SA</td>
<td>2.17</td>
<td>3.16</td>
<td>2.47</td>
<td>357</td>
<td>5</td>
<td>97.63</td>
</tr>
<tr>
<td>NU-SVR-Search</td>
<td>0.60</td>
<td>3.19</td>
<td>1.35</td>
<td>109</td>
<td>1</td>
<td>98.91</td>
</tr>
<tr>
<td>NU-SVR-Search5</td>
<td>2.14</td>
<td>3.14</td>
<td>2.44</td>
<td>494</td>
<td>5</td>
<td>97.66</td>
</tr>
</tbody>
</table>

Four days demand profile has been presented in Fig. 6.4, the actual demand curve is plotted against the curves produced by three best performing models. It can be witnessed that the forecast curves are closely following the actual demand curve. From the results presented in Table 6.2, we can observe that all of the models have produced more than 92% accurate results on NTDCL data set. ANN-FFB-RPROP and NU-SVR-Search based models appear as competitive approaches in this scenario. Hybrid model of feed-forward ANN with bias and trained with RPROP method emerged as most suitable approach for STLF of this kind.
6.4 IESO: Ontario Data Set

This experimentation is performed on 10 years (2002-2012) hourly electricity load demand data sets obtained from Independent Electricity System Operator (IESO) for Ontario State. This data is publicly available and is highly accurate. Based upon high correlation, different input factors like calendar events, demand of last hour, and previous day, weather conditions, and hourly energy price are considered as input parameters in this study. In order to achieve the desired objective of accurate STLF model, it is necessary to compare all three models on same criteria, for that reason data and features sets are identical for all models. For training the neural network models 70% of data is utilized and 30% is used for testing and validation by further dividing the data into two equal parts.
6.4.1 Experiment Setup

This case study covers the simulation of load forecasting techniques on IESO data set. The purpose of this experimentation is to cross validate the forecasting capabilities of different techniques to predict hourly consumption patterns for next few days.

Data Set

Historical demand data was obtained from IESO. This data set contains the cumulative hourly load consumption data for the 11 years from year 2002 to 2012. It contains 24 hourly data entries for a single day. Several other factors have been considered that may affect the load profile fluctuation. Factors affecting short term load forecasting include: Hourly Electricity Price, Weather conditions such as: Temperature, Dew Point, Humidity, etc., Day of Week, Time of the day, and Day type.

6.4.2 Results

The results are obtained as average of 10 runs for each hybrid model. Accuracy comparison is presented in Table 6.3. As, the accuracy is inversely proportional to MAPE of models. Therefore, only the overall percentage accuracy of results is presented. Feed-Forward ANN model is trained by LMA, SA, SCG, RPROP, QN, PSO, and ABC algorithms. Whereas, SVM model is trained with PSO and SA algorithms. Out of these hybrid models, ABC based FFB-ANN model has outperformed others with respect to accuracy and obtained 98.24% accuracy of forecasts. This model is proposed in Chap. 3. Here, we are presenting the comparison of PSO and SA based SVM and ANN models with line graphs. The line graph presented in Fig. 6.5 compares the actual and predicted curves of ANN-FFB-PSO and SVM-PSO models. Whereas, the line graph presented in Fig. 6.6 compares the actual and predicted curves of ANN-FFB-SA and SVM-SA models. In both cases, SVM based models have achieved better accuracy and closely following the actual demand curve.

Table 6.3: Performance Comparison of Models on IESO data

<table>
<thead>
<tr>
<th>Technique</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-FFB-LMA</td>
<td>97.76</td>
</tr>
<tr>
<td>ANN-FFB-SA</td>
<td>97.21</td>
</tr>
<tr>
<td>ANN-FFB-SCG</td>
<td>96.96</td>
</tr>
<tr>
<td>ANN-FFB-RPROP</td>
<td>98.14</td>
</tr>
<tr>
<td>ANN-FFB-QN</td>
<td>95.35</td>
</tr>
<tr>
<td>ANN-FFB-PSO</td>
<td>97.30</td>
</tr>
<tr>
<td>ANN-FFB-ABC</td>
<td>98.24</td>
</tr>
<tr>
<td>SVM-PSO</td>
<td>98.05</td>
</tr>
<tr>
<td>SVM-SA</td>
<td>98.08</td>
</tr>
</tbody>
</table>
Chapter 6. Evaluation of ELF Models

Figure 6.5: Actual v/s Forecasted by ANN-FFB-PSO and SVM-PSO

Figure 6.6: Actual v/s Forecasted by ANN-FFB-SA and SVM-SA
6.5 Medium Term Load Forecasting: Weekly Peak profile

This experimentation is performed to predict weekly peak demand of NTDCL system. Weekly peak demand data is extracted from hourly peak data for years (2007 – 2011). Different techniques are evaluated against their forecasting accuracy and efficiency. The factors considered to predict weekly peak demand are: Year, Month, Week number of the year, Weekly peak load (historical). Table 6.4, shows the peak demand data on weekly basis of 4 months.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Week Number</th>
<th>Weekly Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>12</td>
<td>3</td>
<td>12152</td>
</tr>
<tr>
<td>2007</td>
<td>12</td>
<td>4</td>
<td>11770</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>1</td>
<td>12174</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>2</td>
<td>11735</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>3</td>
<td>11920</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>4</td>
<td>12255</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>5</td>
<td>12098</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>1</td>
<td>11992</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>2</td>
<td>12123</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>3</td>
<td>11960</td>
</tr>
<tr>
<td>2008</td>
<td>2</td>
<td>4</td>
<td>12018</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>1</td>
<td>12093</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>2</td>
<td>12410</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>3</td>
<td>12958</td>
</tr>
<tr>
<td>2008</td>
<td>3</td>
<td>4</td>
<td>13682</td>
</tr>
<tr>
<td>2008</td>
<td>4</td>
<td>1</td>
<td>13635</td>
</tr>
<tr>
<td>2008</td>
<td>4</td>
<td>2</td>
<td>13463</td>
</tr>
</tbody>
</table>

6.5.1 Results

The results are obtained as average of 10 runs for each hybrid model. Accuracy comparison is presented in Table 6.5. As, the accuracy is inversely proportional to MAPE of models. Therefore, only the overall percentage accuracy of results is presented. Feed-Forward ANN model is trained by LMA, SA, SCG, and RPROP, algorithms. Whereas, SVM model is trained with greedy search algorithm. Out of these hybrid models, ANN-SCG model has outperformed others with respect to accuracy and obtained 97.56% accuracy of forecasts. Here, we are presenting the comparison of SVM and ANN models with line graphs. The graph presented in Fig. 6.7 compares the actual and predicted curves of ANN-LMA model. Whereas, the line graph presented in Fig. 6.8 compares the actual and predicted curves of ANN-RPROP model. Overall, data of 60 consecutive weeks is presented. The actual weekly demand line for first 45 weeks is shown and is followed by predicted line. Whereas, predicted demand of next 15 weeks is forecasted by each model and is presented in Fig. 6.8, 6.9, 6.10, and 6.11. However, the comparison of all techniques is presented in Fig. 6.12.
Table 6.5: Accuracy Comparison on Weekly Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-SA</td>
<td>96.663</td>
</tr>
<tr>
<td>ANN-RPROP</td>
<td>97.394</td>
</tr>
<tr>
<td>ANN-SCG</td>
<td>97.566</td>
</tr>
<tr>
<td>ANN-LMA</td>
<td>96.754</td>
</tr>
<tr>
<td>Search-SVM</td>
<td>96.94</td>
</tr>
</tbody>
</table>

Below graphs depicts the overall comparison of all techniques applied on workday hourly data for NTDCL system. Graph lines have been plotted for each technique,
Figure 6.9: Actual v/s Forecasted by ANN-SA

Figure 6.10: Actual v/s Forecasted by ANN-FFB-SCG
Chapter 6. Evaluation of ELF Models

Figure 6.11: Actual v/s Forecasted by SVM-GreedySearch

Figure 6.12: Actual v/s Forecasted by all techniques
6.6 Long Term Load Forecasting: Industrial Profile

In this research work, different techniques have been applied to solve the long term load estimation problem of industrial sector. Based upon data sets provided by National Transmission and Dispatch Company Limited (NTDCL) of Pakistan, we have developed different models with varying predictive performance and accuracy of results. Models are repeatedly evolved unless acceptable results are obtained. This work would help in power planning section of NTDCL to make timely decision to fulfill energy requirement and to help in survival and growth of industrial sector.

The purpose of this experimentation is to determine the forecasting capabilities of different techniques to predict yearly peak load consumption patterns. This is a case of consumer type demand forecasting, in which cumulative load of industrial consumers is predicted.

6.6.1 Data Set

The yearly load profile of industrial consumers is obtained from NTDCL. The input factors include: year wise industrial consumers count, industrial tariff, industrial usage history, and industrial gross domestic product (GDP). This data set contains the yearly industrial demand data for 41 years from 1970 to 2010. Yearly peak data contains only one entry for a single year. Industrial consumption recording is based on the number of units consumed per year by industrial users. Several other factors have been considered that may affect the load profile fluctuation. Factors affecting long term industrial load forecasting include: year, industrial tariff, industrial customer count, and industrial GDP. Original data of industrial load for 40 years is shown in Table 6.6. It contains yearly demand along other affecting parameters industrial tariff and customer count.

6.6.2 Results

The actual demand curve and forecasted curve obtained from ANN-RPROP model is plotted in Fig. 6.13, and forecasted curve produced by PSO based ANN model is shown in Fig. 6.14. The comparison of forecasted curves from both techniques with actual load curve is shown in Fig. 6.15. The accuracy comparison is presented in Table 6.7. It is clear that both techniques have produced competitive accuracy.
**Table 6.6: Industrial consumption data with input factors**

<table>
<thead>
<tr>
<th>Year</th>
<th>IndTariff</th>
<th>Ind-Cust</th>
<th>Ind-Demand (GWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>12.3</td>
<td>61330</td>
<td>1646</td>
</tr>
<tr>
<td>1971</td>
<td>13</td>
<td>64494</td>
<td>1755</td>
</tr>
<tr>
<td>1972</td>
<td>13.8</td>
<td>67056</td>
<td>2109</td>
</tr>
<tr>
<td>1973</td>
<td>14.3</td>
<td>72158</td>
<td>2222</td>
</tr>
<tr>
<td>1974</td>
<td>17.6</td>
<td>78277</td>
<td>2251</td>
</tr>
<tr>
<td>1975</td>
<td>21.2</td>
<td>80733</td>
<td>2244</td>
</tr>
<tr>
<td>1976</td>
<td>28.1</td>
<td>84250</td>
<td>2261</td>
</tr>
<tr>
<td>1977</td>
<td>33.5</td>
<td>91365</td>
<td>2295</td>
</tr>
<tr>
<td>1978</td>
<td>37.4</td>
<td>95036</td>
<td>2596</td>
</tr>
<tr>
<td>1979</td>
<td>46.2</td>
<td>100946</td>
<td>2770</td>
</tr>
<tr>
<td>1980</td>
<td>57.1</td>
<td>107168</td>
<td>3154</td>
</tr>
<tr>
<td>1981</td>
<td>63.1</td>
<td>111484</td>
<td>3482</td>
</tr>
<tr>
<td>1982</td>
<td>67.7</td>
<td>115833</td>
<td>3960</td>
</tr>
<tr>
<td>1983</td>
<td>75.5</td>
<td>119464</td>
<td>4417</td>
</tr>
<tr>
<td>1984</td>
<td>76.5</td>
<td>123508</td>
<td>4708</td>
</tr>
<tr>
<td>1985</td>
<td>78.5</td>
<td>128441</td>
<td>5061</td>
</tr>
<tr>
<td>1986</td>
<td>92</td>
<td>133573</td>
<td>5894</td>
</tr>
<tr>
<td>1987</td>
<td>89.1</td>
<td>139537</td>
<td>6436</td>
</tr>
<tr>
<td>1988</td>
<td>111.1</td>
<td>147439</td>
<td>7236</td>
</tr>
<tr>
<td>1989</td>
<td>133</td>
<td>153042</td>
<td>7579</td>
</tr>
<tr>
<td>1990</td>
<td>150.3</td>
<td>158800</td>
<td>8360</td>
</tr>
<tr>
<td>1991</td>
<td>165.5</td>
<td>162624</td>
<td>9114</td>
</tr>
<tr>
<td>1992</td>
<td>188.8</td>
<td>169436</td>
<td>10213</td>
</tr>
<tr>
<td>1993</td>
<td>198.9</td>
<td>172145</td>
<td>10912</td>
</tr>
<tr>
<td>1994</td>
<td>229</td>
<td>174577</td>
<td>10532</td>
</tr>
<tr>
<td>1995</td>
<td>268.2</td>
<td>179392</td>
<td>10603</td>
</tr>
<tr>
<td>1996</td>
<td>336.3</td>
<td>181092</td>
<td>10330</td>
</tr>
<tr>
<td>1997</td>
<td>374.5</td>
<td>184301</td>
<td>10116</td>
</tr>
<tr>
<td>1998</td>
<td>410.9</td>
<td>186539</td>
<td>10238</td>
</tr>
<tr>
<td>1999</td>
<td>448.4</td>
<td>190084</td>
<td>9945</td>
</tr>
<tr>
<td>2000</td>
<td>416.3</td>
<td>194566</td>
<td>10772</td>
</tr>
<tr>
<td>2001</td>
<td>415.7</td>
<td>195511</td>
<td>11744</td>
</tr>
<tr>
<td>2002</td>
<td>418.7</td>
<td>199849</td>
<td>12637</td>
</tr>
<tr>
<td>2003</td>
<td>442</td>
<td>206336</td>
<td>13462</td>
</tr>
<tr>
<td>2004</td>
<td>446</td>
<td>210296</td>
<td>14476</td>
</tr>
<tr>
<td>2005</td>
<td>425</td>
<td>212233</td>
<td>15567</td>
</tr>
<tr>
<td>2006</td>
<td>425</td>
<td>222283</td>
<td>16596</td>
</tr>
<tr>
<td>2007</td>
<td>517</td>
<td>233162</td>
<td>17603</td>
</tr>
<tr>
<td>2008</td>
<td>568</td>
<td>242401</td>
<td>17299</td>
</tr>
<tr>
<td>2009</td>
<td>748</td>
<td>253089</td>
<td>16035</td>
</tr>
<tr>
<td>2010</td>
<td>894</td>
<td>263507</td>
<td>16372</td>
</tr>
</tbody>
</table>
Table 6.7: Performance Comparison on Industrial Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN-RPROP</td>
<td>99.076</td>
</tr>
<tr>
<td>ANN-PSO</td>
<td>99.647</td>
</tr>
</tbody>
</table>

Figure 6.13: Actual v/s Forecasted by ANN-RPROP

Figure 6.14: Actual v/s Forecasted by ANN-PSO
Chapter 6. Evaluation of ELF Models

Figure 6.15: Actual v/s Forecasted by ANN-PSO and ANN-RPROP
6.7 Evaluation of Proposed Algorithms on Classification Problems

Machine learning algorithms were from the very beginning designed and used to analyze medical datasets. Today, several indispensable tools based on machine learning are being used for intelligent data analysis [122, 123]. The use of data mining approaches is rapidly increasing in medical diagnostic systems due to effectiveness of such approaches in classification and decision making systems. Data mining and machine learning techniques can be used to support medical decisions [122]. An overview of data analysis techniques in medical diagnostics from a machine learning perspective is presented in [123]. Different classifiers and their performance comparisons is provided with respect to appropriateness for medical diagnostic and prognostic problems. Further, these algorithms are applied to diagnose the ischaemic heart disease [123]. The heart disease diagnosis by support vector machines (SVM) and sequential minimization optimization learning algorithm is presented in [124]. Cancer diagnosis is most emerging clinical application of diagnosis methods. A comprehensive evaluation of different classification methods on cancer diagnosis is presented in [125]. Multi-category SVM's were found the most effective classifier and outperformed other machine learning algorithms, such as: ANN and K-nearest neighbors (KNN) based classifiers [125]. This experimentation presents a comparison among different evolutionary optimization algorithms based artificial neural network (ANN) classifiers for the diagnosis of diseases. The experimentation is performed on four diseases data set obtained from UCI machine learning repository. Five learning schemes are used for multi-layer ANN training, which include artificial bee colony (ABC), cuckoo search (CS), particle swarm optimization (PSO), and artificial cooperative search (ACS). We compare the accuracy of classification of all techniques for each disease data set to get the most suitable classifier for each disease data set. The experimental results show that hybrid of multi-layer feed-forward ANN with PSO is superior to the other algorithms for training ANN.

Here, ACS algorithm is used in order to train neural network and its performance is compared with the popular evolutionary algorithms like GA, PSO, CS, differential search algorithm (DSA), and ABC on some known classification problems from medical diagnosis of diseases.

The objective of this study is to analyze the performance of ACS on well-known classification problems. Further, a comparative investigation of the successes stated in [76] is also carried out. For this, four classification problems, discussed in PROBEN1 [77] are used. PROBEN1 is a set of ANN benchmark problems and rules to evaluate the performance of algorithms.

6.7.1 Data Description

In order to evaluate the performance of ACS against PSO, ABC, CS, DSA the following classification problems are used.
Chapter 6. Evaluation of ELF Models

Cancer
Diagnosing the breast cancer. This dataset consists of 9 inputs, 2 outputs, and 683 examples. This dataset is obtained from UCI repository of machine learning databases. The purpose is to classify tumor as either benign or malignant based on cell description obtained from microscopic examination. The dataset is linearly transformed into range 0 and 1 and class labels are encoded as 0 for benign and 1 for tumor.

Heart
Predicting the heart disease. Verdict whether any of four major vessels has condensed its diameter by more than 50%. This binary decision is based on several factors of each individual, such as: description of particular type of patient pain, smoking habits, gender, age, and results of different medical test such as: electro cardiogram and blood pressure results. This data set is obtained from UCI repository, containing 35 inputs, 2 outputs, and 920 samples.

Diabetes
Diagnosing diabetes of Pima Indians. This dataset is based on personal medical examination try to predict whether a Pima Indian individual is diabetes positive or not. The dataset consists of 8 inputs, 2 outputs, and 768 examples. This dataset is obtained from UCI repository of machine learning databases. This dataset is also transformed as above datasets.

Horse
Envisage the fate of a horse with a colic. Predicting whether the horse will be euthanized, will survive, or will die is based on the results of a veterinary examination. This dataset is also obtained from UCI repository of machine learning dataset, it contains 58 inputs, 3 outputs, and 364 examples. The horse survived in 62% of the examples, in 24% it died, and in 14% it was euthanized. This data contains missing values (about 30% overall of the original attribute values).

6.7.2 Algorithm Parameters

ACS Settings: For evaluating ACS against other algorithms following parameters are used in this study. Its population size is 30 and maximum iteration count is 3000.

CS Settings: CS has two control parameters for its learning process, in this experimentation $\beta = 1.50$ and $p0 = 0.25$ are used as recommended in [78].

PSO Settings: The control parameters of PSO include, population count $P$, inertia weight $W$, position limit $L1$, velocity limit $L2$, and two learning rate constants $C1$, $C2$. Here $P$ is 30, $W$ is 0.6, both $C1$ and $C2$ have the same value of 1.80. Here range of $L1$ and $L2$ is $[-1.0, 1.0]$. These parameter settings are recommended in [64].
**ABC Settings:** The number of food sources is 28 which is also the population count of honey bees. The count for each of employed and onlooker bees is 14. The dimension $D$ of the search space is 50 as recommended in [64]. The upper and lower bounds are $[-1.0, 1.0]$.

**DSA Settings:** DSA has two control parameters, $p_1$ and $p_2$. Both have the same value of $3 \times \text{rand}$.

### 6.7.3 ANN Architecture

For comparison 30 runs of each algorithm are made on fixed neural network architecture for a same problem. MSE is used in order to evaluate the ANN performance.

### 6.7.4 Stopping Criteria

During the training of ANN the following stopping criteria is used

1. Stop when the number of epochs reaches 3000.
2. Stop if $\frac{GL(t)}{P_a(t)} > a$ where $a$ is the training strip length and $GL(t)$ is the generalization loss on validation set.

### 6.7.5 Results and Discussion

This section discusses and compares the successes of ACS algorithm with four other computational intelligence based algorithms (i.e. PSO, DSA, CS, and ABC) for numerical optimization problem solving. The algorithms are evaluated on four different real-world problems; Cancer, Heart, Horse, and Diabetes problems. These data sets are obtained from UCI repository of machine learning data sets [126]. Each algorithm is executed 30 times to train ANN. Results are presented as a mean of 30 runs. In order to reveal the optimization capability of each algorithm to reach global minimum, several performance measure of great importance are used i.e. mean epochs of training, mean training error, mean, max, and min classification rates. Mean of these performance measures is also provided next section.

**Cancer Data**

The empirical results of five optimization techniques to build ANN based hybrid models for classification problem tested on cancer data set are presented in Table 6.8. The results are presented as a mean of 30 runs of each algorithm. Since, cancer data has 9 input factors, therefore, our ANN has 9 neurons in input layer, 8 in middle layer, and 2 in output layer. Hyperbolic tangent sigmoid transfer function (tansig) is a neural transfer function. Which is used in this ANN to calculate outputs from both input layers. Mean training error represents the network objective function optimization. Mean epochs shows the ACS has reached its global minimum value in lesser epochs. Still, mean classification rate of other algorithms are better than ACS algorithm. However, the max classification rate is same for all the techniques.
Heart Data
Similarly, the same pool of algorithms are used to train ANN for classification problem of heart disease data set. This data set has 35 input factors. Here, ANN structure contains 35 neurons at input layer, 8 in middle layer and 2 at classification output layer. Soft Max (softmax) is a neural transfer function, which is used here to calculate outputs from both input layers. Results are presented in Table 6.9. Mean training error of 30 runs is given, PSO attained lowest value for this problem. Whereas, CS reached its global optimum in lowest mean epochs. PSO also performed well on mean and max classification rates.

Diabetes Data
Diabetes data set contains 8 input factors, ANN designed for this classification problem has 3 layers. There are 8 neuron at input layer, 4 at second, and 2 at classification output layer. Softmax transfer function is used here to calculate outputs of each layer. Results obtained are presented in Table 6.10. as a mean of 30 runs for each training algorithm, which shows PSO performed better in classification and achieved lowest training error.

Horse Data
This classification problem data set contains high rate of missing values (30%). It has 58 input factors and 3 outputs. The ANN structure for this problem comprise of 58 neurons in input layer and 4 at second and third layer, and 3 output neurons. Here, transig transfer function is used at output layer. The results are presented in Table 6.11, where ABC algorithm outperformed other techniques in terms of mean training error and classification rates.

| Table 6.8: Results on Cancer data |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Learning Scheme | PSO             | ACS             | DS              | CS              | ABC             |
| Mean Training Error | 0.047           | 0.07            | 0.063           | 0.055           | 0.043           |
| Mean EPOCHS      | 186.733         | 135.216         | 216.25          | 176.1           | 427.083         |
| Mean Classification Rate | 98.208         | 96.667          | 97.126          | 97.586          | 98.247          |
| Min Classification Rate | 94.252         | 89.655          | 88.505          | 93.103          | 97.126          |

| Table 6.9: Results on Heart data |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Learning Scheme | PSO             | ACS             | DS              | CS              | ABC             |
| Mean Training Error | 0.097           | 0.125           | 0.116           | 0.118           | 0.107           |
| Mean EPOCHS      | 999             | 666.4           | 997.3           | 656.7           | 986.7           |
| Mean Classification Rate | 80              | 77.913          | 79.087          | 78.347          | 79.391          |
| Max Classification Rate | 82.173          | 80.434          | 80.869          | 80              | 80.434          |
| Min Classification Rate | 75.217          | 75.217          | 76.956          | 76.956          | 77.826          |
Table 6.10: Results on Diabetes data

<table>
<thead>
<tr>
<th>Learning Scheme</th>
<th>PSO</th>
<th>ACS</th>
<th>DS</th>
<th>CS</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Training Error</td>
<td>0.516</td>
<td>0.574</td>
<td>0.558</td>
<td>0.53</td>
<td>0.518</td>
</tr>
<tr>
<td>Mean EPOCHS</td>
<td>743</td>
<td>787.8</td>
<td>1000</td>
<td>517.6</td>
<td>905.6</td>
</tr>
<tr>
<td>Mean Classification Rate</td>
<td>71.093</td>
<td>65.468</td>
<td>67.187</td>
<td>69.531</td>
<td>70.989</td>
</tr>
<tr>
<td>Max Classification Rate</td>
<td>71.875</td>
<td>68.75</td>
<td>69.27</td>
<td>73.437</td>
<td>72.395</td>
</tr>
<tr>
<td>Min Classification Rate</td>
<td>69.27</td>
<td>63.541</td>
<td>64.583</td>
<td>66.667</td>
<td>68.229</td>
</tr>
</tbody>
</table>

Table 6.11: Results on Horse data

<table>
<thead>
<tr>
<th>Learning Scheme</th>
<th>PSO</th>
<th>ACS</th>
<th>DS</th>
<th>CS</th>
<th>ABC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Training Error</td>
<td>0.516</td>
<td>0.767</td>
<td>0.737</td>
<td>0.626</td>
<td>0.602</td>
</tr>
<tr>
<td>Mean EPOCHS</td>
<td>998.2</td>
<td>601.3</td>
<td>967.2</td>
<td>859.1</td>
<td>1000</td>
</tr>
<tr>
<td>Mean Classification Rate</td>
<td>69.56</td>
<td>69.12</td>
<td>68.241</td>
<td>68.241</td>
<td>70.989</td>
</tr>
<tr>
<td>Max Classification Rate</td>
<td>72.527</td>
<td>73.626</td>
<td>75.824</td>
<td>75.824</td>
<td>75.824</td>
</tr>
<tr>
<td>Min Classification Rate</td>
<td>63.736</td>
<td>65.934</td>
<td>64.835</td>
<td>61.538</td>
<td>64.835</td>
</tr>
</tbody>
</table>

6.8 Conclusion

In this study, we contributed by evaluating the performance of leading machine learning techniques to propose a best suitable STLF model for the NTDCL. We have compared ANN and SVM based 20 hybrid models, wherein, nine different optimization techniques are utilized. We propose an accurate ELF model formulation that is fit for indigenous attributes and can fulfill the requirements of NTDCL. Using SVM based models, we have achieve 98.91% accurate forecasts. Whereas, ANN based models have proved more successful to produce comparative accuracy and to avoid over-fitting. This formulation has satisfied our objectives to produce the accurate results of forecast by using state of the art machine learning model development and optimization techniques. This experimentation was performed on cumulative demand data at national level, collected from NTDCL [127]. Similarly, case studies for MTILF and LTILF with higher accuracy are included [67]. Also, the proposed models are applied on classification data set of medical diagnostic domain to cross verify our developed models. In future, we propose to assimilate other power supply companies working at regional level in Pakistan. We aim to take into account medium term (daily, weekly, monthly) and long term (yearly) forecasts on regional and national level.
Chapter 7

A Generalized Load Forecasting Methodology

All generalizations are false, including this one.

Mark Twain

7.1 Introduction

Electric load forecasting (LF) involves the projection of peak demand levels and overall energy consumption patterns to support an electric utility’s future system and business operations. Short and mid-range predictions of electricity load allow electricity companies to retain high energy efficiency and reliable operation. Absence of such prior planning results in a current crisis like situation in Pakistan, where power generation is not up to the mark, its fallout is forced load shedding and voltage instability. To solve the problem of accurate LF, a variety of models is reported in literature. However, the accuracy of modeling techniques is extremely dependent on quality of historical data. Since, recording of data in Pakistan power systems was previously manual. This data contained abnormalities like missing values, outliers, and duplication of records.

The Machine learning algorithms assist not only in model development for forecasting but it also helps in data pre-processing; for example, to fill missing values, detect outliers, and remove duplicated values. Since, the data collection and recording was manual in Pakistan power systems (PPS). There are many human errors while recording data. Similarly, data contains missing records, and duplication of data is often the case. The accuracy of forecasting models heavily depend on the quality of data under study [3]. Observing the aforementioned problems with PPS and scrutinizing the issues with recorded load data, we got motivation to solve the problem of accurate LF for noisy data of indigenous power sector. The hypothesis behind this study is that modeling techniques can perform well on our datasets, when the data is smooth and free of abnormalities.
In order to achieve better accuracy on noisy data sets; this study involves investigation through different statistical and heuristic techniques to suggest an accurate load forecasting model that can fulfill the above mentioned requirements of PPS. For this, we have incorporated machine learning techniques for model development, optimization, and data pre-processing. An analytical study is performed on varying behavior of different techniques on predicating futuristic load utilizing different combination of techniques. In this way authors have modeled the complex behavior between different input factors and consumption behavior to predict future loads. In this chapter, we are formulating a STLF model for the PPS by integrating machine learning techniques. We present a customized LF model formulation that is fit for indigenous attributes and can fulfill the requirements of our power industry. The purpose of this formulation is to produce the accurate results of forecast on noisy data sets with appropriate use of machine learning techniques for data pre-processing, analysis, and model development. This formulation will also help to resolve the challenges listed in Chapter 1 and 2.

In this chapter, a customized LF model formulation is being presented, which incorporates machine learning techniques for data pre-processing, analysis, and model development.

### 7.2 Components of LF Model

The goal of accurate demand forecasting can only be achieved, when data is free from abnormalities before passing it to prediction models. The whole process is divided into four steps, which are input pre-processing, feature selection, normalization, and result interpretation. In order to achieve the desired objective of accurate results of forecasting on noisy data, machine learning techniques are applied on every step of proposed system. The diagram in Fig. 7.1, shows the essential components of this process. These steps are further elaborated in subsequent sections.

![Figure 7.1: Components of LF model](image)

#### 7.2.1 Data Pre-processing and Analysis

Data pre-processing is first and foremost step while building LF models. Pre-processing is applied on data under study to filter outliers, missing, and duplicated values. This effort is carried out to remove irregularities and smooth the load curve, ultimately resulting in more accuracy of forecasts. The pre-processing components are shown in Fig. 7.2, and discussed in subsequent sections.
7.2.2 Outlier Detection

An outlier is defined as an observation that "appears" to be inconsistent with other observations in the data set [128]. In the case of demand data, an addition or removal of one digit can change the demand graph drastically. In this study, different techniques are analyzed to process outliers, including Box Plot, Z-Score, modified Z-Score, 2-Sigma [128]. Though, Box Plot method produced most accurate results.

7.2.3 Treating Missing Values

Missing data is one of the major issues in load forecasting. Missing data arise in almost all serious statistical analyses. Filling missing values is called interpolation, and is a type of regression to treat time series data. After careful observation of LESCO load data, many records are found null or either filled with value '0'. There are various methods for interpolation, including some relatively simple approaches that can often yield reasonable results. In this study, weighted moving average (WMA), regression models, kalman filter, and exponential smoothing methods are utilized [129]. However, WMA and regression models produced more promising results.

7.2.4 Duplicate Entries

Duplication in data is removed by taking average of all values recorded for the same instance of time. This process is performed after removing outliers and missing values. The aforementioned pre-processing methods are applied on load and weather data of LESCO. The graph in Fig. 7.3, shows the five day profile of actual load data before processing and resultant data curve after processing. Actual load data contains missing values represented by '0' and negative values detected as outliers. In the processed curve,
both missing and outliers are smoothed by applying filters. Weather data is processed in similar fashion to remove anomalies.

![Five Day Profile of Load Data with Noise and after Pre-Processing](image)

**Figure 7.3:** Five Day Profile of Load Data with Noise and after Pre-Processing

### 7.2.5 Correlation Matrix

Correlation is a measure of the extent to which corresponding elements from two sets of ordered data are linked together. It is measured by correlation coefficient, which ranges between \(-1\) and \(+1\). It is positive when the values increase together, and it is negative when one value decreases as the other increases. In the case of LF, it helps detecting the most influencing factors on electric load. Such factors are further selected for model development. The formula given in Eq. 7.1, is used to compute correlation coefficient \(r\) of two data series \(X, Y\). In Table 7.1, correlation of different factors with demand curve is provided. It is evident that temperature, time of day, and previous load values have high correlation with current demand. This exercise on each data set can help to obtain most influencing factors affecting load curve variation.

\[
    r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} 
\]

(7.1)

Where, \(n\) is the number of pairs of data.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Hour of day</th>
<th>Temperature</th>
<th>Dew</th>
<th>Humidity</th>
<th>Weekday</th>
<th>Previous Hour</th>
<th>Previous Day</th>
<th>Previous Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation with load</td>
<td>0.582</td>
<td>0.632</td>
<td>0.538</td>
<td>0.313</td>
<td>0.611</td>
<td>0.967</td>
<td>0.830</td>
<td>0.693</td>
</tr>
</tbody>
</table>

### 7.2.6 Over-Fitting

Over-fitting of models can be avoided by many factors: size of training data, model structure, and cross-validation. The larger the train set, the less chances of over-fitting. The model parameters should be carefully selected to achieve better mapping of inputs to outputs. This could be achieved by appropriate selection of neuron connection weights in case of ANN and parameters for the SVM. We have utilized many optimization algorithms from statistical and artificial intelligence domain.
7.2.7 Cross Validation

The cross-validation also called rotation estimation is a statistical method to estimate the performance of a predictive model on independent data. This is a model validation technique to assess its generalization ability of models. To perform the cross-validation, data is spitted into three groups: train, test, and validate. The data may be linearly divided or randomly selected for three subsets. We have used the random and sequential cross-validation techniques for data division.

7.2.8 Performance Evaluation

All of the ANN and SVM based models are ranked on the same criterion, i.e. Mean Absolute Percentage Error (MAPE). It is mostly widely used method to evaluate the accuracy forecasting model [130]. It represents the accuracy as percentage of error. Which is expressed by the formula given in Eq. 7.2.

\[
M = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|
\]  

(7.2)

Where \( n \) is total number of instances, \( A_t \) is the actual load value at time \( t \), and \( F_t \) is the forecasted value for the same time instance.

7.3 Application on Noisy Data Set

This section discuss the application of our proposed methodology on noisy data of one utility. However, the same procedure is adopted for development and testing of all models discussed in previous chapters. This study was carried out on 4 years (2008 – 2011) hourly and 6 years (2005 – 2011) daily peak electricity demand data sets obtained from Lahore Electric Supply Company (LESCO) (http://lesco.gov.pk). Hourly data contained 24 data entries for a single day. Whereas, daily peak data contained only a single entry for each day. As, energy demand on short term is highly dependent on weather conditions. Sudden change in temperature causes fluctuation on demand graph. Based upon high correlation, different input factors like calendar events, demand of last hour, and previous day, and weather conditions were considered as input parameters in this study. In order to achieve the desired objective of accurate LF model, it is necessary to compare all models on same criteria, for this reason data and features sets were identical for all models. For training the LF models, 70% of data was utilized and remaining 30% was divided into two equal halves for testing and validation. Further, the data was normalized before passing to modeling techniques. Here, min-max normalization method was used to scale attribute data set into range \([0, 1]\) [131]. This is defined by the formula given in Eq. 7.3.

\[
\hat{x} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}
\]  

(7.3)
Where \( \hat{x} \) is the normalized value of \( x \), \( x_{\text{min}} \) and \( x_{\text{max}} \) are the minimum and maximum values of \( x \), respectively.

### 7.3.1 Load Data Characteristics

There is a complex and nonlinear relationship between the electricity load and its influencing factors. Electric demand on short-term horizon is affected by calendar inputs, weather conditions, and energy price. The calendar events are time of day, day of week, and holidays. It is important to analyze data to find out correlation and dependencies between dependent (load in this case) and independent variables (load affecting variables). As in graph given in Fig. 7.4, relationship between demand and temperature is shown. Here, 5 day profile is presented. It is clear from the graph that spikes in temperature cause increase in demand. However, time of the day is also a major factor in load fluctuation. Whereas, in Fig. 7.5, association of demand with previous load profiles is presented.

![Figure 7.4: Relationship between Temperature and Demand](image)

### 7.3.2 Model Parameters

In this study, ANN and SVM based models both from statistical and AI domains are evaluated based on their accuracy of results. Four optimization techniques are used to get the optimized set of ANN connection weights and best parameters for SVM based regression model. These learning schemes include RPROP, LMA, SA, and PSO. The standard learning parameters for each technique are used in this experimentation, this include settings of each algorithm, maximum iteration count, and evaluation criteria.

**ANN Model Structure**

ANN model used in this study consists of 3 layers of neurons connected to next layer, namely input layer, hidden layer and output layer. Input layer consists of 10 neurons, 7...
neurons in hidden layer, and one output layer. Hence, total weight connections are 78.

**SVM Model**

In machine learning, SVM's are supervised learning models which analyze and recognize patterns in data. In order to obtain better generalization performance through SVM, its learning parameters should be optimized. In this study these parameters are tuned by SA method.

**RPROP**

This technique has two learning parameters. The initial update value for RPROP algorithm is 0.1 and maximum step size is set to 50. Maximum iteration count for RPROP based model is 3000.

**LMA**

The LMA has two parameters named as lambda scale and lambda maximum value. The initial value of lambda scale is 10.0, and maximum value is $1e^{25}$. The maximum iterations count for this algorithm is also 3000.

**SA**

This algorithm has two control parameters, initial temperature $T_1$ and stop temperature $T_2$. In this study, value of $T_1$ is 10.0, and $T_2$ is set to 0.1; which should always be $T_2 \geq 0$. The maximum iteration count for this algorithm is 1000.

**PSO**

The control parameters of PSO are population count $P$, inertia weight $W$, position limit $L_1$, velocity limit $L_2$, and two learning rate constants $C_1$, $C_2$. Here $P$ is 25, $W$ is 0.7,
both $C_1$ and $C_2$ have the same value of 1.49. Here range of $L_1$ and $L_2$ is $[-1.0, 1.0]$. The maximum iteration count for PSO is 3000.

### 7.4 Experimentation

This section covers the simulations of aforementioned optimization techniques based on ANN and SVM models on hourly and daily peak data. The purpose of this experimentation is to determine the forecasting capabilities of these techniques to predict hourly and daily consumption patterns for next few days.

#### 7.4.1 Hourly Demand Estimation

In this section aforementioned modeling techniques are evaluated and compared with accuracy of results on hourly data sets. Sample data entries of LESCO are shown in Table 7.2. It contains hourly demand along other affecting parameters like weather data and calendar data.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Hour</th>
<th>Temperature (°C)</th>
<th>DewPoint (°C)</th>
<th>Humidity (%)</th>
<th>Demand (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>1</td>
<td>31</td>
<td>68</td>
<td>52</td>
<td>3343</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>2</td>
<td>31</td>
<td>68</td>
<td>52</td>
<td>3298</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>3</td>
<td>33</td>
<td>66.2</td>
<td>43</td>
<td>3228</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>4</td>
<td>35</td>
<td>69.8</td>
<td>44</td>
<td>2945</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>5</td>
<td>36</td>
<td>68</td>
<td>39</td>
<td>3024</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>6</td>
<td>38</td>
<td>68</td>
<td>35</td>
<td>2854</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>7</td>
<td>40</td>
<td>68</td>
<td>31</td>
<td>2844</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>8</td>
<td>41</td>
<td>68</td>
<td>30</td>
<td>2743</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>9</td>
<td>42</td>
<td>68</td>
<td>28</td>
<td>2627</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>10</td>
<td>43</td>
<td>68</td>
<td>27</td>
<td>2611</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>11</td>
<td>43</td>
<td>68</td>
<td>27</td>
<td>2683</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>12</td>
<td>43</td>
<td>64.4</td>
<td>23</td>
<td>3086</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>13</td>
<td>43</td>
<td>64.4</td>
<td>23</td>
<td>3130</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>14</td>
<td>42</td>
<td>64.4</td>
<td>25</td>
<td>3187</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>15</td>
<td>41</td>
<td>64.4</td>
<td>26</td>
<td>3099</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>16</td>
<td>36</td>
<td>68</td>
<td>39</td>
<td>3027</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>17</td>
<td>31</td>
<td>66.2</td>
<td>48</td>
<td>3141</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>18</td>
<td>30</td>
<td>64.4</td>
<td>48</td>
<td>3077</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>19</td>
<td>30</td>
<td>64.4</td>
<td>48</td>
<td>3108</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>20</td>
<td>29</td>
<td>64.4</td>
<td>51</td>
<td>3165</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>21</td>
<td>29</td>
<td>64.4</td>
<td>51</td>
<td>2759</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>22</td>
<td>29</td>
<td>64.4</td>
<td>51</td>
<td>2906</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>23</td>
<td>29</td>
<td>64.4</td>
<td>51</td>
<td>3017</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>29</td>
<td>24</td>
<td>26</td>
<td>64.4</td>
<td>61</td>
<td>3087</td>
</tr>
</tbody>
</table>

#### 7.4.2 Daily Peak Demand Estimation

The purpose of this experimentation is to determine the forecasting capabilities of aforementioned techniques to predict daily peak load consumption patterns for next few days. Most influencing factors have been considered that affect the peak load profile fluctuation. These include daily peak weather conditions like max temperature, and calendar inputs including day of week, week of the year, and day of month. Sample daily peak
data of LESCO is shown in Table 7.3. It contains daily peak demand along other affecting parameters like weather and calendar inputs. Since, experimentation is performed on daily peak demand data; therefore, the average weather conditions on daily basis are considered. For this, we have use the the Feels Like index. It is the apparent temperature value, which is calculated based on temperature, humidity, and wind speed. It is the first temperature determined to denote how hot and cold really feels to human body.

<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Weekday</th>
<th>Week of year</th>
<th>Feels Like (C)</th>
<th>Demand (MW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>14</td>
<td>35.24</td>
<td>2908</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>14</td>
<td>32.73</td>
<td>2543</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>15</td>
<td>34.34</td>
<td>2306</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>15</td>
<td>35.22</td>
<td>2812</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>32.70</td>
<td>2825</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>15</td>
<td>30.06</td>
<td>2808</td>
</tr>
<tr>
<td>2011</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>15</td>
<td>33.90</td>
<td>2643</td>
</tr>
</tbody>
</table>

7.5 Results and Discussion

This study was performed on real electricity demand data collected from LESCO. Several load affecting factors; such as: weather conditions and calendar events were also considered for the analysis. This proposed methodology incorporates data pre-processing, analysis, and LF modeling techniques of machine learning. Five days demand profile has been presented in Fig. 7.6. Where, actual demand curve was plotted against the forecasted curves produced ANN-RPROP model. Actual demand data was plotted for two days and next 3 day profile was estimated. In Fig. 7.7, results of ANN trained by PSO-Jordan method are shown. Forecasted curves of next 3 days is produced by ANN-PSO model. Similarly, results achieved by ANN-LMA are shown in Fig. 7.8. This graph shows how closely the predicted demand values are following the actual one’s. Forecasted values of SVM trained by SA method are shown in Fig. 7.9. To compare the results, estimated demand curves of all the four techniques have been plotted in a single graph given in Fig. 7.10, which shows how accurately, all the techniques are following the actual trend line and then producing the forecasts for subsequent 3 days.

Accuracy comparison is presented in Table 7.4, which is compared on the same criteria by taking MAPE. Training, testing, and validation accuracy on hourly data is presented in Table 7.4. It is evident that SA based SVM has outperformed other techniques in terms of accuracy of results. It produced lowest MAPE of training, testing and validation; reaching 97.50% accuracy of forecasts. This comparison is also presented via bar graph in Fig. 7.11, which confirms that SA-SVM model has achieved the best accuracy on this data set.
Two months daily peak profile is presented in Fig. 7.12. Where, actual demand curve is plotted against the curves produced by RPROP-ANN model. Actual daily peak demand data is plotted with estimates of forty days and next twenty days profile is estimated.
Chapter 7. A Generalized LF Methodology

Figure 7.9: Actual v/s Forecasted demand by SVM-SA

Figure 7.10: Comparison of Actual Load with Forecasted values of four Techniques

Table 7.4: Accuracy Comparison between Different Techniques on Hourly Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>Train Accuracy</th>
<th>Test Accuracy</th>
<th>Validation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPROP</td>
<td>96.95 %</td>
<td>97.22 %</td>
<td>97.06 %</td>
</tr>
<tr>
<td>PSO</td>
<td>96.8 %</td>
<td>96.98 %</td>
<td>96.83 %</td>
</tr>
<tr>
<td>LMA</td>
<td>96.85 %</td>
<td>97.16 %</td>
<td>97.04 %</td>
</tr>
<tr>
<td>SA-SVM</td>
<td>97.36 %</td>
<td>97.65 %</td>
<td>97.57 %</td>
</tr>
</tbody>
</table>

by RPROP-ANN technique. In Fig. 7.12, results obtained by RPROP-ANN are shown. Whereas, in Fig. 7.13, results of ANN trained by PSO-Jordan are shown. Similarly, results achieved by ANN-LMA are shown in Fig. 7.14. Forecasted values of SVM trained by anneal method are shown in Fig. 7.15. Estimated demand curves of all the four techniques are compared in Fig. 7.16.
The accuracy comparison on daily load data is presented in Table 7.5. From the obtained accuracy of results, it is clearly evident that all the techniques have produced above 97% accurate results on daily peak demand data. Still, the SA based SVM has shown more generalization capability. It avoided over-fitting and under-fitting scenarios as there is comparatively little difference between train, test, and validation accuracy. This comparison is also presented by bar graph in Fig. 7.17. The graph also confirms that SA-SVM model has outperformed other techniques in generalization ability and accuracy of results. Hence, this technique is confirmed to produce better results.

Figure 7.11: Accuracy Comparison between Different Techniques on Hourly Data

Figure 7.12: Actual v/s Forecasted Demand by ANN-RPROP
Chapter 7. *A Generalized LF Methodology*

Figure 7.13: Actual v/s Forecasted Demand by ANN-PSO

Figure 7.14: Actual v/s Forecasted Demand by ANN-LMA

Table 7.5: Accuracy Comparison between Different Techniques on Daily Data

<table>
<thead>
<tr>
<th>Technique</th>
<th>Train Accuracy</th>
<th>Test Accuracy</th>
<th>Validation Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPROP</td>
<td>96.61 %</td>
<td>97.78 %</td>
<td>97.48 %</td>
</tr>
<tr>
<td>PSO</td>
<td>96.37 %</td>
<td>97.85 %</td>
<td>97.55 %</td>
</tr>
<tr>
<td>LMA</td>
<td>96.71 %</td>
<td>97.63 %</td>
<td>97.37 %</td>
</tr>
<tr>
<td>SA-SVM</td>
<td>97.17 %</td>
<td>97.13 %</td>
<td>97.19 %</td>
</tr>
</tbody>
</table>
Chapter 7. A Generalized LF Methodology

Figure 7.15: Actual v/s Forecasted Demand by SVM-SA

Figure 7.16: Comparison of Actual Load with Forecasted values of Four Techniques
Figure 7.17: Accuracy Comparison between Different Techniques on Daily Data
7.6 Conclusion

In this study, we have contributed by formulating a STLF model for the PSS by integrating machine learning techniques with improved accuracy. We propose an accurate and generic LF model formulation that is fit for indigenous attributers and can fulfill the requirements of our power industry. This formulation has proved our hypothesis to produce the accurate results of forecast on noisy data sets with appropriate use of machine learning techniques for data pre-processing, analysis, and model development [132]. This experimentation is performed on data collected from one utility. Our proposed methodology has performed well on both cases of hourly and daily peak demand data sets. It has shown better generalization ability while avoiding over-fitting and slow convergence. The generalization ability is confirmed by obtaining lower difference between train, test, and validation accuracy, as well as, by cross applicability of same models on two different data sets with almost same accuracy. In this way, our hypothesis to develop more generalized model by incorporating appropriate machine learning techniques is achieved. This methodology is applied on STLF in this chapter, whereas, it is also applied on MTLF and LTLF on regional and national level in Chapter 6.
Chapter 8

Conclusion and Future Work

The general objective of this work is to propose accurate electric LF models for a reliable and economic operation of electric power systems. Several LF models are formulated using individual and integration of techniques for STLF, MTLF, and LTLF. The major focus however, remained at highlighting the behavior of different modeling and optimization techniques on varying data sets. We developed several LF models for Pakistan with greater accuracy on noisy data sets. The local load affecting parameter are studied and incorporated in this study. Machine learning based algorithms are used for data analysis, pre-processing, classification, and regression tasks. Moreover, we evaluated the performance of each technique for its convergence and generalization abilities towards the lesser error of forecasts.

On the whole, this thesis is comprised of four parts: historical data treatment and analysis, evaluation of existing and new hybrid techniques/ models, development of a customized LF system, design of unique hybrid LF models, and application of proposed models on data sets of other domain for the sake of cross-validation. Moreover, some new models are suggested and presented. Our proposed models obtained better accuracy by incorporating local load factors. They have greater ability of generalization and less prone to stuck in local minima. Several statistical and bench-marking tests are applied to verify our contribution and to validate our models.

We have proposed a unique hybrid model of ABC and FFNN for short term load forecasting [73]. It reveals optimization power of ABC algorithm for training multilayer ANN and finding global minimum in an efficient manner. This is a unique formulation to draw attention of researchers towards the innate optimization capability of ABC algorithm for LF and related problems. FFNN weights are adjusted recursively by ABC algorithm to achieve better accuracy of results. Results are presented and compared in tables and graphs, which proved that ABC based LF model is better approach than PSO and GA. A greater accuracy with MAPE of 1.89% is achieved. Accuracy comparison test proves the significance of the forecasting accuracy improvement from the proposed model. In this way, optimization power of ABC optimization algorithm for training multi-layer
ANN is revealed. It is shown how very simple individuals can collaborate and interact with each other to create dynamic system to solve complex problems. Relatively simple form of solution procedures are defined in these hybrid approaches, whose applicability to the solution of complex optimization problems is shown in this study.

The formulation of hybrid model of ACS with ANN is also a unique contribution presented in this thesis. In this work, five different evolutionary optimization techniques are examined to train ANN’s. All optimization techniques are evaluated on different benchmark functions and real world problem of load forecasting from regression domain. From the obtained results, it can be clearly observed that ACS has shown better optimization capability on benchmark functions as well as on regression data set.

Observing that the solution search equation of ABC algorithm is good at exploration but poor at exploitation. Therefore, we extend the proposed ABC based FFNN model by incorporating several modifications in base ABC algorithm to improve its exploitation aspect. Several improvements are incorporated in base ABC algorithm to improve exploitation and further, applied these evolved versions to predict daily and weekly peak electric load demand. Optimization ability of these modified versions is evaluated by statistical benchmark functions. Modifications are applied in series to obtain an integrated and evolved version of ABC. Further, GO’s are applied to witness its effect on optimization and forecasting accuracy. Base algorithm and all the evolved versions of ABC are used to develop LF models. Results obtained from these models are compared with each other and other modern techniques of optimization for ANN, like PSO and SA. Obtained results portray that base algorithm shown better accuracy when modified in search and probability functions; however, GO’s could not improve results and computation time. Authors achieved a greater accuracy with MAPE of 2.49% which is quite remarkable and is comparable to any other model development technique for LF even on noisy data of such utility companies. Computation time of this approach is quite stumpy when compared to other competitive approaches used with ANN like PSO and SA. In future, we intend to extend and evolve the proposed model further by incorporating different techniques for solution initialization, selection, and probability assignment to improve the exploration and exploitation capabilities of ABC algorithm, to achieve better optimization for improved results.

Moreover, we present the performance evaluation of ANN and SVM based hybrid models. Different case studies are included to cover short-term, medium term, and long term load forecasting scenarios [67]. We contribute by evaluating the performance of leading machine learning techniques to propose a best suitable ELF models for the NTDCL, LESCO, and IESO [127]. We have compared ANN and SVM based 20 hybrid models, wherein, nine different optimization techniques are utilized. From the SVM based models, we have achieve 98.91% accurate forecasts. Whereas, ANN based models have proved more successful to produce comparative accuracy and to avoid over-fitting.
The accuracy obtained using these models is greater than accuracy reported in literature. This formulation has satisfied our objectives to produce the accurate results of forecast by using state of the art machine learning model development and optimization techniques. This experimentation was performed on cumulative hourly demand data at national level, collected from NTDCL. The models developed in this research work can be applied to regional power distribution companies to assess the consumer power demand at micro levels.

The existence of abnormalities in historical demand or weather data effects the accuracy of LF models. The problems such as: missing values and occurrences of outliers in data disturb the actual power demand profiles and hence, produce misleading results. The proper usage of data filtration, pre-treatment, and analysis techniques can produce the improved results. Such techniques are combined to formulate a customized LF model, which is utilized in different case studies. We have contributed by formulating a generalized ELF model for the PPS by integrating machine learning techniques with improved accuracy [132]. We propose an accurate and generic ELF model formulation that is fit for indigenous attributeers and can fulfill the requirements of our power industry. This formulation has proved our hypothesis to produce the accurate results of forecast on noisy data sets with appropriate use of machine learning techniques for data pre-processing, analysis, and model development.

Keeping in view the high potential and optimization ability in nature inspired algorithms; in future, we are highly motivated to explore and exploit the optimization ability of several population based algorithms including: ABC, ACS, CS, and FFA by introducing the modifications for enhanced optimization. Further, we aim at exploring new and unique optimization techniques for ANN and SVM based hybrid model development. Moreover, the proposed models can be applied for generation forecasting from different sources, especially renewable energy generation sources, both at macro and micro levels. We also seek to apply the proposed models on classification and regression problems of other domains, for instance; disease diagnosis, casual inference, and sales prediction. We are also interested to explore the dynamic behavior of consumers at micro level by learning energy consumption behavior of residential dwellings.
References


References


References


